UNIVERSITÀ DEGLI STUDI DI UDINE
DIPARTIMENTO DI MATEMATICA E INFORMATICA
DOTTORATO DI RICERCA IN INFORMATICA

PH.D. THESIS

Design and Evaluation in the Large of Health Apps for the General Population with Case Studies in Mindfulness, Neurological and Psychological Assessment

CANDIDATE:
Andrea Vianello

SUPERVISOR:
Luca Chittaro

Academic Year 2014-2015

Author’s e-mail: andrea.vianello@uniud.it

Author’s address:

Dipartimento di Matematica e Informatica
Università degli Studi di Udine
Via delle Scienze, 206
33100 Udine
Italia

Contents

List of Figures

List of Tables

Introduction

In the last decade, the increasing availability of Internet connection has allowed researchers and health professionals to design and deliver Internet-based interventions aimed at promoting health behavior change (Ritterband and Tate, 2009; Webb et al., 2010). This phenomenon has been referred to as *e-health* and defined as “a field in the intersection of medical informatics, public health and business, referring to health services and information delivered or enhanced through the Internet and related technologies” (Eysenbach, 2001). Several e-health interventions are now available to people and can be classified into three categories (Barak et al., 2009; Proudfoot et al., 2011): (i) on-line interactive therapy programs that provide highly specialized treatment and feedback tailored to the characteristics of the user, delivered with and without guidance from a human therapist, (ii) Internet-operated expert systems in which assessment and therapeutic techniques based on decision rules and behavior change strategies are integrated, and (iii) education programs, digitized information, and online therapeutic communication services such as online counseling, online discussion and support groups also through social networking sites, and ”ask an expert” websites.

These interventions address a variety of psychological, behavioral or physical health conditions and studies in the literature have shown their feasibility and efficacy - see e.g. (Andersson et al., 2011; Grajales III et al., 2014; Griffiths et al., 2006; Moorhead et al., 2013; Webb et al., 2010) for extensive reviews.

In recent years, the diffusion of smartphones and tablets has offered researchers and health professionals new opportunities for developing and delivering e-health interventions. Indeed, mobile devices offer computing capabilities comparable to those of desktop computers and all major mobile platforms, such as iOS and Android, give developers the possibility to build special purpose apps. Mobile platforms let also apps to obtain data from the sensors embedded into mobile devices, such as GPS, accelerometer or barometer, or from external sensors that can be easily connected, such as wearable devices. This data can allow mobile apps to infer where their users are and what they are doing (Klasnja and Pratt, 2012; Miller, 2012).

For these reasons, mobile apps are now exploited by health care professionals for several purposes, such as medical training, disease diagnosis or the delivery of tailored interventions to clinical patients, creating a branch of e-health referred to as *m-health* (Free et al., 2010) - see (Mosa et al., 2012) for a review. These opportunities have also been explored by researchers outside the medical field that proposed a number of mobile apps that are aimed at helping people learn, track and improve health conditions and behaviors (in the following, *health apps*) (Boudreaux et al., 2014; Klasnja and Pratt, 2012), see (Donker et al., 2013; Fiordelli et al., 2013; Free et al., 2013; Klasnja and Pratt, 2012) for reviews. This has also led the availability of a large number of health apps in on-line app stores, such as Apple’s App Store or Google Play (Sama et al., 2014; West et al., 2012). In most cases, these apps exploit the integration with Facebook to give people the possibility to share their health behaviors with others or challenge other people in achieving a health-related goal, e.g. lose weight or reach a specific fitness target (Klasnja and Pratt, 2012). In addition to the mobile context, Facebook can be integrated with desktop on-line apps. As a consequence, a high number of health-apps have appeared also on it (Fernandez-Luque et al., 2010).

However, despite the high availability of health apps for the general population, very few of them have been created by healthcare experts and very few of them have been scientifically evaluated, see e.g. (Abroms et al., 2011; de la Vega and Miró, 2014; Fernandez-Luque et al., 2010; Fiordelli et al., 2013; Hundert et al., 2014; Wolf et al., 2013) for reviews. This scenario points out two problems:

- users do not have a health professional that monitors how they use such health apps. As a consequence, these apps can be used in unpredicted ways. Moreover, too often there is nobody responsible and available if an app is not working as expected or if something goes wrong (de la Vega and Miró, 2014);

- in most cases, such apps do not follow established guidelines or strategies for their health-related purposes. If not scientifically evaluated, it is not possible to know if their use is beneficial to users. Indeed, in the worst case, such apps can be detrimental to people. Moreover, unproven claims about the effectiveness of an app may also lead users to a feeling of helplessness and lack of control about their illnesses (de la Vega and Miró, 2014; Wolf et al., 2013).

The goal of this thesis is to explore how some HCI methods can be used for designing and evaluating health apps that target the general population. To this purpose, we present three case studies in different domains, i.e. (i) mindfulness, (ii) neurological assessment, and (iii) psychological assessment and propose a health app for each domain. Then, to improve the design or assess the efficacy of the app in different contexts of use we employ some of the traditional HCI methods, i.e. quantitative as well qualitative, in the lab as well as in situ, and a recently proposed approach, i.e. research in the large. Research in the large is a methodology that embeds a research apparatus into a mobile or Facebook app and makes it publicly available to attract a possibly large number of users.

Overall, the work presented in this thesis aims at addressing the following research questions:

- *RQ1: Can a mobile mindfulness app be as effective as traditional mindfulness techniques?*
- *RQ2: Can a mobile mindfulness app help people increase their level of mindfulness over time?*
- *RQ3: Can a mobile assessment app provide users with a reliable assessment of their neuromotor performance?*
- *RQ4: How to employ research in the large to evaluate health apps in situ?*
- *RQ5: How to employ research in the large to collect data for scientific studies?*
- *RQ6: What kind of data research in the large can provide with respect to lab evaluations?*

This thesis is organized as follows: in Chapter 1 we provide an overview of the available health apps for the general population and introduce research in the large, discussing also its strengths and weaknesses.

Chapter 2 focuses on AEON, the mobile mindfulness app that we propose. In particular, we first describe the design process of AEON that we carried out by analyzing the difficulties that people could encounter when approaching mindfulness. Then, we illustrate the three evaluations that we conducted to evaluate its effectiveness in helping people practice mindfulness in different contexts of use, i.e. (i) a quantitative lab study, (ii) a qualitative in situ 5-week study and (iii) a quantitative in situ 4-week study with research in the large.

Chapter 3 deals with MotorBrain, the mobile app for neurological assessment that we developed by collaborating with a domain expert. The app offers gamified versions of six motor tests traditionally used by neurologists for the assessment of users' motor skills. Moreover, it aims at collecting a large amount of motor skills data for building a database to be used in neurological studies. In Chapter 3, we describe the first prototype of the app that we developed, as well as the qualitative study we carried out to investigate users' experience with it. Then, we illustrate the second prototype of MotorBrain that we designed. Finally, we describe the quantitative evaluation we employed to investigate whether the data collected by the app can be effectively used for the assessment of users' motor performance.

Chapter 4 focuses on Psicoscrigno, the Facebook app for psychological assessment that we propose. The app offers people a standardized questionnaire for the assessment of problematic Internet use (PIU). In this case study, we exploited the possibilities of research in the large to make the app available to a large number of users as well as carrying out a correlational study to better describe PIU in terms of personality traits. To this purpose, we included into Psicoscrigno also a valid and reliable assessment of time perspective (TP), i.e. an individual difference that has been related to different problematic human behaviors. In Chapter 4 we first describe Psicoscrigno. Then, we describe and discuss the correlational study we carried out with research in the large to study the relationship between TP and PIU.

Finally, the thesis draws conclusion and outlines future work.

1

Thesis Background

In this chapter, we introduce two key topics of this thesis: (i) *health apps*, and (ii) *research in the large*. In particular, in Section 1.1, we briefly outline the scope and purposes of the health apps that are available to the general population. Then, we provide two classifications for such apps, one based on their functions and goals, and one based on the health conditions and behaviors that they target. In Section 1.2, we first provide the original definition of research in the large and the one we propose to extend its scope. Then, we describe the advantages and the possible challenges of using such study methodology.

1.1 Health Apps for the General Population

Currently, a growing number of health apps is available on on-line app stores (Sama et al., 2014; West et al., 2012), such as Apple's App Store and Google Play, as well as on social networks (Fernandez-Luque et al., 2010; Grajales III et al., 2014) such as Facebook. Such apps address a wide range of health conditions and behaviors, such as headache or healthy eating, and offer people the possibility to (i) learn about health, wellness and fitness, (ii) track health conditions or behaviors, and (iii) improve health outcomes (Boudreaux et al., 2014; Klasnja and Pratt, 2012). To offer their functionalities, health apps can exploit the fact that mobile devices are always with people. Indeed, mobile apps can collect data from the smartphone's internal sensors, e.g. GPS or accelerometer, or from external devices, such as wearable technologies to infer where they users are and what they are doing. This information can be used to trigger clues for behavior change at the most appropriate time or offer people methods to easily collect their health-related data. Finally, the integration with social networks such as Facebook offers users of desktop and mobile apps the possibility to share their health conditions or behaviors with others or challenge other people in achieving a health-related goal, e.g. lose weight (Klasnja and Pratt, 2012; West et al., 2012).

Users can easily find and install health apps by browsing on-line app stores or Facebook's App Center, i.e. a web page where all the integrated mobile and desktop apps can be found. In particular, on-line app stores, such as Apple's App Store and Google Play, offer users a specific category, called "Health & Fitness", where all the available health apps are listed. According to a recent market research, in 2013 there were 97000 health apps on on-line app stores (research2guidance, 2013). Recently, a number of web sites that offer users a list of the available health apps have emerged, e.g. My Health Apps (Myhealthapps.net, 2014) - see (Boudreaux et al., 2014) for a review. In particular, such web sites are aimed at evaluating the usability, functionality, accuracy of the content and efficacy of health apps to be used by consumers, healthcare providers and healthcare organizations (Boudreaux et al., 2014).

However, despite the high availability of health apps, very few of them have been scientifically evaluated or created by healthcare experts. Moreover, very few of them adhere to well-established strategies for health behavior change, e.g. the Theory of Planned Behavior (Ajzen, 1991), or practices employed in traditional health interventions. The only health apps that underwent such scientific scrutiny are those created ad-hoc for research purposes, but very few of them are available to the general population, see e.g. (Abroms et al., 2011; de la Vega and Miró, 2014; Donker et al., 2013; Fiordelli et al., 2013; Hundert et al., 2014; Klasnja and Pratt, 2012; Weaver et al., 2013) for reviews.

In the following, we employ the two classifications proposed by West et al. (2012) to briefly survey the existing types of health apps for the general population. Such classifications were originally proposed for mobile health apps but they can be applied also to desktop health apps on Facebook. In particular, the first classification considers the functions or goals of the available health apps and is adapted from the Precede-Proceed Model (PPM) for health promotion planning and evaluation (Green and Kreuter, 1991). The second classification considers the health conditions or behaviors that are targeted by the available health apps and is based on the Health Education Curriculum Analysis Tool (HECAT) (US Department of Health and Human Services, 2008).

1.1.1 Functions or Goals of the Available Health Apps

The Precede-Proceed Model is a framework for the evaluation and design of effective health interventions (Green and Kreuter, 1991). According to the PPM model, a health behavior is influenced by both individual (Precede) and environmental (Proceed) factors.

West et al. (2012) focus on the individual factors of the PPM model for classifying the available health apps according to their functions and goals. Such factors are: (i) *predisposing factors*, (ii) *reinforcing factors*, and (iii) *enabling factors*, and they can be used to classify a health app as a (i) *predisposing app*, (ii) *enabling app*, and (iii) *reinforcing app*. According to West et al. (2012), the three categories are not mutually exclusive. For example, a health app can be both predisposing and enabling. In the following, we briefly describe such factors and the health app categories they define.

Predisposing and enabling factors are antecedents to behavior. In particular, predisposing factors include knowledge, attitudes, beliefs, personal preferences, existing skills, and self-efficacy towards the desired behavior change (Green and Kreuter, 1991). Predisposing apps are those health apps that precede a target behavior and aim at: (i) offering people knowledge or fostering awareness of health conditions or health outcomes, e.g. by providing information on risks of alcohol abuse, (ii) providing health information, e.g. by presenting information regarding ways to prevent sexually transmitted diseases, (iii) influencing users' beliefs, values, or attitudes, e.g. by discussing common reasons to avoid tobacco in an effort to assist the user in quitting smoking, and (iv) enhancing users' confidence or motivation, e.g. by convincing people that they can change their diet (West et al., 2012).

Enabling factors are skills or physical factors, such as the availability of resources, or services, such as tracking capabilities, that can facilitate the achievement of motivation to change behavior (Green and Kreuter, 1991). Enabling apps are those health apps intended to be used at or around the same time as the desired behavior. In particular, such apps are aimed at facilitating a health behavior change through: (i) teaching a skill, e.g. by showing pictures and instructions on how to perform healthy stretching, (ii) providing a service, e.g. by geo-locating places for physical activity, and (iii) tracking progress and/or recording behavior, e.g. by offering a calorie counter or by automatically tracking the distance that the user has run (West et al., 2012).

Finally, reinforcing factors include factors that reward or reinforce the desired behavior change, including evaluations or feedback, social support, and rewards (Green and Kreuter, 1991). Reinforcing apps are those health apps that (i) interface with a social network, e.g. by automatically uploading to Facebook the distance that the user has run, (ii) provide encouragement from trainers or coaches, e.g. by offering an easy communication with a coach, and (iii) include an evaluation based upon the user's self-monitoring, e.g. by providing feedback about user's reports of his/her physical activity (West et al., 2012).

1.1.2 Health Conditions or Behaviors Targeted by Health Apps

The Health Education Curriculum Analysis Tool (HECAT) is an assessment tool developed by the U.S.'s Centers for Disease Control and Prevention, and contains guidance, appraisal tools, and resources to conduct a clear, complete, and consistent examination of health education curricula (US Department of Health and Human Services, 2008). In particular, the HECAT defines eight health education content areas, i.e. (i) *alcohol, and other drugs*, (ii) *healthy eating*, (iii) *mental and emotional health*, (iv) *physical activity*, (v) *safety*, (vi) *sexual health*, (vii) *tobacco*, and (viii) *violence prevention*.

West et al. (2012) merged some content areas, i.e. (i) alcohol, and other drugs with (ii) tobacco, and (iii) safety with (iv) violence prevention. Then, they employed the resulting seven categories to classify the available mobile health apps. Also in this case, the categories are not mutually exclusive. We describe such categories in the following.

1.1.2.1 Alcohol, Tobacco, and other Drugs

The health apps in this category are those aimed at supporting people in the reduction or recovery from alcohol, tobacco or drug addiction. To this purpose, such apps can offer people information on the risk of substance abuse and on how to find and share meetings with experts or consultants. Other strategies that these apps employ include tracking capabilities, e.g. for letting people log the days since they quit smoking, triggering recommendations or advice, and giving motivational enhancement if the user is succeeding in reducing a substance use (Abroms et al., 2011; Cohn et al., 2011; Weaver et al., 2013).

Extensive reviews of the available mobile health apps related to alcohol consumption are provided in (Cohn et al., 2011; Weaver et al., 2013), while Abroms et al. (2011) provide an overview of the available mobile health apps for smoking cessation.

1.1.2.2 Healthy Eating

The available health apps for healthy eating are those aimed at helping people follow a healthy diet and/or lose weight. To this purpose, such apps provide users with information on healthy meals or drinks, nutritional breakdown of specific food items, and healthy recipes and cooking tips. These apps can also suggest people local groceries or restaurants where to buy or consume healthy food, and persuade them to perform physical activities. Moreover, these apps can offer users tracking capabilities, e.g. calories counters, journals of eaten food or logs of physical activities performed. Finally, they can provide users with the possibility to set a weight loss or maintenance goal and share their progress with friends, e.g. on Facebook, or with an expert (Azar et al., 2013; Breton et al., 2011; Pagoto et al., 2013; West et al., 2012).

Extensive reviews of mobile health apps in this category are provided in (Azar et al., 2013; Breton et al., 2011; Pagoto et al., 2013).

1.1.2.3 Mental and Emotional Health

The health apps in this category address mental or emotional health symptoms or disorders, such as stress, gambling, binge eating, depression or anxiety (Donker et al., 2013; West et al., 2012). Such apps can offer people information on a specific disorder and reference tests or diagnostic tools, e.g. an interactive questionnaire to assess anxiety. In Chapter 4, we provide a survey of the desktop on-line apps that offer an assessment of a particular mental health condition, i.e. problematic Internet use.

Health apps in this category can also provide people with self-help treatments and therapies to reduce a specific disorder, e.g. by presenting teaching materials through audio or videos. Such treatments can also offer people the possibility to contact an expert. A survey of the available mobile health apps for depression is provided in (Martínez-Pérez et al., 2013).

Recently, a number of mobile apps for mental health that are based on meditation and mindfulness techniques have appeared, see (Plaza et al., 2013) for a review of mindfulness apps. In Chapter 2, we provide a survey of those mobile apps that exploit some form of interactivy to support people practice midfulness.

1.1.2.4 Physical Activity

Health apps in this category are those aimed at promoting physical activities among users. To this purpose, they can provide information and guidance for performing physical exercises, suggest places or facilities to perform them, and provide a calendar of local races and competitions. Other strategies that such apps can employ include the monitoring and automatic logging of physical activities, e.g. by collecting data from

GPS or wearable devices, the provision of performance feedback to users, and the possibility to contact a coach or a trainer. Finally, such apps can offer users the possibility to set specified goals to achieve, can give motivational enhancement, e.g. if intermediate goals have been achieved, and let them share their progress with friends, e.g. through Facebook (Cowan et al., 2012; Middelweerd et al., 2014; West et al., 2012).

Extensive reviews of the available mobile health apps for physical activity are provided in (Cowan et al., 2012; Middelweerd et al., 2014).

1.1.2.5 Violence Prevention and Safety

The available health apps in this category focus on (i) first aid and emergency preparedness, and (ii) injury prevention (West et al., 2012). To this purpose, such apps can employ video, interactive quizzes, or offer step-by-step guides to teach people how to handle common emergencies, such as an asthma attack, or to avoid common injuries, such as snakebite.

A review of mobile apps for sports injury prevention is provided in (van Mechelen et al., 2014).

1.1.2.6 Personal Health and Wellness

Health apps in this category focus on hygiene, oral or skin care, sleep, and specific diseases, e.g. anemia or migraine (West et al., 2012). Such apps can offer information on how to maintain a health behavior such as oral hygiene, how to prevent a specific disease such as cancer, and how to check its symptoms. Apps in this category can also offer possible remedies based on traditional or complementary and alternative medicine, e.g. by offering videos of exercises for alleviating physical pain. Other features that such apps can offer include methods for managing a particular health condition, such as reminders for medications and tracking capabilities for medications and disease-related parameters, e.g. level of insulin or headache intensity.

Several reviews are available for the health apps in this category. For example, Martínez-Pérez et al. (2013) surveyed the available mobile health apps for iron-deficiency anemia, hearing loss, migraine, low vision, asthma, diabetes mellitus, osteoarthritis, and unipolar depressive disorders. Other reviews focus on a specific health condition, such as diabetes (Chomutare et al., 2011; El-Gayar et al., 2013), cancer (Bender et al., 2013), melanoma (Wolf et al., 2013), headache (Hundert et al., 2014), and physical pain (de la Vega and Miró, 2014; Rosser and Eccleston, 2011).

1.1.2.7 Sexual and Reproductive Health

The health apps in this category focus on sexuality, reproduction, pre and postnatal care, and sexually transmitted diseases (STDs) (West et al., 2012). In particular, the apps that target sexuality and reproduction offer users automatic or manual methods for logging sexuality activities, ovulation and menstrual cycles (Lupton, 2014). The apps that target STDs aim at informing the user on how to reduce sexual transmission risks or where to finding testing centers. Such apps can also offer self-diagnosis tests for STDs. Other apps in this category specifically target people diagnosed with HIV or other STDs, and offer them medication adherence tracking and reminders, medication interaction information, medical appointment calendars and reminders, doctor/clinic names and locations, and symptoms/side effects trackers (Muessig et al., 2013).

A review of the available mobile health apps for the care and prevention of HIV and other STDs is provided in (Muessig et al., 2013), while Lupton (2014) reviews mobile health apps for sexuality and reproduction self-tracking.

1.2 Research in the Large

The term *research in the large* originates from the HCI community and indicates a methodology that embeds a research apparatus into a mobile app and makes it publicly available on app stores, such as Apple's App Store or Google Play, to attract a possibly large number of users (Poppinga et al., 2012). Mobile apps can collect data from users as well as from devices and, through the integration with Facebook, capture also users' social activities. Indeed, by offering users the possibility to log in, Facebook allows a mobile app to collect data about their social behavior, such as how they chat or become friends. In this way, researchers can gain data for statistical analysis, run studies with a heterogeneous sample of participants and observe behavior in naturally occurring user contexts (Böhmer and Krüger, 2014; Henze and Pielot, 2013; Miller, 2012; Wilson et al., 2012).

In addition to mobile apps, Facebook allows also desktop on-line apps to be integrated into its platform and collect data about users social behavior (Wilson et al., 2012).

In this thesis, we broaden the scope of research in the large, considering it as "a methodology that embeds a research apparatus into a mobile or desktop on-line app (in the following, *research app*) and makes it publicly available on app stores or Facebook".

Research in the large has been used in several studies within the HCI community, see (Böhmer and Krüger, 2014) for a review, but also in the field of psychology, e.g. (Bless et al., 2013; Dufau et al., 2011; Killingsworth and Gilbert, 2010) or social science, see (Błachnio et al., 2013; Wilson et al., 2012) for reviews. In the following, we focus on those studies that investigated the research approach itself with the aim of providing guidelines for other researchers and identifying possible challenges. Such studies are summarized in Table 1.1.

Table 1.1: Studies that investigate advantages and challenges of using research in the large

Study	Study goal and brief description
(Blunck et al., 2013)	Investigate the source of data heterogeneity when using research in the large. Description of three case studies with three mobile apps that collected sensors' data.
(Coulton and Bamford, 2011)	Study the effects that the behavior of an app store has on a mobile app lifecycle. Description of the design, implementation and distribution of three mobile games.
(Cramer et al., 2010)	Summarize advantages and challenges of research in the large.
(Ferreira et al., 2012)	Identify advantages and challenges of using research in the large. Description of a 4-week long study of a mobile app that collected data on users' habits with battery charging.
(Henze et al., 2011a)	Identify advantages and challenges of using research in the large. Description of five experiments carried out with five mobile apps.
(Henze and Pielot, 2013; Henze et al., 2013)	Summarize advantages of using research in the large based on the authors' previous experience. Provide a step-by-step guide for conducting large-scale studies with research in the large.
(Kranz et al., 2013)	Investigate users' behavior with app updates and the adoption of near-field communication technology (NFC) by publishing a gamified app.
(Schleicher et al., 2011)	Identify advantages and challenges of using research in the large. Description of a 4-week study conducted with a mobile app that let users share real-time opinions during the soccer 2010 world cup.
(McMillan et al., 2013)	Provide a set of ethical guidelines for conducting studies with research in the large.

Table 1.1: Studies that investigate advantages and challenges of using research in the large

Study	Study goal and brief description
(Miller, 2012)	Identify advantages and challenges of using research in the large for running psychological experiments.
(Miluzzo et al., 2010)	Identify advantages and challenges of using research in the large. Description of the development, distribution and support of a mobile app for creating social networks by using the smartphone's onboard sensors.
(Wilson et al., 2012)	Identify advantages and challenges of using research in the large for studying human social behavior.

In the following, we point out the advantages and challenges of using research in the large that have been identified by the studies summarized in Table 1.1.

1.2.1 Advantages of Research in the Large

We briefly illustrate the advantages offered by research in the large according to the following classification: (i) *opportunities*, i.e. what are the opportunities offered by research in the large, (ii) *ecological validity*, i.e. what are the aspects that can increase the ecological validity of a study carried out with research in the large, and (iii) *external validity*, i.e. what are the factors that can enhance the external validity of an experiment carried out with such research approach.

1.2.1.1 Opportunities

Almost all studies summarized in Table 1.1 agree that research in the large gives the possibility to reach a large number of users. Indeed, by publishing an app on on-line app stores or Facebook, thousands of participants can be recruited for a study, thus not limiting the recruitment on those participants available in the local area of researchers.

In this way, a lot of data can be gathered that can be used for:

- remotely evaluate an app, e.g. test the performance of an algorithm (Cramer et al., 2010);
- answer specific research questions or run controlled experiments (Ferreira et al., 2012; Henze and Pielot, 2013; Henze et al., 2013; Kranz et al., 2013; Miller, 2012; Wilson et al., 2012). For example, the concept behind a research can be verified by analyzing the data collected when an app is actually used by people. Alternatively, two different versions of an app, each one representing a different experimental condition, can be compared and users can be randomly assigned to one of the two conditions;
- develop and refine an app (Kranz et al., 2013; Ferreira et al., 2012; Miluzzo et al., 2010). For example, an early version of an app can be published. Then, by analyzing the data collected by the app and the feedback provided by the users, bugs could be identified and resolved, and new functionalities added. Such modifications can be delivered by publishing an app update and the whole process can be iteratively repeated.

1.2.1.2 Ecological Validity

Another important aspect of research in the large is that it allows to increase the ecological validity of experiments (Cramer et al., 2010; Henze and Pielot, 2013; Miller, 2012; Wilson et al., 2012). Ecological validity concerns the extent to which findings of an experiment generalize to other settings, e.g. from the laboratory to a field setting or from one field setting to another (Coolican, 2009). With research in the large, it is possible to unobtrusively observe how people use a mobile app or a smartphone in their everyday life and also capture details about their context of use, e.g. by tracking how users interact with other apps or where they use the smartphone through GPS data. More insight on user's context can also be obtained by combining qualitative and quantitative data (Cramer et al., 2010; Henze and Pielot, 2013; Miller, 2012). Finally, if a research app exploits the integration with Facebook, research in the large makes it possible to capture human social behavior, e.g. how people chat and become friends, that was difficult to assess with traditional studies (Wilson et al., 2012).

1.2.1.3 External Validity

In addition to higher ecological validity, research in the large allows to enhance the external validity of studies (Ferreira et al., 2012; Henze et al., 2011a; Henze and Pielot, 2013; Miller, 2012; Wilson et al., 2012). External validity refers to the extent to which the findings of an experiment generalize beyond the exact experimental setting, e.g. to other people or other cultural contexts (Coolican, 2009). With a large number of users from different countries, it is possible to (i) compare data among different cultural contexts or demographic groups and (ii) derive conclusions that are not limited to a particular country or culture but apply to the global population (Henze and Pielot, 2013; Wilson et al., 2012).

However, there are a lot of challenges that have to be addressed when using research in the large, which are described in the following section.

1.2.2 Challenges of Research in the Large

We briefly illustrate the challenges of using research in the large according to the following categories: (i) *internal validity*, i.e. what are the factors that can threaten the internal validity of the experiments carried out with such research methodology, (ii) *data heterogeneity*, i.e. what are the sources and the types of heterogeneities that can be present in the data collected, (iii) *qualitative data*, i.e. what are the reasons that make it difficult to obtain qualitative data, (iv) *demographic data*, i.e. what are the factors that make it difficult to obtain and verify demographic data, (v) *ethical*, i.e. what are the ethical principles that can be violated when using research in the large, and (vi) *apps adoption and usage*, i.e. what are the causes that could impede a high adoption or a long-term use of a research app.

1.2.2.1 Internal Validity

One of the main challenges of research in the large is the fact that the internal validity of the experiments could be threatened (Henze and Pielot, 2013). Internal validity reflects the extent to which there is a causal effect between the independent and dependent variable of an experiment. It is high when the manipulation of the former can be held responsible for the observed effects in the latter (Coolican, 2009). When carrying out experiments with research in the large, it is impossible to directly observe users. As a consequence, users could use a research app in an unpredictable way, creating another experimental condition, or they could directly manipulate the independent variable, for example by turning on or off a certain app feature that serves as the experimental condition (Ferreira et al., 2012; Henze et al., 2011a; Miller, 2012; Miluzzo et al., 2010). Moreover, there could be environmental conditions that could act as confounding factors, which are typical of HCI studies conducted outside the lab (York and Pendharkar, 2004). For example, users can use an app when riding a bus, a condition that can affect touch input capabilities, or they can use an app outdoor and the sunlight can affect screen readability. However, as pointed out by Henze and Pielot (2013), the large number of users that can be reached with research in the large can factor out individual differences and contextual effects, allowing studies to maintain a reasonable internal validity.

1.2.2.2 Data Heterogeneity

Other challenges of research in the large reside in the large amount of data that can be collected. In particular, when using a mobile research app there could be a lack of homogeneity due to the different hardware and operating systems of smartphones (Blunck et al., 2013). Since users are not directly observed and their qualitative feedback could be missing, a lot of effort could be needed to understand such data (Cramer et al., 2010; Ferreira et al., 2012; Henze et al., 2011a; Miller, 2012; Schleicher et al., 2011). Blunck et al. (2013) identified three types of data heterogeneity that can emerge when collecting data from mobile apps, but can be applied also to desktop on-line apps: (i) *user heterogeneity*, (ii) *device heterogeneity*, and (iii) *project heterogeneity*.

User heterogeneity refers to the data heterogeneity caused by users. Indeed, as pointed out before, they could use a research app or their smartphone in unpredictable ways, generating different types of data. In addition, they might customize their smartphone or web browser in manners that impact the collection of data, e.g. by enabling and disabling location or communication features, such as Wi-Fi or Bluetooth, or modifying access rights for apps.

Device heterogeneity refers instead to the data heterogeneity that is specifically due to the characteristics and capabilities of the mobile devices used for the data collection. Nowadays there is a high range of smartphone models that present crucial dissimilarities across their hardware, such as GPS, network chips and sensors, and the availability of such hardware. For this reason, it could be difficult to directly compare sensing data from one smartphone to another, even if the devices are of the same model, as they could have different hardware or firmware revisions. Other sources of data heterogeneity can be due to the different operating systems, even different versions of the same one. For example, a particular sensor might not be accessed from a specific platform, or a service may behave differently across different operating system revisions.

Finally, project heterogeneity indicates the data heterogeneity that could emerge when different versions of the same mobile app are available. For example, an app upgrade might introduce new features, such as use of hitherto unused sensors, user input elements, or change data representations.

For all of the above described reasons, it could be difficult to understand the data collected by a research app or compare data generated by different users (Miller, 2012).

1.2.2.3 Qualitative Data

As pointed out in the previous section, another challenge of research in the large is the fact that it could be difficult to get qualitative feedback from users, since they are not met in person. For this reason, it could be hard or impossible to understand participants' behavior or answer a particular research question (Ferreira et al., 2012; Henze et al., 2011a; Miluzzo et al., 2010; Schleicher et al., 2011).

For mobile research apps, a source of qualitative data could be the comments that users write on the app's description page inside the app store. However, not all app stores allow developers to directly reply to users, and thus clarify the aspects that are not clear. Furthermore, users comments and ratings cannot be deleted. As a result, since they are particularly trusted among users, the bad comments and low ratings could affect users' willingness to download a research app (Ferreira et al., 2012; Kranz et al., 2013). For desktop research apps, users can only provide a rating by default and, also in this case, it cannot be deleted. However, Facebook allows developers to connect a description page to their apps. When available, users can use such page to send a direct message to developers or post a comment. In this case, users' comments can easily be deleted by developers.

Another source of qualitative data for mobile research apps could be the e-mail address that on-line app stores require developers to indicate when publishing an app. However, this method tend to be used very rarely by users (Ferreira et al., 2012; Henze et al., 2011a).

As a possible solution, originally proposed for mobile apps but applicable also to desktop apps, Ferreira et al. (2012) and (Henze et al., 2011a) suggest to use a feedback form directly into the research app. This mechanism could make it easy for users to send their comments on the app and, at the same time, possibly avoid that they provide low ratings or write bad comments on the on-line app stores or on a Facebook page. Moreover, to encourage users, the app can reward them each time they provide feedback, e.g. by unlocking new features. However, it is impossible to avoid that users provide feedback only to get the rewards, thus obtaining nonsense words. As suggested by Henze et al. (2011a), if deep qualitative data is needed, other study methodologies should be considered.

1.2.2.4 Demographic Data

Other than qualitative data, also demographic data could be difficult to obtain when using research in the large. Indeed, on-line app stores do not share demographics of users and also when such data can be collected it is not possible to verify it, as users are not met in person (Cramer et al., 2010). The situation might be different in the case of Facebook, since it requires users to indicate their demographic data at the time of registration and an integrated app can ask users the permission to access it. Moreover, several studies in the literature have found that the information provided by users on their Facebook profile tend to be accurate - see (Wilson et al., 2012). As pointed out by Wilson et al. (2012), this might be explained by the fact that people become friends on Facebook after being off-line friends. Thus, if a user presents inaccurate information on the Facebook profile, his/her on-line friends, who are also friends in the real world, would realize that the user was not telling the truth.

However, when demographic data can be obtained, other challenges must be taken into consideration:

- smartphone or Facebook users could not reflect the general population (Cramer et al., 2010; Ferreira et al., 2012; Henze et al., 2011a). For example, as they have to use a smartphone or a desktop computer, they could be more tech-savvy people;
- users of different mobile platforms can be very different. For example, as pointed out by Cramer et al. (2010), "a common stereotype is that more wealthy users or design-oriented users are associated with the iPhone platform, while geeks with the Android one".

For these reasons, the results of an experiment carried out with research in the large methodology cannot easily be generalized to the world population. However, this situation may change in the near future given the increasingly adoption of smartphones and Facebook by users. Indeed, a recent market research pointed out that the number of smartphone users will reach 2.73 billion by 2018, with a penetration by population of 36.25%. Similarly, Facebook users will reach 1.56 billion by 2018, with a penetration by population of 20.9% (EMarketer, 2014).

1.2.2.5 Ethical

As in traditional HCI user studies, there are some ethical principles that must be taken into consideration when using research in the large. Several studies listed in Table 1.1 agree that users must be informed about the nature of the study and the data collection purposes and process. This information should be included into a Terms and Conditions (T&C) page shown after the first launch of a research app, or in a web page whose link is provided by the app, and possibly written in different languages. Moreover, the research app should only collect data related to its purposes. Users should be able to opt-out from a study and they should have the possibility to ask for the deletion of their collected data. Finally, data must be anonymized, and sent to the researchers and saved in a secure manner. Given the increasing adoption of research in the large, sets of ethical guidelines have emerged, e.g. (McMillan et al., 2013), which synthesize the principles above described.

These principles are implicitly implemented when exploiting the integration with Facebook. Indeed, Facebook requires each app to explain what information it will collect. Moreover, it requires developers to indicate a web page that explains how the data will be saved and used (Wilson et al., 2012). However, studies in the literature pointed out that very few users read the T&C page and thus they are not aware that their data will be collected (Cramer et al., 2010; McMillan et al., 2013). Thus, given the lack of in-person meetings with users, it could be very difficult to brief them about the nature of the study and to see whether they are capable to take part, e.g. they are an adult of sound mind. Debriefing could also become impossible when participants dropped out from the study by deleting a certain research app or by not using it anymore (McMillan et al., 2013).

1.2.2.6 Apps Adoption and Usage

Other challenges of research in the large are related to users app adoption. Since on-line app stores and Facebook offer a high number of apps, and they do not have a category dedicated to research apps, it could be difficult for users to find and install a certain research app (Ferreira et al., 2012; Schleicher et al., 2011).

Other problems that could affect the adoption and usage of a research app can be due to the specific platform characteristics and behavior. For example, the Apple's App Store requires a review time before a submitted app will be made publicly available. This can delay the diffusion of an app and of its subsequent updates (Cramer et al., 2010; Miluzzo et al., 2010; Schleicher et al., 2011). Moreover, if the platform of an app store or Facebook encounters some problems, a research app could not work as expected or be available any more. For example, Coulton and Bamford (2011) published their apps on Nokia WidSets, an early app store run by Nokia until 2009. At a certain point during one of their study, the app store was updated to provide new functionalities to the apps. However, a bug in the update made the app under evaluation not work correctly and as a consequence all participants abandoned it. Furthermore, in a subsequent moment, the app store was taken offline, putting an end to the study.

Another challenge regards the fact that research apps are used only for a short period of time (Coulton and Bamford, 2011; Ferreira et al., 2012; Henze et al., 2011a; Miluzzo et al., 2010; Schleicher et al., 2011). For example, Coulton and Bamford (2011) found that generally users of their two research apps used them only on the day of the download. Similarly, Schleicher et al. (2011) pointed out that in their study, 42% of participants of used the research app occasionally or just once, with only 27% of users were still active at the end of the 4-week study period. This could due to the fact that users are not accustomed to research apps and to participate in a study by installing them and need to perceive they are receiving clear benefits from using them (Ferreira et al., 2012; Miluzzo et al., 2010). For these reasons, it is difficult to carry out long term studies with research in the large (Coulton and Bamford, 2011; Henze et al., 2011a; Schleicher et al., 2011).

To encourage a high adoption and long-term usage of research apps, some researchers have suggested to:

- advertise a research app, e.g. inside Facebook or the app store (Cramer et al., 2010; Schleicher et al., 2011). However, as the advertisements can require costs, researchers should plan in advance a budget for that (Cramer et al., 2010);
- design apps that are very good-looking, similar to commercial apps (Coulton and Bamford, 2011; Kranz et al., 2013; Schleicher et al., 2011). Moreover, apps must be bug-free, stable, well-performing and, in the case of mobile app, seamlessly integrated with the operating system appearance and behavior (Ferreira et al., 2012; Henze et al., 2011a; Miller, 2012; Miluzzo et al., 2010);
- use fun and competitive elements and devise an incentive mechanism, such as in-app rewards, to keep users engaged (Coulton and Bamford, 2011; Ferreira et al., 2012; Henze and Pielot, 2013; Miluzzo et al., 2010). Moreover, users should be provided with methods to communicate, share opinions and see the performance of other users (Coulton and Bamford, 2011; Schleicher et al., 2011). Examples of desktop on-line apps that employed fun elements to attract users and collect a large amount of data are the Games With A Purpose that we survey in Chapter 3.

2

Design and Evaluation of Mobile Mindfulness Apps

In the first case study, we focus on the *Mental and Emotional Health* category (see Section 1.1.2). Among the possible topics in this category, we considered mindfulness, due to the positive effects it brings on people. Extensive reviews of the positive effects of mindfulness are provided in Brown et al. (2007), Chiesa and Serretti (2011) and Keng et al. (2011).

Originally associated to specific meditation techniques (Kabat-Zinn, 1990), a more recent definition considers mindfulness as a psychological process that consists of two components: *orientation to experience*, which refers to an orientation of curiosity, openness and acceptance toward one's experience, and *self-regulation of attention*, which refers to the non-elaborative awareness of mental events, i.e. thoughts, feelings and sensations, as they arise (Bishop et al., 2004). Some authors refer to self-regulation of attention as (i) *decentering*, defined by Safran and Segal (1996) as "the ability to observe one's thoughts and feelings as temporary, objective events in the mind, as opposed to reflections of the self that are necessarily true", or *detached mindfulness*, defined as "a state of awareness of internal events, without responding to them with sustained evaluation, attempts to control or suppress them, or respond to them behaviorally" (Wells, 2006). Decentering is considered particularly important in the literature because it can be helpful in reducing negative emotional states, such as worry and ruminative thinking, by helping individuals realize that their thoughts are impermanent events in the mind - see (Querstret and Cropley, 2013) for a review. Typically, decentering is achieved by practicing techniques that require individuals to be aware of their thoughts and to observe them while they pass by, without acting or grasping on them, and without trying to suppress them (hereinafter, we refer to this kind of practice as *distancing from thoughts*). Unfortunately, distancing from thoughts, as any other mindfulness practice can be difficult for people with no or minimal experience with meditation (in the following, *naive meditators*) (Kabat-Zinn, 2005; Segal et al., 2002). This can discourage them to start practicing or lead them to abandon the practice soon. There is thus the need to explore new and simpler ways to bring the benefits of mindfulness to people. This has captured the attention of the HCI community that has started proposing new, computer-based interactive approaches to help people practice it, e.g. (Thieme et al., 2013; Vidyarthi et al., 2012). However, such approaches rely on special hardware and settings, making them scarcely accessible to the general public and restricting the times and locations in which they can be used. These limitations could be overcome by mobile apps that run on common smartphone hardware without needing additional equipment. Currently, an increasing number of mobile mindfulness apps is available on online markets such as Apple's App Store or Google Play, see (Plaza et al., 2013) for a recent review on Android apps. However, none of these apps has been formally evaluated (Plaza et al., 2013) and thus it is not possible to understand whether they are effective in helping people achieve mindfulness, or could instead be ineffective or detrimental.

For this case study, we developed a mobile *enabling app* (see Section 1.1.1 for the PPM classification) that aims at supporting people *during* the practice of distancing from thoughts. Then, we carried out three different evaluations to assess its short- and long-term effects on people, as well as deeply investigate users' perception of it. The proposed app is called AEON (smArtphone basEd thOught distaNcing).

In this chapter, we first provide an introduction to the concept of mindfulness and survey the major modern approaches to mindfulness (Section 2.1). Then, we review related work on computer-supported mindfulness techniques. In Section 2.2, we illustrate the motivations and the design process that led us to the development of AEON, which is then described. Then, we motivate the three different evaluations we carried out: i.e. (i) a quantitative lab study (Chittaro and Vianello, 2014) (described in Section 2.3) , (ii) a qualitative in situ long-term study (Section 2.4), and (iii) a quantitative in situ long-term study based on research in the large (Section 2.5).

2.1 Related Work

The origins of the first mindfulness techniques can historically be traced back to Eastern philosophies. In particular, they are central to Buddhist traditions which attribute the first teachings of mindfulness to the Buddha himself (Gunaratana, 2002). Unlike Eastern traditions, Western conceptualizations of mindfulness are generally independent of any specific circumscribed philosophy, ethical code, or system of practices (Keng et al., 2011). In such conceptualizations, mindfulness was initially defined as a particular way of paying attention, a way of looking deeply into oneself in the spirit of self-inquiry and self-understanding (Kabat-Zinn, 1990). Although other definitions have been proposed over the years, the recent review on scientific studies of mindfulness by Keng et al. (2011) points out that most surveyed research follows the operational definition proposed by Bishop et al. (2004). This definition considers mindfulness as "a process of self-regulation of attention in order to bring a quality of non-elaborative awareness to current experience within an orientation of curiosity, experiential openness, and acceptance". Also, Bishop et al.'s definition supports the decoupling of mindfulness from meditation. Indeed, as pointed out by Hayes and Shenk (2004), if mindfulness is considered as a psychological mode or process, then any technique that is effective in producing that mode or process can be considered as a mindfulness technique. As a consequence, new mindfulness practices are emerging in addition or as an alternative to traditional meditation techniques.

In the following, we first summarize the most established mindfulness-based interventions followed in medical and psychological contexts. Then, we illustrate new, recently proposed computer-based approaches.

2.1.1 Mindfulness-Based Interventions

Although they often include techniques taken from Eastern traditions, modern Western approaches to mindfulness are clinically oriented and emphasize standardization and manualization to facilitate scientific study and empirical research (Chiesa and Malinowski, 2011). A large and growing number of studies assessed the effects of these approaches on different aspects of well-being (Keng et al., 2011). The first standardized approach to be introduced was the Mindfulness-Based Stress Reduction program (MBSR) (Kabat-Zinn, 1990) which was developed by Kabat-Zinn in 1979 and has been the subject of several studies among clinical and non-clinical populations (Keng et al., 2011). MBSR is a group-based intervention for populations with a wide range of stress-related disorders or chronic pain and is offered in hospitals and clinics around the world, as well as in schools, workplaces, corporate offices, law schools, adult and juvenile prisons, inner city health centers and a range of other settings (Kabat-Zinn, 2003). The program consists of an 8- to 10-week course in which a group of up to 30 participants meets weekly for 2- 2.5h sessions together with an all-day (7-8 h) intensive session usually held around the sixth week. During the sessions, participants receive instructions and training about three kinds of mindfulness techniques, which they have to practice also at home: *sitting meditation*, *body scan* and *mindfulness yoga* (Baer, 2003). Sitting meditation consists of different exercises, such as *mindful breathing* and *distancing from thoughts*. In mindful breathing, participants learn how to direct their attention to the sensations of breathing. They have to notice when their mind wanders away, observing it nonjudgmentally and bringing it back to breathing (Baer, 2003). In distancing from thoughts, participants shift their awareness to the process of thinking itself. They have to try to perceive thoughts as "events" in their minds. In particular, they have to note the thoughts' charge and possibly not be drawn into them, but just maintain the "frame" through which they are observing the process of thought. Participants have to be aware that each individual thought does not last long, i.e. it is impermanent, and that some thoughts keep coming back. Thus, participants have to act as a "non-judging observer" and to note how they feel about their thoughts (Kabat-Zinn, 1990). The body scan technique requires instead participants to sequentially direct their attention to the different parts of their body. They have to note the sensations arising from each part of the body and, as in the mindful breathing exercise, bring their attention back when their mind wanders away. Finally, mindfulness yoga consists of a series of postures to learn mindfulness of bodily sensations during gentle movements and stretching (Carmody and Baer, 2008). Participants are instructed to practice the above techniques at home for at least 45 minutes a day, six days per week, and are provided with CDs containing spoken instructions to be used as a support to mindfulness practice in the early weeks (Baer, 2003). Finally, participants are encouraged to engage in informal mindfulness practice by doing everyday activities (such as eating, walking, washing the dishes) with full awareness of the associated movements, sensations, cognitions and feelings that may be present. The goal of this out-of-class practice is to bring the capacity of mindfulness in everyday life, together with its associated benefits (Carmody and Baer, 2008).

The same techniques of MBSR are included in the Mindfulness-Based Cognitive Therapy program (MBCT) (Segal et al., 2002), which is an eight-week manualized intervention program developed to prevent depressive relapse in formerly depressed individuals. Unlike MBSR, MBCT incorporates also techniques and exercises derived from Cognitive Therapy that aim at helping participants view thoughts as mental events rather than as facts and thus change one's awareness of and relationship to thoughts and emotions (Keng et al., 2011; Teasdale et al., 2000).

While MBSR and MBCT are meditation-oriented approaches to mindfulness, Dialectical Behavioral Therapy (DBT) (Linehan, 1993a) and Acceptance and Commitment Therapy (ACT) (Hayes et al., 1999) do not involve formal meditation (Chiesa and Malinowski, 2011; Keng et al., 2011). DBT was first developed as an intervention for patients who meet criteria for borderline personality disorders and teaches mindfulness mainly in the context of group training as a way of helping patients increase self-acceptance and as an exposure strategy aiming to reduce avoidance of difficult emotion and fear responses (Linehan, 1993b). The exercises require visualizing thoughts, feelings and sensations as if they were clouds passing by in the sky (i.e., a distancing from thoughts technique), observing breath by counting or coordinating with footsteps, which is similar to mindful breathing, and bringing mindful awareness into daily life activities (Keng et al., 2011). ACT is aimed at fostering individuals' acceptance of unwanted thoughts and feelings, and to stimulate actions that contribute to an improvement in circumstances of living (Hayes, 2005). The program is based on the premise that psychological distress is often associated with attempts to control or avoid negative thoughts and emotions, which often paradoxically increase the frequency, intensity, or salience of these internal events, and result in further distress and inability to engage in behaviors that would lead to valued long-term goals (Keng et al., 2011). ACT consists of six core treatment processes: acceptance, defusion, contact with the present moment, self as context, values, and committed action (Hayes et al., 2006). Mindfulness is taught in the first four processes and consists of various exercises aimed at enhancing awareness of an observing self and foster the deliteralization of thoughts and beliefs, such as being in non-judgmental contact with environmental events as they occur (Keng et al., 2011). ACT has been delivered in both individual and group settings, with duration varying from 1 day to 16 weeks (Keng et al., 2011).

In addition to the above described mindfulness-based interventions, other techniques aimed at promoting mindfulness have begun to appear in non-clinical settings. In this chapter, we focus on computer-supported mindfulness, with an emphasis on approaches that require a user to *interact* with the computer application *while* practicing mindfulness techniques (in the following, *interactive practices*).

2.1.2 Studies of Computer-Supported Mindfulness

In most studies of mindfulness that involved computer support, mindfulness was a component of a broader therapeutic intervention delivered at distance. In particular, some studies investigated the use of web pages, e.g. (Boggs et al., 2014; Krusche et al., 2012, 2013; Thompson et al., 2010), web-enabled smartphones, (Kristjánsdóttir et al., 2011; Nes et al., 2012) or smartphone apps (Ahtinen et al., 2013; Lappalainen et al., 2014; Ly et al., 2014; Morris et al., 2010; Yang et al., 2014) for presenting mindfulness-related teaching materials. However, such courses do not offer any interactive practice to users (in the following, we refer to them as *non-interactive practices*).

Only the web-based course proposed by Glück and Maercker (2011) contained a small interactive practice. The web application showed participants a blue sky with a cloud that slowly wandered out of sight when they pressed the spacebar. Participants had to recognize any distressing thought, feeling or sensation that arose in their mind, label it non-judgementally (e.g. acknowledge that one feels angry by simply labeling the internal image with “anger”) and imagine placing it on the cloud, watching it wandering out of sight. However, participants found this interactive practice to be more difficult than the traditional non-interactive techniques that were taught in the intervention. This could be due to the fact that this interactive practice provided only a very primitive level of support, giving users the burden to carry out most of the assigned task mentally.

Other studies employed virtual environments (VEs) to support mindfulness. Baños et al. (2012) included a mindfulness technique in VEs for mood induction on elderly people aimed at increasing joy and relaxation. The mindfulness technique required participants to watch the VE, paying full attention to everything they heard or saw. Moreover, participants were told that if any thoughts or feelings arose during the practice, they should acknowledge and accept them, and that they should try to limit their attention only to the stimuli of the VE. Results of the study indicated significant increases in joy and relaxation and significant decreases in sadness and anxiety. Participants also reported low levels of difficulty of use and high levels of satisfaction. The employed VEs did not provide users with interactive practices: participants had to watch the VE without interacting with it during the practice.

Recently, the HCI community has started proposing computer-based approaches that include interactive mindfulness practices.

The Meditation Chamber (Shaw et al., 2007) and Sonic Cradle (Vidyarthi et al., 2012) propose interactive practices for mindful breathing. They both consist of immersive installations in which users can interact with their respiration (and also with other physiological parameters in the case of the Meditation Chamber) to control visual or audio content respectively. In this way, they offer users a tangible target to focus their attention on and invite its re-direction if it has driven away. The Meditation Chamber, which comprises also muscle relaxation techniques, was shown to be effective at promoting relaxation, see (Shaw et al., 2007), while the qualitative study described in (Vidyarthi and Riecke, 2014) pointed out that by using Sonic Cradle participants experienced some subjective elements typical of mindfulness meditation, such as reduced thought and clarity of mind. Moreover, participants described their experience with Sonic Cradle as relaxing and desirable, while experienced meditators suggested it was easier to engage with Sonic Cradle.

The Mindfulness Sphere (Thieme et al., 2013) relies on heartbeat rather than breathing perception as an object of user’s attention. The system is specifically aimed at introducing mindfulness in an intervention targeting women with a dual diagnosis of Learning Disability and Borderline Personality Disorder. It consists of a 12-cm diameter sphere that can sense the heartbeat of the user who touches it and translates it into visual and tactile feedback though multicolor LEDs or soft vibrations. However, this interactive practice was not formally evaluated and thus its effectiveness in promoting mindfulness remains unknown.

Finally, Yu et al. (2012) proposed two systems to support users in practicing *walking meditation*, i.e. a mindfulness practice that focuses attention on breathing combined with walking. More specifically, the technique requires users to slowly walk by lifting the foot with heel first while breathing in, and land the foot with toes first while breathing out. The first system (Walking-Aware System, WAS) aims at enhancing users’ awareness of walking and consists of a pair of shoes equipped with three force sensors, while the second system (Breathwalk-Aware System, BAS) aims at fully supporting walking meditation by introducing also respiratory sensors. For both systems, the interactive practice is supported by a mobile app that provides walking (WAS and BAS) and breathing (BAS) guidance and feedback. Results of two studies (Yu et al., 2012) showed the effectiveness of WAS and BAS in increasing user’s awareness of walking and support the practice of walking meditation, respectively.

Today, a number of mobile, smartphone-based applications aimed at supporting people in practicing mindfulness is available on on-line markets, such as Apple’s App Store or Google Play, see (Plaza et al., 2013) for a recent review on Android apps. However, none of these apps has been formally evaluated (Plaza et al., 2013) and thus it is not possible to understand whether they are effective in helping people achieve mindfulness, or could instead be ineffective or detrimental.

Recently, one of the publicly available apps, i.e. Mindfulness App (MindApps, 2014), has been used in a literature study cited above, i.e. (Ly et al., 2014). However, the study was not aimed at evaluating the effectiveness of the app. Instead, it focused on comparing two smartphone-delivered psychological treatments, one based on behavioral activation and other on mindfulness. In particular, the Mindfulness App was employed in the mindfulness treatment only to present audio guides to participants.

In the following section, we briefly survey the mobile mindfulness apps that are currently available on the app stores.

2.1.3 Smartphone-Supported Mindfulness

A high number of mobile apps related to mindfulness is available on on-line stores such as Apple's App Store and Google Play. We first focus our attention on the apps that do not target distancing from thoughts. Such apps can be divided in two groups, i.e. those that do not include interactive practices (in the following, *non-interactive apps*) and those that include interactive practices (in the following, *interactive apps*).

Some of the non-interactive apps offer only timers and sounds to time user's meditation practices, e.g. Lotus Bud Mindfulness Bell (Sager, 2014) and ZenFriend (Small-n-Tall, 2014). Other apps feature instead audio guides for supporting several mindfulness techniques, such as mindful breathing or mindful walking. Such guides are:

- adapted from the MBSR program or led by mindfulness teachers, e.g. in Mindfulness App (MindApps, 2014) or Headspace (HeadSpace, 2014);
- targeted to the mental, physical and emotional status that the user enters, e.g. in Stop, Breathe & Think (Tools for Peace, 2014);
- specific for a particular context or situation that the user can choose, e.g. if (s)he is walking or exercising as in Buddhify2 (21awake, 2014);
- adapted for a particular health condition or behavior, e.g. losing weight as in Lose Weight (Rodriguez Fornies, 2014), or quit smoking as in Craving to Quit (Goblue International, 2014).

A few non-interactive apps offer video animations for helping users perform the mindful breathing exercise, e.g. Universal Breathing: Pranayama (Saagara, 2014). Finally, several of the above mentioned apps allow also users to set reminders to meditate, keep track of their mood and sessions as well as share them with other users.

The three available interactive apps in this category are all focused on the mindful breathing exercise. Such apps are:

- Mindfulness TS (Mindfulapps, 2012) that exploits the touchscreen to offer the interactive practice. Indeed, it requires users to focus their attention on their breath and tap the screen each time they breath out. The user can specify the length of the session, and at the end the application provides him/her with summary statistics about his/her level of attention, based on the number of times the user tapped the screen. In a similar exercise, the user has to think of a word or a short sentence related to a positive feeling and has to tap the screen each time (s)he repeats it.
- BreathTheWaves Full (BreathTheWaves, 2013) that exploits instead the smartphone's motion sensors. For this reason, the app must be used when the user is lying down with the smartphone placed on his/her abdomen. In the interactive practice, after several breaths, the app recognizes the user's breathing pattern and starts to create synchronized sounds to help him/her remain focused on it. Users can set the length of the sessions and choose among different soundscapes, e.g. rolling wave or whale sounds. Finally, they can visualize statistics on their respiration rate.
- Muse Calm (InteraXon, 2014a) that requires an external brain-sensing headband, i.e. Muse (InteraXon, 2014b) to offer its functionalities. For the interactive practice, users must wear on the headband and connect it to the smartphone through the Bluetooth. Then, they have to close their eyes and count their breaths. The headband tracks the user brain activity and is able to detect if the user is focused or distracted. If it detects that the user's brain if focused, the app plays a gentle sound. Otherwise, if the

headband detects that the user's mind has wandered away, the app plays sound of winds and crashing waves. This real-time feedback is aimed at helping the user identify the distraction and bring back his/her attention to the breath. Also this app lets users specific the length of the meditation sessions and offer statistics of their performance.

Overall, to offer an interactive practice, the above described apps are bound to physical cues, e.g. the falling and rising of the belly, or on a fixed object of attention, e.g. the counting of the breaths. However, such features are not suitable for supporting a mental practice as distancing from thought that requires users to focus on intangible and impermanent objects, i.e. the thoughts. Finally, Muse Calm requires an ad-hoc hardware for the interactive practice, i.e. the brain-sensing headband, which can be scarcely accessible to the general population.

Considering the apps that focus on distancing from thoughts, one of them, i.e. Let Go (Araaya, 2012), aims at helping users become aware of their negative feelings and the reactions they usually adopt for them. To this purpose, the app offers a (customizable) list of predefined negative feelings, such as sadness and anxiety, and when the user experience one of them, (s)he has to tap on the corresponding label. Then, the app plays a piano melody and shows images of nature and quotes to help the user let go of the experienced negative feeling. When the user taps again on the screen, the app shows a list of common reactions, such as "letting go" or "keep a lot" and the user can select the one (s)he has adopted. Finally, the app offers users statistics on the frequency of feelings they experienced and the reactions they chose.

Other apps aim instead at helping users let go of their thoughts or worries through some form of interactivity. However, two of these apps, i.e. Just Let Go (Wolffram and TrueSelfSoft, 2011) and ThoughtBox (Hartnett, 2013), offer a limited support, requiring instead the user to do most of the task mentally. Indeed, Just Let Go shows a mandala, i.e. a geometric figure representing the universe in Hindu and Buddhist symbolism. The user has to think of a thought (s)he wants to let go and press the center of the mandala while breathing out for five seconds. Then, (s)he has to release the mandala and breathe in. ThoughtBox lets instead users enter their thoughts or worries. However, to let a thought go, users have to tap on it and take a deep breath, as in Just Let Go.

Five other apps offer instead some animations for helping people let go of their thoughts or worries. Such apps are:

- Throw Your Worry Away! (Keru, 2011) that lets user enter a text, which is supposed to describe a worry, and associate it to a rocket. When the user touches the screen, the rocket is fired into the sky, and is shown leaving Earth and going off-screen.
- Let It Go - stress relief (LIGSR) (Inside Outlier, 2014) in which users can enter a thought or a worry on a label that is attached to a balloon. Then, they can release the balloon and see it fly away.
- Worry Bubble (Embodied, 2014) that lets users choose among seven worry domains, such as health and money, and enter a related worry. Then, the app shows a bubble that contains an icon of the chosen worry domain and a white dot that is supposed to be the worry entered by the user. The user can touch such bubble to make it burst. In a subsequent screen, another animation shows the icon of the chosen worry domain fly away.
- Worrydoll Lite (Dontworrycompany, 2011) in which users can enter and assign a text (which is supposed to be a worry statement) to each of the four dolls displayed in the garden of a house. Once tapped, each doll will go inside the house and start to worry for the user about the statement assigned to it. Users can tap on the doll to see again their entered thought and know the amount of time the doll has been worrying about it. This process lasts until the app is closed.
- Let It Go: Vent and Release (LIGVR) (Nature Applications, 2014) that allows users to enter a thought or a worry on a list. The user can then choose to let it go and the app shows an animation of a beam of light that goes up and down the screen. The length of such animation is randomly chosen by the app and can last between 1 and 72 hours.

Overall, although the above described apps can offer users a metaphor for helping them let go of their thoughts or worries, none of them lets users see their entered thoughts or worries while they are going away.

The only app that offers users such visualization is The Shredder (Bowers, 2011). Indeed, it proposes to use a virtual shredder as a fun way to reduce negative thinking, stress and anxiety. The user can enter texts describing unpleasant thoughts, feelings or situations and drag them inside the virtual shredder to see them destroyed. However, the user does not have control on the graphic process.

In conclusion, none of the available mindfulness app implements a metaphor for distancing from thoughts that offer users both (i) the visualization of the entered thoughts while they are going away, and (ii) the control of such process.

It must be noted that we started our survey of the available mobile mindfulness apps in the first half of 2012. For this reason, some of the described apps are not available in stores anymore, i.e. Throw Your Worry Away! and Worrydoll Lite.

2.2 The Proposed Application

In this section, we describe the motivations and the design process that led to the development of AEON. Then, we illustrate the app in detail.

2.2.1 Design Process and Motivations

To design our interactive practice, we first studied the difficulties that naive meditators could encounter when approaching mindfulness. To this purpose, in addition to a thorough analysis of the literature, we enrolled in an actual mindfulness program to experience ourselves possible difficulties and to hear opinions from other participants in the program. Considering the approaches described in Section 2.1.1, we first looked for courses based on MBSR or ACT, because they do not target a specific set of individuals, while MBCT is specifically addressed to people who suffer repeated bouts of depression and the aim of DBT is to treat people with borderline personality disorders. Then, we chose MBSR because of its larger diffusion and more generic scope than ACT. The mindfulness program we followed was held at our local hospital and led by an expert neurologist and meditator. The program lasted 8 weeks with a weekly 1.5h session and the participants were naive meditators. It included all MBSR techniques with the only exception of yoga. At the end of each meeting, ample discussion time was devoted to allowing participants to share their experience and the difficulties they encountered during the current and the homework mindfulness sessions. Each author of this thesis participated to a different edition of the course. The organization of the course and the instructor did not change, but the two groups of participants were different. In this way, we had the opportunity to hear discussions involving a larger number of participants. Moreover, by not being together, there were no opportunities to influence each other. We independently collected our observations on paper and integrated them only at the end of the two courses.

In both courses, we observed that participants reported problems with practicing all the techniques they were learning, but distancing from thoughts appeared more difficult than mindful breathing and body scan. In particular, some participants explicitly reported that it was difficult to mentally visualize their thoughts and feelings coming and going away, while for others it was difficult to be aware of their thoughts without being drawn into them.

These observed difficulties could be explained by the fact that distancing from thoughts requires to perform a more abstract task than the other two considered techniques. Mindful breathing and body scan are grounded in bodily activities and physical sensations which could provide users with tangible targets to direct their attention on. Indeed, in mindful breathing, the user can concentrate on interoceptive cues, i.e. the feeling that the breath creates when passing through the nostrils or the physical rising and falling of the belly with the in-breaths and the out-breaths. Body scan relies on a rich set of somatosensory cues, combining tactile and thermal perception, nociception (i.e. perception of pain) in the different parts of the body, and proprioception (i.e. the ability to sense the position of the body in physical space). Unfortunately, distancing from thoughts cannot rely on any of the above summarized physical cues and requires mental effort to be able to visualize and observe one's thoughts.

The difficulty of distancing from thoughts is also recognized in the literature. For example, Kabat-Zinn (1990) acknowledges that the exercise requires great concentration and suggests that it should be done for short periods of time in the early stages of mindfulness practice. Difficulties in trying distancing from thoughts can foster a sense of failure, which is common in naive meditators (Kabat-Zinn, 2005) and could discourage them to continue the practice. For these reasons, we chose to focus our work on distancing from thoughts and to develop an interactive practice to support it.

We examined the interactive practices proposed in the literature that were available when we started out research work on mindfulness (see Section 2.1.2) to define our initial design choices. The choice of developing for a mobile platform was motivated by the following considerations about the literature. Interactive practices such as (Glück and Maercker, 2011) that rely on a desktop computer with keyboard and Internet connection limit the possibility of practicing to those situations in which the user is in front of such a system. Applications which require more sophisticated software and hardware, such as touchscreens to interact with a VE (Baños et al., 2012) or ad-hoc systems such as the Meditation Chamber (Shaw et al., 2007) make it even more difficult for users to access the interactive practices. Mobile phones are instead always available and can make it much easier for a user to initiate an interaction. Therefore, we chose to focus on a mobile app to increase the number of opportunities to practice and the contexts in which it is possible to be supported by the application in practicing mindfulness. This is especially important for naive meditators, because they are not used to practicing regularly. Restricting the places and times in which a naive meditator can practice can be detrimental to the goal of making his/her practice regular.

In designing our specific mobile interactive practice, we first examined in detail the available mobile distancing from thoughts apps (described in Section 2.1.3), deriving the following considerations. Just Let Go uses a mandala as a metaphor for distancing from thoughts and requires a user to think and keep in mind the thought (s)he wants to let go. Considering the difficulties naive meditators often experience in mentally visualizing their thoughts, we decided that the interface of our app should externalize thoughts to make it easier to practice distancing.

Throw Your Worry Away, WorryDoll Lite and The Shredder allow a user to enter his/her thoughts or worries in the application. However, the first two apps do not provide any means of visualizing the entered thought while it is going away. In WorryDoll Lite, users see an animation of a doll which is supposed to worry for them. They can only tap on the doll to see again their entered thought and know the amount of time the doll has been worrying about it. Moreover, this process lasts for a long amount of time, i.e. until the app is closed, and this could not give users the awareness of the impermanence of each thought. In Throw Your Worry Away, after entering a thought, a user cannot see it on the screen, which only shows a rocket leaving Earth and going off-screen. Moreover, the funny style of the graphics and the animation can distract the user from his/her entered thought. For our app, in addition to showing the thought so that a user can focus on it, we decided to visualize the process of progressive disappearance of the thought.

The Shredder allows a user to see his/her entered thought while it is destroyed by a virtual shredder. Even if this app implements a metaphor which comprises both the thought and its disappearance, it does not give the user control on the graphic process that makes the thought progressively disappear. Moreover, the shredding sound might be perceived as unpleasant. Therefore, we decided that our app should give users control on the disappearing of the entered thought and, at the same time, foster a pleasant experience.

To choose a metaphor that could foster a pleasant experience in which to remain focused on the thought and its disappearance, we looked at natural elements. We were initially inspired by Linehan (1993b), who pointed out that an important consequence of mindfulness practice is the realization that most sensations, thoughts, and emotions fluctuate, or are transient, passing by "like waves in the sea" (cited in (Baer, 2003)). Elaborating on the idea of waves, we then associated them to their ability of dissolving things as they pass over them, as for example on the seashore. In our interactive practice, each thought is visualized as written over a parchment placed under water and, as the waves pass over the thought, they progressively dissolve the ink. Users control the process, by physically playing with water and causing the formation of waves. In this way, the user can focus on each thought, observing it while (s)he makes it disappear at his/her own pace. Moreover, the propagation of waves might metaphorically represent the detachment of thoughts from their associated emotional charge, as the latter is symbolically carried away with them.

2.2.2 The AEON App

The proposed application is organized in two screens. The “My Thoughts” screen (Figure 2.1) shows the list of thoughts entered by the user and two status bars. The status bar at the top contains the buttons to enter or delete thoughts from the list.

Figure 2.1: The “My Thoughts” screen.

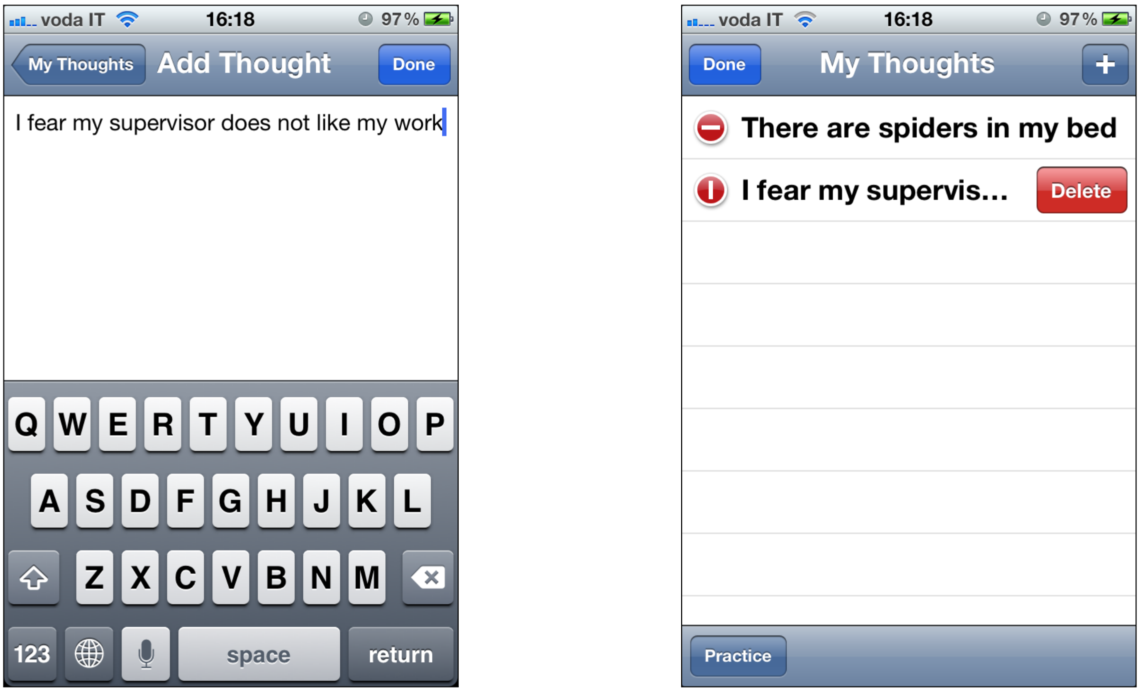
A thought can be entered by pressing the “+” button and writing in the text-area that appears (Figure 2.2a), while it can be deleted from the list by tapping the “Edit” button and selecting the corresponding row (Figure 2.2b).

The status bar at the bottom of the “My Thoughts” screen contains the “Practice” button, which starts the distancing from thoughts practice. Before practice, the user has to select the thoughts (s)he wants to distance herself/himself from, by touching the corresponding rows. A check is shown on each selected thought (Figure 2.3).

When the user presses the “Practice” button, the app switches to the “Practice” screen, which displays a parchment under water, with the first selected thought written in ink on the parchment (Figure 2.4).

The user can then interact with the water by touching any point on the screen and moving his/her fingers anywhere over the screen. User’s actions trigger waves in the water that propagate over the entire parchment. The simulated water behaves consistently with user’s experience of the natural element: the wave triggered by a tap is circular (Figure 2.5) while waves triggered by moving the finger on the screen are more chaotic (Figure 2.5). The user has control on the timing and speed of the process: (s)he chooses when and where to trigger waves and how strong the waves are (dissolving more or less ink).

After the user has distanced himself/herself from a thought by making it completely disappear, the app allows him/her to move to the next thought by swiping with two fingers from the right to the left border of the screen. When the last of the selected thoughts has disappeared, the swipe action makes the app return to the “My Thoughts” screen.



(a) (b)

Figure 2.2: Entering a thought (a) and deleting a thought (b).

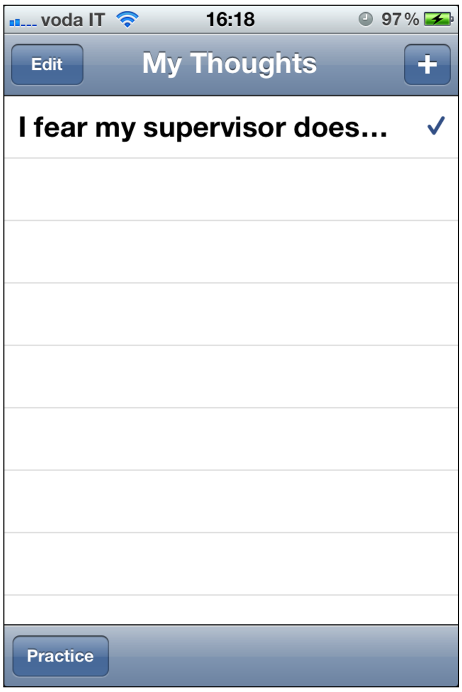


Figure 2.3: Selecting thoughts.

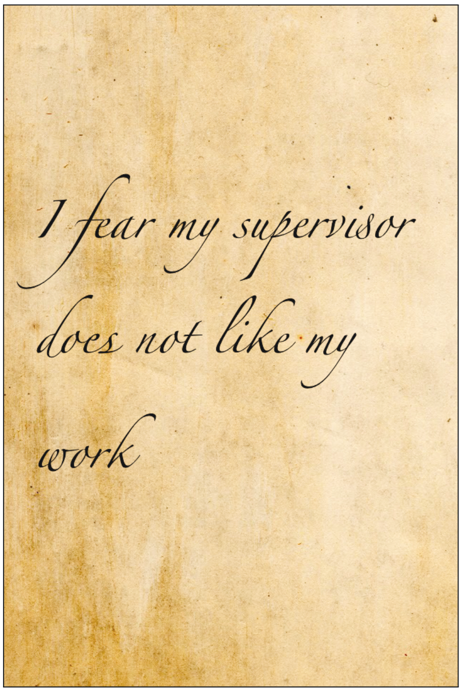
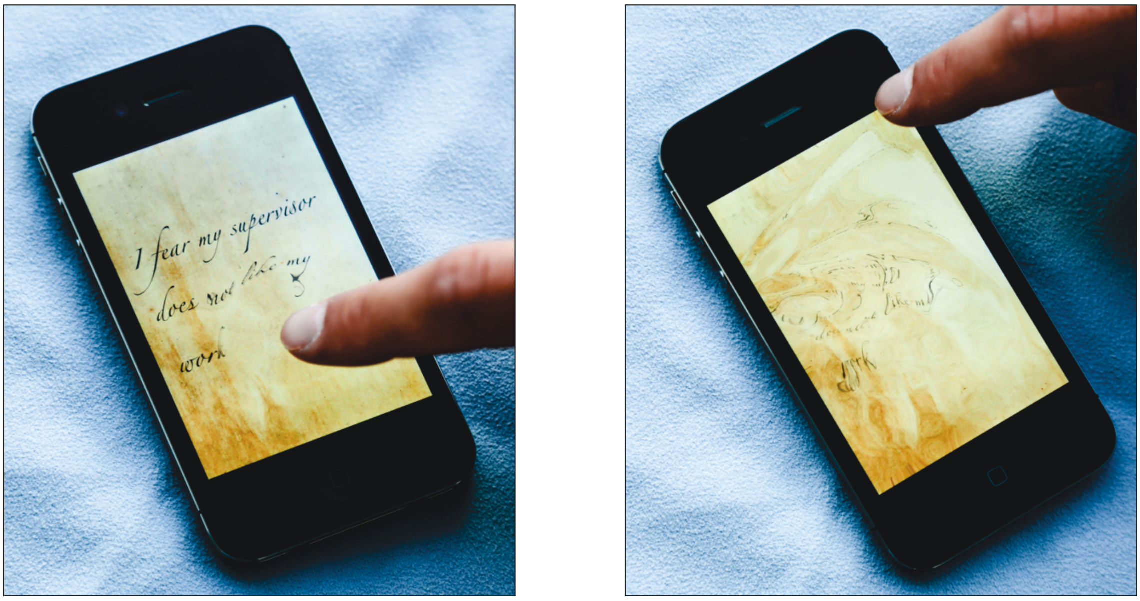


Figure 2.4: “Practice” screen.



(a) (b)

Figure 2.5: Triggering a circular wave (a) or more chaotic waves (b) in the “Practice” screen.

AEON can provide users with new opportunities for approaching and learning mindfulness. Indeed, as mobile devices are always at one's disposal, they can make it much easier for users to engage in practice sessions. Moreover, in contrast with abstract distancing from thoughts techniques, an interactive practice can provide a visual representation of thoughts, making the practice of distancing from thoughts easier for users. Finally, by employing graphics and animation, these practices could offer an engaging and enjoyable experience to users, which could persuade them to practice mindfulness more frequently.

However, we cannot exclude that our interactive practice could introduce possible elements of distraction, because a user could be more focused on the physical interaction with the device or on the water simulation rather than on his/her entered thoughts. We also cannot know a priori if the interactive practice is actually effective in supporting mindfulness and if it could possibly be more effective than traditional practices.

To evaluate all the above mentioned aspects, we carried out three different studies of AEON. First, given the novelty of our proposed approach, we wanted to assess if it could actually support naive meditators in practicing mindfulness and achieving a mindfulness state. Thus, we contrasted it with two traditional mindfulness techniques not based on technology in a controlled lab setting. Such settings have the advantage of more easily isolating mindfulness practice from other elements typically present in clinical intervention programs, thus allowing greater control over independent variables, scientifically measure the outcomes produce by the practice and draw stronger conclusions about causal effects (Keng et al., 2011).

Second, since AEON is intended to be used by people in their everyday contexts, we carried out an in situ long-term qualitative study. Indeed, as pointed out by Klasnja et al. (2011), a critical contribution to the evaluation of technologies for health and well-being is to deeply understand how a proposed system is used by its target audience in situ and this understanding, together with the design knowledge that results from it, is arguably the biggest contribution that HCI can make to the development of effective systems. To this purpose, we carried out a 5-week study on a sample of naive mediators who were asked to practice with AEON at least once a day. Then, we interviewed them at the end of the study period.

Third, we wanted to deepen the investigation of the long-term effects of AEON outside the lab and, at the same time, extend the sample of our study. To this purpose, we employed a quantitative approach, by including a mindfulness questionnaire into the app, and the research in the large methodology (see Section 1.2) to carry out a 4-week study. Participants could participate in the study by downloading the app from Apple's App Store and Google Play and answer the mindfulness questionnaire after the installation and the acceptance of the study condition, two weeks after and four weeks after the acceptance of the conditions.

In the following, we describe the three evaluations of AEON that we carried out, i.e. (i) the lab quantitative study (Chittaro and Vianello, 2014) (Section 2.3), (ii) the in situ long-term qualitative study (Section 2.4), and (iii) the in situ long-term quantitative study by employing the research in the large paradigm (Section 2.5).

2.3 Quantitative Lab Study

To evaluate the effectiveness of AEON in helping naive meditators achieve mindfulness, we contrasted it in a lab setting with two traditional distancing from thoughts techniques:

1. *cloud imagery* (CLOUD, for short), i.e. a mental imagery task in which people imagine their thoughts as written on clouds floating in the sky, allowing them to occupy their own space and observe them as they pass by (Wells, 2006);

2. *card-tossing* (CARD, for short), i.e. a task in which people pick up cards (with their thoughts written on) one at a time, look at the thought on the card and then toss the card into a wastepaper basket (Hayes et al., 1994; Leahy, 2006).

We chose to include the CLOUD technique in the study because it is a well-known traditional approach for practicing distancing from thoughts. However, since naive meditators could find a mental technique difficult to start with, we included in the study also a technique that is based on manual interaction and offers thought externalization, i.e. CARD.

The three techniques (AEON, CLOUD and CARD) were compared in a controlled laboratory setting. This is consistent with a recent line of scientific research on mindfulness, which has examined the immediate effects of brief mindfulness interventions on a variety of emotion-related processes - see (Keng et al., 2011). In these studies, participants are typically asked to practice one or more mindfulness techniques. Then, after each practice, they have to answer one or more questionnaires which assess the possible outcomes produced by the considered techniques. For example, the studies described in (Feldman et al., 2010; Lau et al., 2006) employed the mindful breathing technique, while Erisman and Roemer (2010) used an additional exercise (mindfulness of emotions) which asked participants to be aware and let go of their emotions while the experimenter read them a poem. In all of these studies, the level of decentering achieved by participants was one of the assessed outcomes. Other studies employed distancing from thoughts, e.g. (Singer and Dobson, 2007; Perlman et al., 2010), but without measuring decentering.

Considering that this study focuses on distancing from thoughts techniques, we considered the achieved level of decentering as an important outcome to assess. As previously explained, the pleasantness and the difficulty of the technique can influence the motivation of a naive meditator, so we included also their assessment in the study.

2.3.1 Hypotheses

As already mentioned in Section 2.2.1, naive meditators tend to experience difficulties with traditional mindfulness techniques. We hypothesize that the proposed interactive practice better supports naive meditators with respect to traditional techniques, resulting in measurable higher levels of achieved decentering. Moreover, we predict the interactive practice to be perceived as more pleasant and less difficult than the traditional ones. As a result, we also expect the interactive practice to be preferred by participants.

2.3.2 Participants

Participants were recruited among graduate and undergraduate students through direct contact, asking them if they were willing to participate in a study of three different techniques aimed at distancing from worries. To identify naive meditators, first we provided candidates with a definition of meditation as in the study by (Feldman et al., 2010). Then, we employed a questionnaire made of three items. The first item was the question used in (Feldman et al., 2010), which asked participants about their meditation frequency. The available responses were “I meditate at least once a day”, “I meditate at least once a week”, “I meditate once per month”, “I do not meditate regularly”. Participants who chose one of the first three answers were also asked the second and third questions which required them to specify when and how long they meditate to define in this way a precise period of time regularly devoted to meditation. Following Thompson and Waltz (2007), determining this period of time for participants who claim to meditate regularly is important to distinguish participants with formal meditation practice.

Candidates who chose “I do not meditate regularly” (20 participants) or chose other responses but then indicated very short periods of time devoted to meditation (2 participants) were considered as naive meditators, and thus included in the study.

A total of 32 candidates were recruited, and the 22 (10 M, 12 F) of them who met the above described criteria to participate in the study formed our sample. The age of participants ranged from 19 to 28 ($M = 23.95, SD = 2.15$). All of them were right-handed. On a self-report scale ranging from 1 (low familiarity) to 7 (high familiarity), participants were very familiar with mobile devices ($M = 6.36, SD = .73$) as well as mobile touchscreen devices ($M = 5.95, SD = 1.13$).

2.3.3 Measures

2.3.3.1 Decentering

The participants’ achieved level of decentering was assessed with the 7-item *Decentering* subscale of the textitToronto Mindfulness Scale (TMS) (Lau et al., 2006).

The subscale asks participants to express how well what they experienced is described by items such as “I experienced myself as separate from my changing thoughts and feelings” or “I approached each experience by trying to accept it, no matter whether it was pleasant or unpleasant". Items are rated on a 5-point Likert scale (0="not at all", 4="very much"). Scores on the subscale are summed and the total score ranges from 0 to 28. The subscale was translated into Italian and its internal reliability was measured with Cronbach's alpha, $\alpha = .78$ (CLOUD), .79 (CARD), .72(AEON).

2.3.3.2 Pleasantness

The degree of pleasantness of each technique was measured with the Pleasure dimension of the *Self-Assessment Manikin* (SAM) (Bradley and Lang, 1994), which employs 5 graphic depictions that range from a smiling, happy figure to a frowning, unhappy figure. Pleasantness is rated on a 9-level Likert scale, composed by the 5 depictions and the 4 spaces between them, where the happy figure corresponds to 9 and the unhappy figure to 1.

2.3.3.3 Difficulty

The level of difficulty of each technique was assessed by a three-item questionnaire ("I found it difficult to practice this technique", "I found it demanding to practice this technique", "I found it complicated to practice this technique"), rated on a 7-point Likert scale (1="strongly disagree", 7="strongly agree"). To obtain a composite measure, the sum of the three items was averaged for a single mean score, $\alpha = .90$ (CLOUD), .80 (CARD), .70 (AEON).

2.3.3.4 Preference

Preference was assessed by a question that asked participants to indicate which was the technique they preferred to practice.

2.3.4 Materials and Apparatus

To allow participants to practice the CARD technique, decks of 21×10cm cards were prepared. Each deck consisted of three numbered white cards interleaved by two card-shaped sheets of carbon paper. Each deck was held together by two removable clips. Each participant received a deck in which the cards were numbered with a "1", a deck numbered with a "2" and a deck numbered with a "3". Usage of these materials is described in Section 2.3.5

The AEON app was run on an Apple iPhone 4S equipped with a 3.5", 960×640 pixel touch screen. During usage, the device was in portrait mode and placed over a mat to avoid sliding. Participants interacted with the device by using the fingers of their dominant hand. During the evaluation, participants were seated in a 44cm-high chair in front of a 72cm-high table.

2.3.5 Procedure

The study was based on a within-subjects design with distancing from thoughts technique (CLOUD, CARD, and AEON) as independent variable. The order of presentation of the experimental conditions was counterbalanced to prevent learning effects.

Participants were individually taken to a quiet room and briefed about the nature of the experiment. Afterwards, they were instructed to think of three worries they had been having in that period of their life, without disclosing them to the experimenter. Participants were then provided with the three decks of cards and asked to write the first worry on the upper card of deck numbered "1", the second worry on the upper card of deck numbered "2" and the third worry on the upper card of deck numbered "3". Thanks to the carbon paper in the decks, this produced three written cards for each worry. Participants had then to remove the clips on each deck and organize the cards in three new decks, each one made by three cards with the first worry on top followed by the second and the third worry beneath. All the written worries in each deck faced downward and the three identical decks were placed on the table separately. To let participants freely express any kind of personal worry, they were previously informed that the experimenter would have been seated in a position from which it was impossible to read the worries on the cards and they could take away all the cards with them at the end of the experiment. After the preparation of the decks was completed, the experimenter showed participants the AEON app and explained to them how to enter worries and delete them. They were then asked to enter their three worries following the same order followed with the cards. Also in this case, the experimenter was unable to see the worries they entered and participants were informed that at the end of the experiment they could delete the worries from the app. These preparation activities were carried out before the execution of the experimental tasks so that at the beginning of each condition participants had all the necessary materials ready to practice distancing from thoughts. Before each condition, the experimenter explained in detail the technique to practice and was available to clarify possible doubts.

Participants were asked to practice the technique on the three worries for three times: following in each repetition the same order they decided for the worries at the beginning of the experimental procedure. As a result, in each condition they practiced the distancing from thoughts technique three times on each worry. Since in the CLOUD condition the experimenter had no way of observing if participants possibly skipped a repetition of the three worries, they were asked to say the number of each completed repetition at the end of it. To avoid introducing a confounding factor in the experiment, they were asked to do the same also in the other two conditions. For the CARD condition, in each repetition, participants used one of the deck of cards they had previously organized, picked up one card at a time from it, looked at the worry on the card and then tossed it into a wastepaper basket.

At the end of each session, participants were asked to fill out on a computer the questionnaires for measuring decentering, pleasantness and difficulty. At the end of the last condition, participants were also asked to indicate their preferred technique. We collected questionnaire data with the computer to avoid possible transcription errors. Moreover, to ensure participants' privacy, at the beginning of the procedure each participant picked up a printed random-generated code from a box and entered it into the computer as a unique identifier. In this way, the collected data was stored in anonymized form.

Finally, participants were briefly interviewed to possibly get comments on the techniques. More specifically, they were asked to freely express any difficulty or impression concerning each technique. The relevant comments are illustrated in the Discussion section. After thanking participants for their participation, to reassure them that the worries they wrote remained private, they were invited to take with them the cards they had previously tossed in the basket and to delete their worries from the application. Overall, carrying out the procedure took about 45 minutes per participant.

2.3.6 Results

2.3.6.1 Decentering

Figure 2.6 shows the mean achieved level of decentering for the three conditions. The data was subjected to a Shapiro-Wilk test of normality, which revealed no significant deviation from the normal distribution. A one-way ANOVA was then carried out, which pointed out a significant effect ($F(2,42) = 3.50, p < .05, \eta^2 = .14$). The effect was then investigated by carrying out a t-test pairwise comparison with Bonferroni correction. The post-hoc analysis revealed a significant difference ($p < .05$) between CARD ($M = 13.64, SD = 5.44$) and AEON ($M = 17.18, SD = 4.99$), with participants achieving a higher level of decentering with AEON. The average decentering for CLOUD ($M = 15.55, SD = 5.25$) was in between CARD and AEON.

2.3.6.2 Pleasantness

Figure 2.7 shows the mean degree of pleasantness for the three conditions. The scores were analyzed with Friedman's test, which pointed out a significant effect, $\chi^2(2, N = 22) = 15.90, p < .001, W = .36$. We then employed a Wilcoxon signed-rank test pairwise comparison with Bonferroni correction to investigate the effect. The analysis pointed out a significant difference ($p < .01$) between AEON ($M = 6.69, SD = 1.64$) and CLOUD ($M = 5.22, SD = 1.77$) and a significant difference ($p < .01$) between AEON and CARD ($M = 5.34, SD = 2.12$), with participants perceiving AEON as more pleasant to practice than the other two techniques.

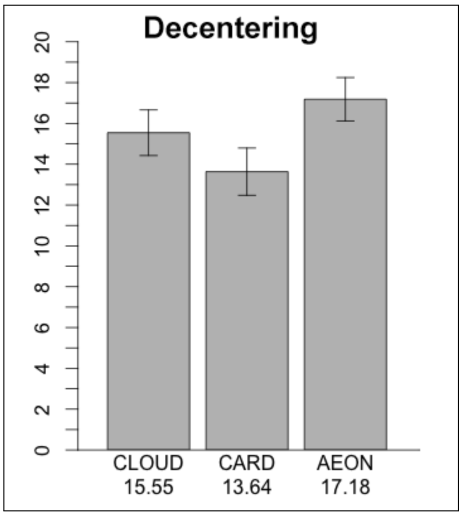


Figure 2.6: Mean achieved level of decentering (capped bars indicate $\pm 1SE$).

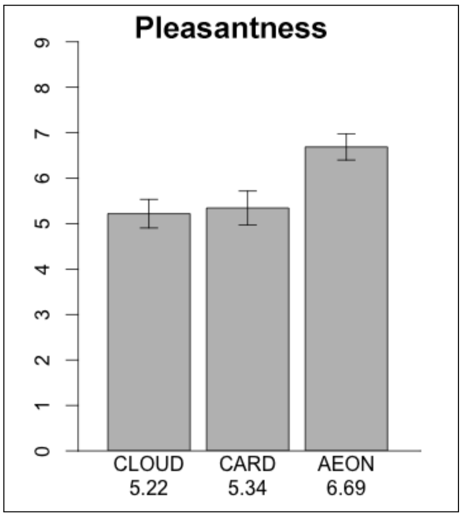


Figure 2.7: Mean degree of pleasantness (capped bars indicate $\pm 1SE$).

2.3.6.3 Difficulty

Figure 2.8 shows the mean level of difficulty for the three conditions. The data was subjected to a Shapiro-Wilk test of normality, which revealed a significant deviation from the normal distribution. Thus, we employed Friedman's test, which pointed out a significant effect, $\chi^2(2, N = 22) = 34.82, p < .001, W = .79$. We then performed a Wilcoxon signed-rank test pairwise comparison with Bonferroni correction. The post-hoc analysis revealed a significant difference ($p < .001$) between CLOUD ($M = 3.64, SD = 1.84$) and CARD ($M = 1.33, SD = .50$) and a significant difference ($p < .001$) between CLOUD and AEON ($M = 1.15, SD = .32$), with participants perceiving AEON and CARD as less difficult to practice than CLOUD.

2.3.6.4 Subjective Preference

A Chi-Square test was performed on subjective preference data (frequencies are shown in Figure 2.9). The analysis revealed a significant effect ($\chi^2(2, N = 22) = 9.90, p < .01, w = .67$), and AEON was the preferred approach.

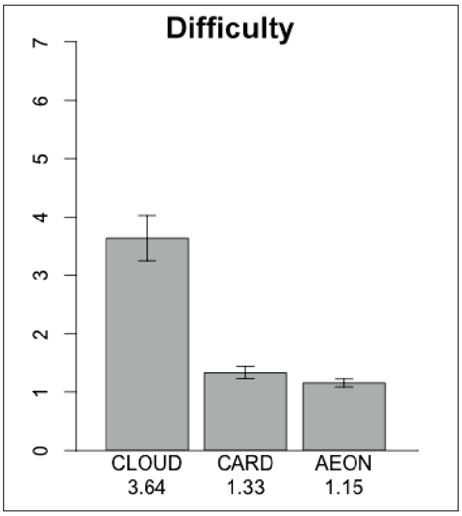


Figure 2.8: Mean level of difficulty (capped bars indicate $\pm 1SE$).

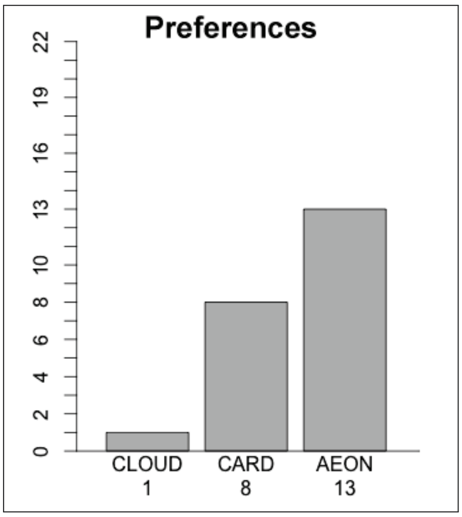


Figure 2.9: Preference frequencies.

2.3.7 Discussion

2.3.7.1 Difficulty

AEON received the best ratings in terms of difficulty and the post-hoc analysis showed that AEON was perceived as significantly less difficult to practice than CLOUD. A possible explanation of this result is that, compared to CLOUD, AEON could provide a higher degree of *computational off-loading*. This term refers to "the extent to which different external representations reduce the amount of cognitive effort required to solve informationally equivalent problems" (Rogers, 2004) and is part of the theory of external cognition (Scaife and Rogers, 1996). In particular, the better performance of AEON could be especially related to the *temporal and spatial constraining* dimension of computational off-loading, which regards the way different representations can make relevant aspects of processes and events more salient when distributed over time and space (Rogers, 2004). Indeed, AEON allows a user to see his/her worries, interact with them and visualize them as they progressively disappear. On the contrary, CLOUD requires a user to mentally visualize his/her worries as clouds passing by. This consideration is reflected in most participants' comments, which pointed out that it was hard to visualize the clouds. Participants were also very familiar with mobile touchscreens which could have contributed to further make them at ease with AEON.

This result looks apparently in contrast with (Glück and Maercker, 2011), in which participants found the interactive practice to be difficult. However, in that study, the interactive practice required participants to mentally label their thoughts and then imagine the labels as if they were written on the clouds on the screen, and thus provided less support than AEON to participants.

The post-hoc analysis pointed out that also CARD was perceived as less difficult to practice than CLOUD. Again, this result could be explained by a higher degree of computational off-loading that CARD could provide with respect to CLOUD. Indeed, CARD lets a user see his/her worries externalized on the paper cards.

Although the difference between AEON and CARD was in the hypothesized direction, it was very small and the post-hoc did not reach statistical significance. This could be due to the fact that, as mentioned above, the two techniques were both able to provide a good degree of computational off-loading. In particular, considering the *cognitive tracing* design concept of computational off-loading, i.e. "the way users develop their own understandings and external memories of a representation of a topic by being allowed to modify and annotate it" (Rogers, 2004), both AEON and CARD provided users with a concrete way to externally manipulate their worries, by means of different types of manual interaction. This is reflected in the comments of some participants, for example one emphasized that he felt to have a more direct contact with his worries with AEON and CARD rather than with CLOUD.

2.3.7.2 Decentering

Participants achieved the highest level of decentering while using AEON and the post-hoc analysis showed that AEON was significantly more effective than CARD. In mindfulness practices, the ability to focus attention without being distracted, i.e. to maintain a sustained attention, is one of the necessary components to gain awareness on the current experience (Bishop et al., 2004; Shapiro et al., 2006). Through multimodal interaction, AEON could have involved participants more in carrying out the practice and attracted more their attention to the worries on the screen, helping them in keeping focused. The water simulation also provided visual feedback on the distancing process. Perception of the effectiveness of AEON emerges clearly in some participants' comments. In particular, one user pointed out that AEON made his worries seem less important, while the manual activity of CARD mainly focused him on the will to actively drive away worries (rather than changing his perception of worries). Another user stated that AEON gave him the feeling of actually deleting his worries.

On the contrary, CARD could have at times focused participants attention more on the control of the motor activity required to successfully toss the cards into the wastebasket without missing it (and on the associated perception of the room environment) rather than on the worries themselves. This is reflected in some participants' comments. For example, one participant explicitly pointed out that he found more difficult to concentrate when practicing CARD and CLOUD than AEON. Another participant suggested that for her the manual activity of CARD might be more suited for other purposes such as letting go of anger.

Although the difference between AEON and CLOUD was in the hypothesized direction, the post-hoc did not reach statistical significance. CLOUD might not have suffered from the problem of CARD above described, thus letting users remain focused on their worries. We also observed that most participants closed their eyes during the practice of CLOUD, thus limiting the possibilities of being distracted by the environment.

2.3.7.3 Pleasantness

AEON was perceived to be significantly more pleasant to practice than the other two techniques. In addition to the considerations about computational off-loading, other factors that could have contributed to this result are the visual stimuli provided by the application which could be aesthetically pleasing, and the possibility to tactilely interact with a simulated natural element (water) to dissolve worries. This would be consistent with several users' comments, which described the water simulation as beautiful, relaxing and enjoyable. In particular, one participant stated that he found himself smiling while practicing this technique, while another one claimed that it fostered a feeling of personal wellbeing. Similarly, one participant claimed that AEON helped him a bit to let off steam.

The other two techniques could not have elicited the same feelings in users, due to their required abstract mental task (CLOUD) or less attractive motor activity (CARD). In one case, a participant reported that CARD made her nervous even after the practice ended.

This result, together with the fact that AEON was also perceived as less difficult, can be an important factor in helping naive meditators approach mindfulness as well as encourage its prolonged practice. Providing a simple and pleasant way to practice can help overcome the barriers naive meditators could encounter.

2.4 Qualitative In Situ Long-Term Study

The quantitative lab study we described in Section 2.3 shows that AEON obtained better results in terms of level of mindfulness achieved, perceived level of difficulty and degree of pleasantness than the two traditional techniques, suggesting that it could be a novel and effective way to help naive meditators approach mindfulness. However, as we outlined above, a critical contribution to the evaluation of technologies for health and well-being is to deeply understand how a proposed system is used by its target audience in situ (Klasnja et al., 2011).

For these reasons, we conducted a second study of AEON, which differs from the previous one in various ways: it is qualitative rather than quantitative, it was carried out in situ instead of the lab, it is aimed at investigating users' experience over a five-week period rather than a strictly structured use in a short lab test. Thus, while the quantitative study was aimed at finding if the app could actually help people in achieving a mindful state, this study is aimed at gaining a deeper understanding of users' perceptions in using the app for ameliorating worry, as well as discovering possible different patterns of use and design opportunities.

The five-week in situ study described in this section was carried out on naive meditators, who are the target users of AEON. Participants were asked to use the app with their worries at least once a day. Then, they were interviewed at the end of the five weeks.

2.4.1 Participants

Participants were recruited through direct contact. They were asked if they were willing to try a mobile app that might help in attenuating worries, on their mobile phones for a period of five weeks. They were informed that the app could run on iPhones (iOS 5 or higher) or Android smartphones (Android 2.2 or higher), and that they could keep the app as compensation for their participation. To identify naive meditators, we followed the criterion by Lau et al. (2006), who screened participants to ensure that they had no experience with any form of meditation (including yoga, tai chi, and qi-gong). They defined experience with mindfulness meditation as having at least 8 weeks of experience of daily practice. Thus, we asked participants if: (i) they have ever attended a course on or are practicing any form of meditation, e.g. mindfulness, yoga, tai-chi, qi-gong; (ii) they have ever practiced or are practicing daily meditation techniques for at least eight weeks. Participants who answered affirmatively both questions were not considered naive meditators and thus not eligible for enrollment in the study.

In total, we recruited 21 participants and 18 of them met the previously described criterion, thus forming the sample of our study. After five weeks of usage of AEON, they were contacted again for the interviews through the email address, Facebook contact or phone number they were asked to give us at the beginning of the study, but three of them did not show up: one participant told us that he did not use the app at all, one participant had accepted a job in another city, and one participant did not answer to our requests sent to her contact address. Among the 15 participants (5 M, 10F) who showed up for the interviews, 12 were students (5 undergraduate, 4 graduate, 3 PhD students; 9 computer science, 1 engineering, 1 business administration, 1 foreign languages), two had other occupations (primary school teacher, psychologist), and one was unemployed at the time of the study (his level of education was diploma of secondary education).

The age of participants ranged from 22 to 29 ($M = 25.47, SD = 2.39$). Two participants were left-handed and 13 were right-handed. On a self-report scale ranging from 1 (low familiarity) to 7 (high familiarity), participants reported to be very familiar with mobile touchscreen devices ($M = 6.47, SD = .83$).

2.4.1.1 Materials and Apparatus

The AEON app was originally developed for the iOS platform as described in Section 2.2.2. To have more recruitment possibilities, we ported it to the Android platform before starting the recruitment of participants for the current study. Figure 2.10 shows the "My Thoughts" screen in the Android version of the app.

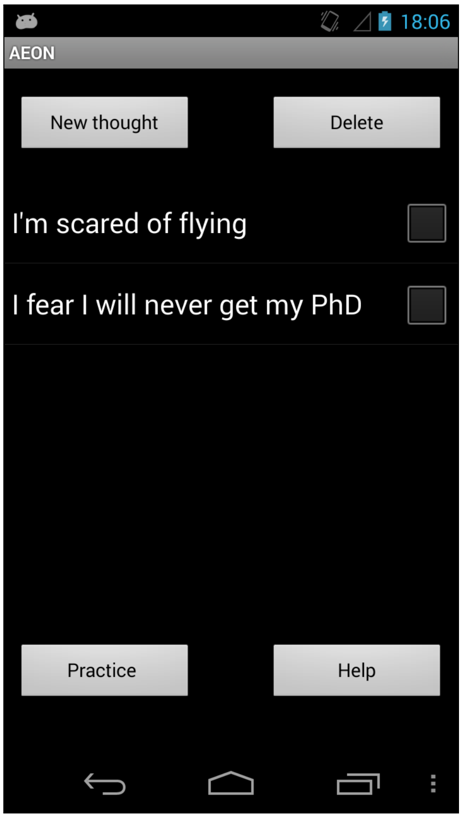


Figure 2.10: "My Thoughts" screen in the Android version of the app.

To make this screen as familiar as possible to users, we used the default themes and widgets (e.g. buttons) available in the Android development platform. Although this screen presents some differences in visual appearance on the two operating systems (see Figure 2.1 for the iOS version), the functions it offers and the way they are used are the same, as we described in Section 2.2.2. We made only two small changes in both versions of the app: (i) we added a button at the bottom right of the screen to allow users receive information about app usage (shown in a "Help" screen), and (ii) we restricted the length of the thought that a user could write to a maximum of 140 characters. This choice was inspired by the length that popular communication tools, such as Twitter, offer for entering single thoughts in a computer.

Unlike the "My thoughts" screen, the visual appearance and behavior of the interactive practice are identical in the iOS and the Android versions of the app. In particular, the behavior of the "Practice" screen is the one that we described in Section 2.2.2 with the only addition of a gesture that allows the user to return back to the "My Thoughts" screen at any moment, i.e. by swiping with two fingers from the bottom to the top border of the screen. Finally, to ensure that the interactive practice was the same for both versions of the app, we disabled the back button (available on Android devices) in the "Practice" screen.

The AEON app was installed on the participants' own mobile phones. Nine participants had an iPhone and six had an Android smartphone. Table 2.1 shows the main specifications of the devices used by participants.

Table 2.1: Main specifications of the device used by each participant.

Participant	Device model	Operating system	Screen size (in inches)	Resolution (in pixels)
P1	LG Optimus Pro	Android	2.8	240×320
P2	iPhone 4	iOS	3.5	640×960
P3	Sony Xperia S	Android	4.3	720×1280
P4	Samsung Galaxy Next Turbo	Android	3.14	240×320
P5	iPhone 4	iOS	3.5	640×960
P6	iPhone 4S	iOS	3.5	640×960
P7	iPhone 4S	iOS	3.5	640×960
P8	iPhone 5	iOS	4	640×1136
P9	Samsung GT-S5369	Android	3	240×320
P10	Samsung Galaxy S3 mini (i8190)	Android	4	480×800
P11	iPhone 4	iOS	3.5	640×960
P12	Samsung Galaxy Gio	Android	3.2	320×480
P13	iPhone 4S	iOS	3.5	640×960
P14	iPhone 4S	iOS	3.5	640×960
P15	iPhone 3GS	iOS	3.5	320×480

During the study, AEON collected usage time for each session of app use, and number of different thoughts used in the "Practice" screen during the entire period. Data was saved on the mobile device during the study and sent to a secure server when participants came to the interview.

2.4.1.2 Procedure

The experimental design included two meetings, one at the beginning and one at the end of the study period, and five weeks of usage of AEON in situ.

In the first meeting, participants were individually taken to a quiet room. They were informed that the study had the purpose of evaluating a mobile app which might help them attenuating their worries. No specific mindfulness concepts such as decentering were introduced or mentioned because we wanted to assess if users could possibly report the perception of such mindful state even if that was not an explicitly stated goal. Participants were also asked to sign a written consensus which, together with the purposes of the study and a description of the collected data, explained the measures we adopted to guarantee their privacy (data anonymization and safe server storage). Then, the experimenter installed AEON on the participants' mobile device, showed how to use it and remained available to clarify possible doubts. The experimenter also reassured participants that their entered worries were going to remain private and we had no way to receive or see them. This decision was taken because, while making the entered worries available to the experimenters would have been useful for further analyses, we reasoned that it could have prevented some participants from taking part in the study and could have led other participants to use the app in a limited, less natural way to avoid having other people read their more intimate worries.

Participants were then asked to pick up a printed randomly generated code from a box and enter it into the app. They were also asked to enter the same code on a computer and then fill out the demographic questionnaire. We collected questionnaire data with the computer to avoid possible transcription errors. After the completion of the demographic questionnaire, the computer generated and showed another code which participants had to enter into the AEON app to start the study. These codes were used to avoid storing the names of the persons in our database and into the app for data collection. Participants were encouraged to use AEON at least once a day and to contact the experimenter if they had any doubt or question. Participants were informed that after five weeks the app would stop working, showing them a screen (the "End Study" screen) inviting them to contact the experimenter to conclude the study.

Five weeks later, participants were taken again to a quiet room for the second meeting. They were asked to enter on a computer the code shown on the "End Study" screen of the AEON app. In response, the computer displayed another code which participants had to enter in the "End Study" screen. This made the "Send" button appear on the mobile device. Participants were then asked to press that button in order to send the data collected by the app to our server. They were informed that AEON was going to stop collecting data at that moment, but the app would remain available and working on their phones.

Participants were interviewed following a semi-structured approach to gather information about their experience (see Appendix A for the interview protocol). If necessary, to examine interesting issues spontaneously raised by participants, further questions were asked. Participants were asked if they agreed to have their interview recorded. To ensure participants' privacy, the interviews were saved with the code showed on the "End Study" screen. At the end of the interview, participants were thanked for their participation and reminded that they could keep AEON as compensation.

2.4.1.3 Data Analysis

The interviews we collected were transcribed verbatim. We then employed thematic analysis to identify and organize common and salient themes which emerged from the transcripts, following the steps outlined in (Braun and Clarke, 2006). We adopted an inductive approach as no assumptions were made prior to analysis on the themes that might emerge. The analysis involved: (i) reading and re-reading the transcripts to familiarize with the data, (ii) coding interesting features in all the dataset and collating data relevant to each code, (iii) combining all the codes into potential themes, gathering all data relevant to each potential theme and organizing themes into different levels (e.g., main overarching themes or sub-themes within them), (iv) checking if the themes and sub-themes made sense in relation to the coded extracts and to the entire data set, and (v) refining each theme and sub-theme, generating clear definitions and names. For step (iii), a theme was divided into one or more sub-themes when it was particularly large or complex.

These steps were carried out by the two authors of the thesis. Themes and sub-themes were reviewed together following an iterative process which ended when we reached agreement.

To assess the validity of the themes we identified, we then involved two external and independent coders. We created a codebook that they used to code the data, following (DeCuir-Gunby et al., 2011). The codes in the codebook identified the themes and sub-themes that were pointed out by the thematic analysis. For each code, we provided a label, a full definition, i.e. an extensive definition that provides inclusion and exclusion criteria, and an example extracted from the data set. The independent coders were briefly introduced to the concept of decentering. They were also told that they could apply two or more codes to the same fragment of text. Finally, as the coding process was carried out with a Computer Assisted Qualitative Data Analysis Software (CAQDAS), the independent coders were instructed to use a specific tool, i.e. Coding Analysis Toolkit (CAT) (Texifter, 2007), an open-source web-based CADQAS. In total, there were three coders, as we counted as one.

2.4.2 Results

2.4.2.1 Collected Data

We removed sessions shorter than 5 seconds, because we had previously determined that to launch the application, consider at least one thought, and exit the application requires more than that time. Sessions shorter than 5 seconds indicate that the user launched and then quickly exited the app, without practicing distancing from thoughts. Table 2.2 shows the resulting usage data for each participant. Figure 2.11 shows the distribution of the percentage of sessions throughout the day. Such percentage reaches the highest value during evening hours (10.9% of sessions during 22:00-22:59) and 46.8% of sessions occur from 19:00 to 02:59. Figure 2.12 shows the trend of the app mean usage per day over the five weeks of the study. The mean usage shows a quick drop after the first 3 days and then decreases only slowly until the end of the study.

Table 2.2: Usage data for each participant.

Participant	**Usage (days)**	**Usage (sec)**	**Mean usage per day**	**Considered thoughts**	**Sessions**	**Mean usage per session**
P1	14	3018	215.57	16	24	125.75
P2	2	106	53.00	1	2	53.00
P3	35	8947	255.63	95	125	71.58
P4	29	7448	256.83	37	39	109.97
P5	32	12753	398.53	71	40	318.26
P6	18	7693	427.39	25	65	118.35
P7	12	1146	95.50	10	19	60.32
P8	10	1287	128.70	5	20	64.35
P9	24	2546	106.10	7	53	48.04
P10	20	3174	158.70	34	30	105.80
P11	4	580	145.00	5	5	116.00
P12	26	4235	162.88	21	45	94.11
P13	20	4233	211.65	17	35	120.94
P14	8	1173	146.63	13	12	97.75
P15	21	10020	477.14	67	57	175.79
Mean	$M = 18.33$	$M = 4557.27$	$M = 215.95$	$M = 28.27$	$M = 38.10$	$M = 117.44$
Standard Dev.	$SD = 9.96$	$SD = 3895.28$	$SD = 127.11$	$SD = 28.13$	$SD = 30.42$	$SD = 69.53$

Usage (days): number of different days in which AEON was used; Usage (sec): total amount of AEON usage in seconds; Mean usage per day: Usage (sec)/Usage (days); Considered thoughts: number of different thoughts on which the user has practiced distancing from thoughts; Sessions: total number of sessions (after excluding those shorter than 5 seconds); Mean usage per session: Usage(sec)/Sessions.

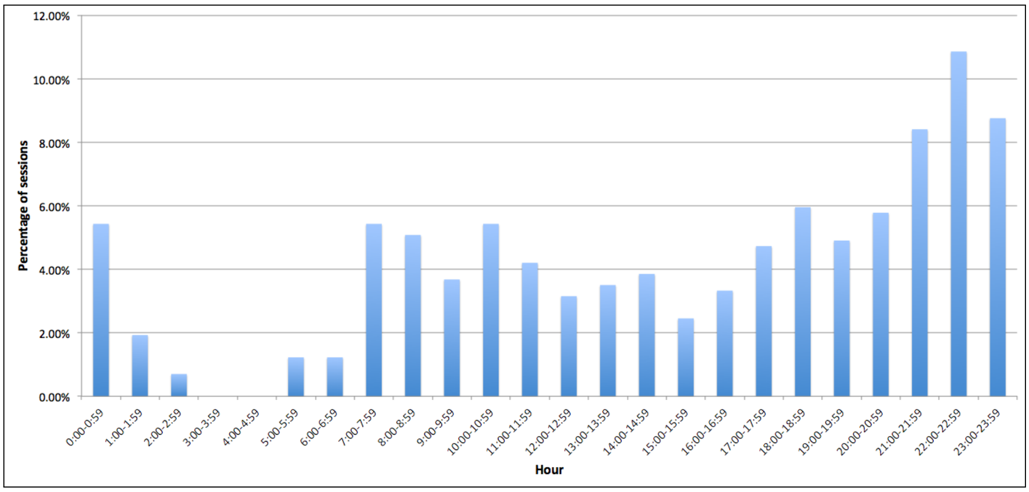


Figure 2.11: Distribution of the percentage of sessions throughout the day.

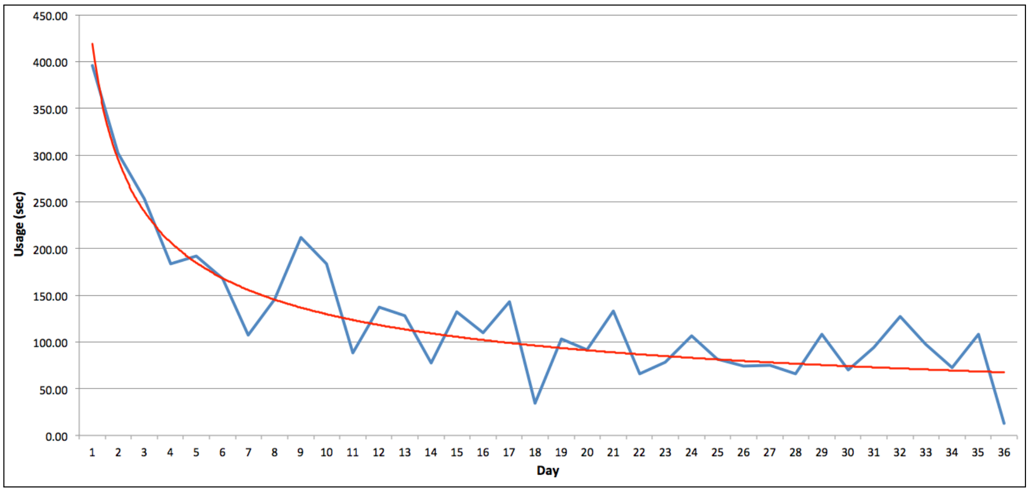


Figure 2.12: App mean usage per day. The red line indicates the trend of the mean.

2.4.2.2 Qualitative Data

Prior to thematic analysis, we looked at the usage statistics for each participant (Table 2.2). We discarded the interview data from participant P2, because she used AEON in an extremely limited way in terms of time (less than two minutes, split over 2 days) as well as considered thoughts (just 1). Moreover, during the interview, she provided only a few generic words that sounded as they were aimed at pleasing the interviewer.

After the coding of the dataset, we analyzed the level of agreement among coders using Fleiss' kappa (Fleiss, 1971). Results pointed out an overall kappa coefficient of .68, which indicates substantial agreement (Landis and Koch, 1977).

In the following, we present the results of the thematic analysis, organizing them into four topic areas: (i) "Decentering", which contains the themes that concern decentering from worries described by participants, (ii) "Feelings", which contains the themes that capture the feelings elicited by AEON in participants, (iii) "Patterns of use", which contains the themes that describe participants' particular patterns of use of AEON during the study, and (iv) "App features", which contains the themes that regard participants' perception of AEON features, including the problems they encountered and their suggestions for improvement.

In the following sections, we describe each theme providing sample extracts from the interviews. In the extracts, the parts in round brackets are the additional questions raised by the experimenter who led the interview, while the parts in square brackets are words or phrases added to clarify the meaning.

Decentering

Table 2.3 summarizes the themes and sub-themes that belong to the "Decentering" topic area. For each theme or sub-theme, the table indicates the participants to which it applies.

Table 2.3: Themes and sub-themes of the "Decentering" topic area. The last column indicates the participants for which each theme or sub-theme applies.

Theme	Sub-theme	Description	Participants
Focus on worries		AEON helped participants focus on their worries.	
	Focus on entered worries	When using AEON, participants thought about the worries they were entering into the app or about those worries already entered.	P1, P5, P6, P11, P15
	Focus on negative thoughts of the day or on seriously worrying thoughts, but not already entered into the app	When using AEON, participants focused on worries, bad feelings or emotions of the day, or on something that really worried them, but not already entered into the app.	P3, P4, P12, P14
	Focus on how to solve worries or to ease the tension	When using AEON, participants focused on how to solve the problems related to the worries they entered into it or on how to ease the tension caused by them.	P9, P13
Participants experienced decentering		AEON helped participants experience decentering from their worries. Participants point out different aspects of decentering.	
	Externalization, acceptance	AEON helped participants see their worries as external. Moreover, AEON helped participants better accept their worries.	P4, P8, P9, P10, P13, P14, P15

Table 2.3: Themes and sub-themes of the "Decentering" topic area. The last column indicates the participants for which each theme or sub-theme applies.

Theme	Sub-theme	Description	Participants
	Letting go	AEON helped participants let go of their worries.	P1, P4, P6, P13
	Impermanence of thoughts	AEON evoked in participants a reflection on the impermanence of worries.	P4, P12
Participants experienced decentering slightly or for a short time		AEON helped participants experience a stance of objectivity only slightly or only for a short period of time.	P3, P5, P8, P11
Participants did not experience decentering		AEON did not help participants experience decentering from some or all their worries.	
	No decentering	Using AEON, participants did not experience decentering from their worries.	P3, P7, P12
	No decentering with serious or long term worries	Using AEON, participants did not experience decentering from their serious or persistent worries. Participants provide an explanation.	P4, P6, P7, P13, P14

Focus on Worries

The majority of participants reported that, when using AEON, they focused on their worries. In particular, five of them said that they focused on the worries they were entering into the app or on those already entered. For example:

[When using AEON] I thought of my worries, mostly...yes [I thought] of what I was entering into it. - P5

Similarly:

[When using AEON] I thought about what was...into the app, [about] the worries I had entered into it. - P15

Instead, other four participants reported that, when using AEON, they focused on the worries, emotions or feelings of the day not already entered into the app. For example:

I thought about the bad feelings, the worries I had during the day. - P3

Two of these participants also referred to the difficulty to find thoughts to enter:

First of all, I had to think of something that really worried me, sometimes it was easier, like, I don't know, when I was more stressed, then in the days I was more normal it was more difficult to find the sentences to enter into [AEON], It was a burden. - P14

Finally, two participants pointed out that they thought about how to solve the problems related to the worries they entered into AEON or about how to ease the tension caused by them. For example:

[I thought] about what I had written [into the app] and [I thought] about how to solve it in real life more than in the technological life of... of the app. - P13

Participants experienced decentering

Almost all participants highlighted that they experienced decentering from their worries when using AEON. In particular, seven participants explicitly reported the experience of a stance of acceptance and objectivity. For example:

I would say I learned to stop worrying so much, I mean, I was aware of the worries I had, but, I would say that I handled them with more philosophy like... I was more quiet. - P8

Similarly, P15 said she was more confident to accept and overcome those difficulties she previously considered much more negatively.

For three participants, AEON was useful to experience decentering with worries they considered to be less serious. For example:

For stupid things, which however really bother me, like things related to my work, maybe yes [I related to my worries in a different way]. - P14

Similarly, P4 pointed out:

[. . .] while for other worries, those on affectivity or interpersonal relationships, you write [them inside the app] and then maybe you forget them, you. . . I don't know what it is due to, it's like either [the worry] was something external, something that wasn't part of you anymore, you see it farther, more objectively, or it's like digesting it and then saying ok, there was it [the worry] but I can go on, I can overcome it.

Like P4, P9 explicitly highlighted that her experienced stance of objectivity and acceptance was due to the fact that with AEON she could see her worries externalized. Indeed, she said:

[. . .] maybe seeing it [a worry] written is different than only thinking about it, maybe reducing it to one or three words, you say "ok, it's nothing to write home about" [...] as I said before, maybe seeing them [the worries] written, they were something more, ..., more concrete than you may expect.

In a subsequent moment of the interview, P9 reported also that the externalization of worries helped her to be less anxious.

The interviews revealed also that AEON helped four participants not to react in response to some thoughts, but to let them go. For example, P13:

[. . .] there were less difficult problems, that, without knowing how to deal with them, I was not able to solve, but dissolving a thought with the water, it was easier, sometimes even without writing it into the app, but just by thinking about how the app worked - I was able to let some thoughts go away more easily compared to other thoughts without using the app. I had the idea of the app in my mind.

Interestingly, P6 said that he used AEON to help him in letting go his urge to smoke:

In that period, I quit smoking and perhaps every time I had the urge to smoke, I wrote the thought "no smoking" and it could be a deterrent, in short, it could help me, I dissolved the thought, I imagined myself and I let go the urge to smoke (Are you saying that AEON helped you a little bit to quit smoking?) yes, because you're focused in that minute while you are dissolving "I want to smoke" and you intensively think about the cigarette and when you've finished thinking and dissolving, maybe you would focus on something else you need to do and the urge to smoke has gone.

Finally, two participants pointed out that the list of the entered worries shown in the "My Thoughts" screen evoked a reflection on their impermanence. For example:

[. . .] having the history of all the worries I had in the five weeks, it was useful to see how those which were giant worries two weeks ago, at the end were gone by. - P12

Participants experienced decentering slightly or for a short time

Two participants said that, when using AEON, they experienced a decentered stance on their worries only slightly, mainly due to their low interest and confidence in the approach offered by the app. For example:

It [AEON] helps a little bit to minimize [a worry] but maybe [. . .] I believe that if a person tends to worry too much on things, he would not say "ok it [AEON] completely calmed me", maybe it helps a little bit but not too much. - P3

Similarly, P11 pointed out:

I detached myself [from my worries] a little bit, but, maybe it's because I'm not influenced by these methods, but I did not... I detached myself a little bit, but it is not that... (so, did you detach a little bit from your worries?) yes, I detached myself a little bit, but it is not that I did not think of them [my worries] anymore [. . .]

Three participants (P11, P5, P8) said that the sense of detachment lasted for a short period of time, e.g.:

[I detached] when I was using it [AEON], i.e. just after I used it, but, it [the feeling of detachment] lasted about 10 minutes. - P11

Participants did not experience decentering

Seven participants pointed out that, when using AEON, they did not experience decentering from their worries, or from some of them. For example, P3, who reported the experience of a slight decentering for some worries, said:

Maybe I try to let myself go.. but at the end my approach to the worry did not change at all.

Similarly, P12, who highlighted a reflection on the impermanence of her worries, said that when using AEON she did not change her approach to the most two recent worries she entered. She suggested that this could be due to the fact she was stressed during the period of the study:

The application was one of the things to throw in during these days, however, I did not felt relieved by the application, I was a little bit stressed in general.

Five of the seven participants pointed out that they did not experience a decentered stance on their serious or long term worries when using AEON. For example, P4, who was almost at the end of her PhD, said:

[...] instead, worries related to my studies remain, in my opinion.

Two of these participants explicitly expressed their skepticism on the method offered by the app to attenuate their worries, for example:

I do not think that it [AEON] can actually drive away a worry unless it is only a trivial one, let's say, by simply swiping your finger on the screen and letting a sentence poetically go away with the water, to be honest it [AEON] does not make [a worry going away] a lot. - P7

Similarly, P6, who found AEON useful in helping him let go of his urge to smoke, said:

I always approached things as they were, I do not know, (and by writing and dissolving them?) To be honest as usual, because if I'm looking for a job, which is my worry, although I write and dissolve it, my interest to look for a job remains, and not having a job gets me down.

Feelings

Table 2.4 summarizes the themes and sub-themes that belong to the "Feelings" topic area. For each theme or sub-theme, the table indicates the participants to which it applies.

Table 2.4: Themes and sub-themes of the "Feelings" topic area. The last column indicates the participants for which each theme or sub-theme applies.

Theme	**Sub-theme**	**Description**	**Participants**
Positive feelings		AEON elicited positive feelings in participants for a short period of time (when they were using AEON and/or immediately after), attributed to its graphics or water simulation.	
	Types of positive feelings	AEON elicited a feeling of relaxation, well-being, pleasantness, and enjoyment in participants.	P1, P3, P4, P8, P9, P10, P12, P13, P14
	Causes of the positive feelings	The positive sensations elicited by AEON were attributed to its graphics or to the simulation of the natural element.	P4, P8, P9, P12, P14
No feelings		AEON did not elicit any particular feeling in participants when they were using it or in general during the period of the study.	P3, P7, P11, P14
Negative feelings		AEON elicited negative feelings in participants when they were using it.	
	Worries	The negative feelings were elicited by thinking about the worries entered into AEON.	P1, P4, P5, P13, P15

Table 2.4: Themes and sub-themes of the "Feelings" topic area. The last column indicates the participants for which each theme or sub-theme applies.

Theme	Sub-theme	Description	Participants
	Burden	The negative feelings were due to the fact that they had to force themselves to use AEON.	P4, P6

Positive feelings

The usage of AEON elicited a range of positive sensations in nine participants. The majority of them referred to relaxation, such as P1:

I was quite relaxed, although the app is meant to be used to write sad thoughts.

Another participant, P10, pointed out that the sense of relaxation helped her fall asleep during the evening, while P9 said that the usage of AEON was also enjoyable.

Other two of these participants referred instead to a sense of well-being:

[sometimes] it was like a relief valve, I liked a lot thinking that by deleting a thought I was able to get over some aspects of my life. - P13

P4 provided instead this explanation:

[. . .] But then I felt better, that is, once you focused on a worry and also played with it, in the end I would not say that it is like overcoming it but, maybe at the end of the day you say "I made it, I'm ok" (so, was it a good feeling?) yes, it was a nice sensation, almost relaxing, yes, like "I'm closing a bad thing and maybe I do not burden other people."- P4

Interestingly, five of these participants said that the positive sensations elicited by AEON were due to its graphics and the simulated natural element. For example, P14 explicitly mentioned the beauty of the graphics, while P8 indicated the water simulation:

I thought to be calm where I was sitting and I thought about the waves which were moving, thus it gave me a sense of relaxation.

Finally, the experience described by P12 is worth noting:

[. . .] it was..., the water was mesmerizing, indeed, although the written thought disappeared, I continued to move my finger because it was wonderful [. . .] I don't know if this mesmerizing was due to the thought I was dissolving but, actually, I felt attracted.

No feelings

Four participants pointed out they did not experience any particular feeling during the usage of AEON or in the long term. In particular, P3, who described a pleasant experience when using the app, reported that she was more worried in the period of the study, but said it depended on external circumstances unrelated to the app. Similarly, P14, who highlighted that the usage of the app was relaxing, pointed out that it could be difficult to understand whether she was more or less relaxed in general. Unlike these two participants, P11 provided this explanation:

I did not feel attracted, maybe it's because I'm very introverted, let's say, I tend to keep my worries for myself, I tend to tell them to nobody [...].

Similarly, P7, who pointed out he felt indifferent during the period of the study, also revealed to be skeptical about the approach offered by the app to attenuate his worries.

Negative feelings

Six participants reported that the usage of AEON elicited in them negative feelings. For four participants, these negative feelings were due to the worries they entered into the app. For example, P5 pointed out that she felt worried about the things she had to do in her life, which were the worries she entered into the app, while P1 said that sometimes her happiness turned to sadness when she read the worries she had previously entered into AEON. In line with these negative feelings, P13, said: *"70% of times [when I was using AEON] I felt depressed, due to the worries I was writing"*, while in *"the other 30% of times"* he felt a sense of well-being.
Another participant, i.e. P15, pointed out that she felt nervous when distancing from thoughts:

I felt a little bit nervous due to the fact of seeing the thoughts [written]... maybe dissolving them made me feel a little bit less nervous, but still, making contact... having to write them made me nervous [. . .] not always, but very often [. . .] I am not able to tell why [. . .]

Unlike these participants, P4 indicated the time needed to dissolve some worries in the "Practice" screen to be the cause of the negative sensations she felt (the necessary time needed to dissolve a thought in the app does not actually change with the thought). She pointed out:

[. . .] other [worries] needed instead more time [to be dissolved] and they elicited a little anxiety in me.

Other two of these participants pointed out that the negative feelings they felt were due to the fact that they had to force themselves to use the app. For example, P6, who at the beginning of the study felt confident about the possibility of alleviating his worries by using AEON, pointed out:

It [the usage of AEON] was a little bit boring at the end. I had to force myself to use it, because I used it when I had 5 minutes in which I did not know what else to do, but it was not the case that I woke up thinking "I have to use the app".

On the contrary, for P4 these feelings were elicited only at the beginning of the study, while towards the end the usage of AEON became a pleasant habit.

Patterns of Use

Table 2.5 summarizes the themes and sub-themes that belong to the "Patterns of use" topic area. For each theme or sub-theme, the table indicates the participants to which it applies.

Table 2.5: Themes and sub-themes of the "Patterns of use" topic area. The last column indicates the participants for which each theme or sub-theme applies.

Theme	**Sub-theme**	**Description**	**Participants**
Particular uses		AEON was used in particular moments and with particular strategies.	
	Use during the evening	AEON was mainly used during the evening, at the end of the day.	P3, P4, P5, P12
	Use during free time or anxious periods	AEON was mainly used during free time or anxious periods.	P1, P9
	Use as a diary or record of worries	AEON was used as a diary, to keep record of worries.	P4, P5, P12, P15
Forget to use		Participants forgot to use AEON.	P4, P7, P8, P9

Particular uses

Six participants pointed out that they used the app during particular moments. For example, one participant revealed that she used the app when she was more anxious:

I noticed that I used it [AEON] more, even like two or three times per day, when . . . when there were those days of total anxiety. - P9

Unlike P9, P1 said that she used the app during the empty moments of the day. Other four of these participants revealed they mainly used AEON during the evening and, for two of them, the app was useful to summarize the day they just spent. For example:

I used it [AEON] during the evening, because usually I did not have enough time [during the day], so it's when you review the things you've done during the day, so I would still use it during the evening. - P5

Considering how they used the app, it came out from four participants that they employed it for a purpose that was not part of the original design, i.e. they used the list of entered worries in the "My Thoughts" screen as a diary of their worries, for example:

It was like a diary [. . .] I saw this app as a diary. - P5

For two of these participants, i.e. P4 and P12, the list stimulated a reflection on the impermanence of their worries (as already described in the "Decentering" topic area).

Forget to use

Four participants revealed that sometimes they forgot to use the app and two of them provided an explanation. Indeed, P8 pointed out that this could be due to the period in which he tried the app, though he did not explain why the period was so peculiar. Instead, P4 said:

[. . .] I had to remember to use the app, because I did not use it exactly when there was a stressful moment, as in those moments I had to manage other things [. . .].

App Features

Table 2.6 summarizes the themes and sub-themes that belong to the "App features" topic area. For each theme or sub-theme, the table indicates the participants to which it applies.

Table 2.6: Themes and sub-themes of the "App features" topic area. The last column indicates the participants for which each theme or sub-theme applies.

Theme	Sub-theme	Description	Participants
Beautiful and easy-to-use app		AEON is a beautiful and easy-to-use app.	P1, P4, P5, P8, P9, P10, P11, P12, P13, P14
Usability issues		Usability issues.	
	Length of water simulation	The time required to delete a worry is too long.	P4, P6, P7
	Gestures	Problems related to the gestures needed in the "Practice" screen to go to the next worry or back to the "My Thoughts" screen.	P4, P7, P9, P12
	Other	Other usability issues.	P3, P5, P12
Suggestions for improvement		Suggestions for improvement.	
	Effects/Animation	Provide new effects or animations to dissolve worries.	P5, P10, P13, P14
	Background	Provide the possibility to change background color or image inside an effect.	P4, P5, P10, P11

Table 2.6: Themes and sub-themes of the “App features” topic area. The last column indicates the participants for which each theme or sub-theme applies.

Theme	Sub-theme	Description	Participants
	Music	Provide background music or ambient sounds.	P6, P7, P13
	Other	Other improvements.	P4, P5, P8, P9, P13

Beautiful and easy-to-use app

Eight participants pointed out that they liked the app. Some of them referred to the overall app, as, for example P11:

Interesting, let's say it's a novel app, a novel idea [. . .] maybe I will use it in the future.

Other participants explicitly referred to its graphics or to the animation:

The interface is nice and, in my opinion also the idea of water. - P14

Similarly, P4 said:

[. . .] the water effect is wonderful!

P4 and P14, together with other three participants, also pointed out that the app was easy to use. In particular, P4 highlighted the fact that the “Help” screen was useful to assist her when she did not remember how to use the app, while P13 said:

It's a nice idea [. . .] more than anything else, the strength of this app relies in the fact that it's simple to use.

Usability issues

Some participants pointed out they encountered usability issues when using the app. One concerned the water effect. Indeed, P4 said she had the subjective impression that sometimes the deletion of a worry lasted for a long time, while for other times it was shorter. She also pointed out that she felt a little bit anxious in the former case.

Other two participants suggested that the deletion of a worry should be shorter. For example, P6 said:

[. . .] maybe the deletion of a worry should be a little bit shorter, I mean, the water effect is too strong, the water should calm down in a shorter time.

Other problems were related to the gestures required to go to the next selected worry or to go back to the “My Thoughts” screen. For example, for three participants these gestures were not intuitive and thus were often performed in the wrong way. As a result, these participants triggered other waves instead of going back to the “My Thoughts” screen or go to the next selected thought.

P12, who was a Computer Science master student, formulated this hypothesis:

I think there is a usability problem [in the Practice screen], If I may say so, because I've a Samsung but that screen requires iPhone's gestures. I did always the wrong gesture to close [go back to the “My Thoughts” screen], because the Samsung requires the arrow [the button] [. . .] and also the scrolling [of the worries], because I'm not accustomed to use two fingers on the screen, maybe for those who use the iPhone it's normal, but it is not for those who use the Samsung.

P12 revealed also that she did not like the fact that every time she launched the app some worries were already selected in the list:

[. . .] because if I don't notice that the last worry in the list is selected, I then later notice there are more worries [in the Practice screen] than those I selected.

Similarly, when she pressed the “Delete” button in the “My Thoughts” screen:

When I had to delete [some worries], they were already selected... The list should have been empty.
Problems related to the list emerged also from another participant. Indeed, P3 said that she would have liked to have a worry automatically deleted from the list when she dissolved it in the "Practice" screen.

Suggestions for improvement

Participants provided different suggestions to improve the app. One of these is the possibility to use simulations of other natural elements to dissolve a worry. In particular, P5 and P13 suggested using fire. However, P5 said that fire could not be useful to provide relaxation, while P13 said:

An idea could be to use fire, but in my opinion, with fire the app could lose the aim for which it has been created.

Another suggestion, given by four participants, referred to the possibility to select among different colors or images for the background. For example, P4 would like to have a cyan background, while P11, explicitly suggested:

[I'd like to have] a little stones background rather than the parchment.

For three participants, background music or ambient sound could be an interesting feature. In particular, P13 suggested that a water sound might be relaxing.

Other suggestions were related to the worries. Indeed, P5 would like to have the possibility to enter longer worries, while P4 would like to enter and delete also positive thoughts. P9 suggested instead visualizing the worries in a different style in the "Practice" screen:

I would put the worries in bold, so you notice them more on the parchment and they stand out better when you are dissolving them.

P13 suggested sharing the worries with other users:

I would have shared some stupid worry with a friend of mine. For example, I would have said to [the person] "dissolve this thought with me!"

This participant pointed also out that the possibility to organize worries in categories could be an interesting additional feature.

Another participant wanted more privacy for the worries she entered:

When you launch the app you have all your thoughts and I think it would be appropriate to protect the app with a password... because you could leave the phone unlocked and someone launches the app and sees all your thoughts and... because I see the app as a diary, I would not want someone else to see them. - P5

Finally, P8 suggested that the app could provide some usage statistics, which could persuade people to use it more:

I would provide some data about the usage, which could give benefits to users [. . .] maybe an analysis of stress based on how a user moves his/her finger on the screen, maybe after one month you see that a user behaved in a different way [. . .] giving something more [usage statistics] could persuade users, because I forgot sometimes to use the app. - P8

2.4.3 Discussion and Design Opportunities

2.4.3.1 Mean Daily Usage

The analysis of the temporal data collected by the app over the five weeks pointed out that mean daily usage saw a quick drop after the first three days and then it decreased but very slowly until the end of the study period (see Figure 2.12). The longer time of daily usage in the first days of the study could be partially explained by the fact that participants had to enter their worries that were memorized in the app for later use as well as familiarize with the app itself. The app's novelty effect on participants could also have contributed to this result. The fact that mean daily usage does not increase towards the end of the study excludes the parking lot compliance effect (Stone et al., 2003), i.e. participants did not try to create the appearance of complying with the study protocol.

2.4.3.2 Decentering

The analysis of the interviews pointed out that AEON helped participants focus on their worries, which are part of their internal experience. Indeed, when using the app, some participants focused on the worries already entered, while others focused on something that really worried them but was not entered yet into the app. This result suggests that the usage of AEON can help naive meditators focus on their internal experience, which is the first step to achieve decentering from it. As pointed out by Shapiro et al. (2006), the capacity to bring attention to internal experience is a fundamental component of mindfulness and its practice could enhance the ability to inhibit secondary elaborative processing of thoughts, feelings and sensations, that is the ability to achieve decentering. In our case, the interviews pointed out that 13 participants experienced decentering from their worries or from some of them when using AEON. In particular, some participants referred to a stance of objectivity and acceptance toward their worries, while others pointed out that the usage of the app helped them let go of their worries. Four of these participants experienced decentering only slightly or for a short period of time, but also this result is interesting if we consider the fact that they were unfamiliar with mindfulness and decentering.

In addition to the considerations about the focus on internal experience, another factor that could have contributed to participants' experience of decentering is the high degree of *computational off-loading* that the app can offer to users. Computational off-loading refers to "the extent to which different external representations reduce the amount of cognitive effort required to solve informationally equivalent problems" (Rogers, 2004) and is part of the theory of external cognition (Scaife and Rogers, 1996). As pointed out by some participants, AEON offered them a way to see their worries and their progressive disappearance through an external visualization, which was helpful to perceive their worries as distant objects. This specific result is in line with the comments that participants provided in our lab evaluation of AEON (see Section 2.3) and suggests that the usage of AEON can help naive meditators in achieving a detached awareness of their internal experience, which according to Wells (2006) is one fundamental component of decentering.

Two participants pointed out that the list of the entered worries shown in the "My Thoughts" screen evoked in them a reflection on the impermanence of worries. This was an unexpected result, as the list was designed to be only a support for users to conveniently select the worries on which to practice. As suggested by the two participants, a possible explanation of this finding is that, when looking at the older worries, they realized that those worries were no longer important for them as they were tied to the particular past period in which they were entered into the app. This suggests that a more explicit highlighting of temporal information in the list of worries could be another factor that can help users experience decentering from their worries, as the awareness that thoughts and feelings are impermanent is one of the desired outcomes of mindfulness (Linehan, 1993b) and particularly decentering (Safran and Segal, 1996; Teasdale et al., 2002; Wells, 2006).

The analysis also pointed out that seven participants did not experience decentering from their most serious or persistent worries or from all their worries. Some of these participants explicitly pointed out their skepticism about the approach offered by AEON to attenuate their worries, which could explain this outcome. Indeed, this attitude might have had a negative impact on their intentions for using the app. As pointed out by Shapiro et al. (2006), the intention to practice is another necessary component of mindfulness, which sets the stage to achieve benefits from practicing. In addition to this consideration, it must also be noted that to experience decentering from serious or persistent worries a disciplined formal as well as informal mindfulness practice is needed on a daily basis, see e.g. (Kabat-Zinn, 2003). In this context, AEON is designed to be of help to naive meditators in one of the set of activities that need to be learned in the early stage of the practice of mindfulness, and the fact that the majority of participants in the study experienced decentering from some of their worries using the app is an encouraging result.

2.4.3.3 Feelings

AEON elicited different feelings in participants, both positive and negative. Considering the positive ones, when using the app nine participants felt a sense of enjoyment, well-being or relaxation, which was the most frequent feeling (reported by six of the nine participants). Interestingly, five of these participants explicitly pointed out that the positive feelings they felt when using the app were due to its graphics and simulation of the natural element.

This result can be explained by the fact that experiences with nature and natural elements, including passive ones, such as staring out a window or at a computer screen that displays images of nature, can have a restorative impact on people, e.g. stress reduction, relaxation and an overall restoration in energy and well-being (Bates and Marquit, 2011). Thus, in our case, the water simulation can have provided participants with a virtual experience of the corresponding natural element that was able to bring its restorative effect. Moreover, AEON allowed participants to interact with the simulated natural element with their hands, a factor that could have contributed to make the experience more realistic and pleasant.

This result suggests that the simulation of a natural element can be appropriate for the purposes of mindfulness apps. Moreover, the pleasant experience users derive from the interaction could persuade them to practice mindfulness more frequently.

On the other hand, for five participants the usage of AEON elicited negative feelings (worry, sadness, nervousness or anxiety), which were due to the worries they entered into the app. As pointed out by Baer (2003), the mental observation of internal conditions might induce phenomena in contrast to relaxation, such as autonomic arousal or racing thoughts. In our case, AEON could have offered participants a new and more concrete way to be in touch with their worries, and this might contribute to eliciting negative feelings. However, it must be noted that these participants also reported the experience of decentering from their worries or from some of them when using AEON. Thus, these negative feelings could be only an initial side effect that could be mitigated as users learn to practice mindfulness.

Finally, four participants pointed out that they did not experience any particular feeling when using the app or in general during the period of the study. In particular, two of them pointed out that they were skeptical about the method offered by AEON to attenuate their worries. This attitude towards the app might explain their low involvement.

2.4.3.4 Patterns of Use

Results pointed out different patterns of use of AEON. In particular, four participants indicated that they mainly used the app during the evening. This result is consistent with the data collected by the app (see Figure 2.12), and, as explicitly pointed out by two of these participants, this can be due to the fact that the evening is seen as a right time to briefly summarize what happened during the day. Moreover, this finding is in line with the suggestions of some mindfulness teachers, e.g. (Gunaratana, 2002; Kabat-Zinn, 2005), who point out that evening is a good moment to practice mindfulness, as it can help people to free their mind from the burden they accumulated during the day. Also, it can help them sleep better, and this was explicitly remarked also by one of the study participants.

The analysis revealed that four participants used the app, and in particular the list of worries in the "My Thoughts" screen, as a diary or a record of their worries. As pointed out in Section 2.4.2.2, for two of these participants looking at the list evoked a reflection on the impermanence of worries.

This unexpected result has interesting implications, suggesting that a distancing from thoughts app should offer users an easy way to record and see all the thoughts they entered also from a temporal point of view. A possible solution could be to organize the list with a calendar view. Indeed, as recently proposed by Hund et al. (2014), a continuous list representation of successive days (Figure 2.13) could let users be faster in searching for dates and for specific days when they are highlighted rather than with the standard grid layout, e.g. the one offered by iOS or Android.

In our case, the list view calendar could be useful to highlight the days in which users entered thoughts or practiced mindfulness with them (see Figure 2.14 for a possible adaptation).

Moreover, the list view could offer users the possibility to select a thought and trigger an animation which shows on a timeline when and for how long they practiced or not mindfulness with it. The animation could begin from the day in which the user entered the thought into the app (Figure 2.15a). While the timeline is scrolling, the animation could highlight the periods of time during which the user did not practice mindfulness with the selected thought (see example in Figure 2.15b) as well as the periods of time during which (s)he practiced on the thought (see example in Figure 2.15c). As a result, the animation could offer

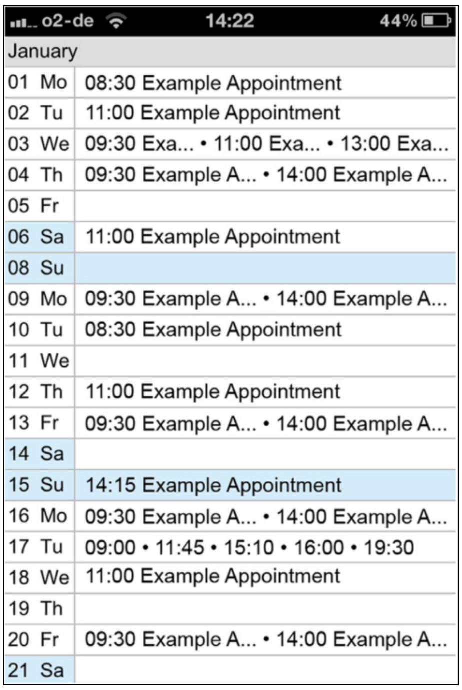


Figure 2.13: List view calendar (Hund et al., 2014)

users a simple way to see how the practice with their thoughts changed over time, which can foster a reflection on their impermanence.

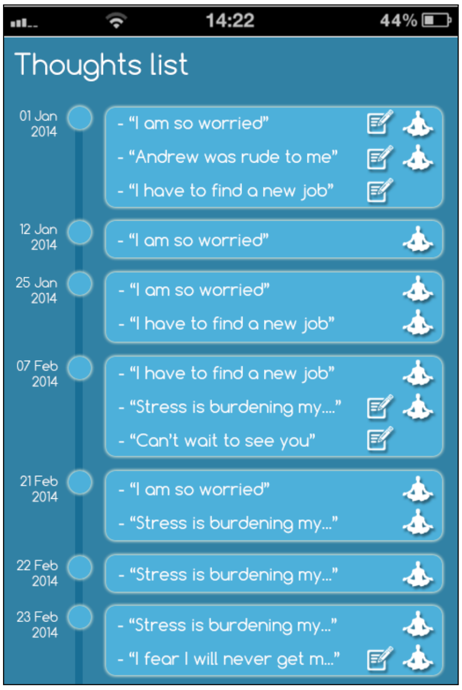


Figure 2.14: Possible adaptation of the list view calendar for a mindfulness app. The calendar shows the days in which the user entered thoughts (first icon) or practiced mindfulness with them (second icon).

(a) (b) (c)

Figure 2.15: Example of the timeline animation. (a) The animation begins from the day in which the selected thought has been entered into the app by the user. This day is highlighted in the timeline. (b) While the timeline is scrolling, the animation informs the user about the number of days in which (s)he has not dissolved the selected thought. (c) The animation reaches the first of the three consecutive days in which the user dissolved again the selected thought. These days are highlighted in the timeline.

Finally, the analysis pointed out that sometimes participants forgot to use AEON. This result could be due to the fact that, as the app proposed users a novel behavior, they were not being triggered to perform the behavior (Fogg, 2009). Although some participants pointed out that they mainly used AEON when they were more anxious, the addition of automatic prompts to use the app could be useful to help users develop a regular mindfulness practice.

To make the prompts effective, one should consider that users already receive a large number of notifications from their apps, which can be interrupting and frustrating. Thus, as suggested by Sahami Shirazi et al. (2014), it is necessary to find a balance between disrupting the user and providing valuable information. In our case, a possible solution could be based on two (optional) notifications per day which invite users to practice mindfulness. In general, a mindfulness app could provide users with some hints for setting notification times, following the suggestions of Gunaratana (2002) and Kabat-Zinn (2005), who point out that two good moments to practice mindfulness are at the beginning and at end of a day. Thus, for the first daily notification, the app could suggest users to indicate a time which closely follows the one they set for their smartphone's alarm clock. In this case, the notification could prompt users with messages like: "Begin your day with a mindful attitude". However, if users choose another time for the notification, as for example an afternoon hour, the prompt could be: "This is the moment you chose to practice mindfulness, take your time and begin the observation of your thoughts".

For the second daily notification, the app could suggest users to indicate a period of time just before they usually have dinner or before they go to bed. In these two cases, the prompts could respectively be like "Let's free the mind of the mental stress you accumulated during your day" and "Let's get rid of the burdens before you sleep". Overall, allowing the user to choose how many notifications to receive and tailor their time could persuade users to use more the app, as pointed out for such kind of tailoring by (Fogg, 2009).

2.4.3.5 App Features

The analysis pointed out that almost all participants liked the app. This could be due to the fact that AEON offered users a visual experience that could be aesthetically pleasing and, as pointed out before, provided them with the possibility to tactilely interact with the simulated natural element. Participants also found AEON easy to use probably because it requires just a few simple steps to enter a thought and practice distancing from it.

The analysis also pointed out that participants experienced some problems when using the app. One of them concerned the gestures required by the app to change worry or to go back to the "My Thoughts" screen, which four participants considered as counterintuitive. Interestingly, all but one of them had an Android device. These participants could be more familiar with the use of a physical button to navigate between different screens of an app, typical of Android devices, than swipe gestures. However, we did not want to introduce differences in the distancing from thoughts practice in the Android and iOS versions of the app used in the study. More tailoring to the platform conventions should be considered for interactive practice that reach the commercial phase. In distancing from thoughts, different methods should be used to implement the gestures to change thought or to exit from the practice. For example, a possible solution could consist in the gradual appearance of two buttons at the bottom of the screen, i.e. one to change worry and one to exit the practice, when a thought has completely disappeared. Although these buttons could be perceived as extraneous elements unrelated to the practice of mindfulness, they could make the usage of the app easier.

Other suggestions for improvement emerged from the analysis. Most of them concerned the possibility to select among different simulations of natural elements to dissolve a worry or to choose among different background images or colors. Finally, other participants suggested using a background music or an ambient sound. For example, the simulation of wind could provide an alternative to water: a thought could be displayed as written on the sky with a cloudy style; then, as users move their fingers on the screen, they trigger gentle air blows which make all the words of the thought spread out on the sky and progressively disappear.

However, the effectiveness of new natural element simulations in helping users achieve decentering will have to be carefully evaluated.

Another suggestion concerns the possibility to provide users with usage statistics. This could be an interesting feature for a mindfulness app and, for this purpose, we can take inspiration from personal informatics and quantified self systems, i.e. tools that help people collect personally relevant information for the purpose of self-reflection and gaining self-knowledge to help people become more aware of their own behavior, take better decisions, and change behavior (Li et al., 2010, 2011). However, a trade off must be found between the mainly qualitative experience of mindfulness and the quantified self-tools that can be included in the app. Indeed, the introduction of such monitoring capabilities could focus users mainly on the process of data collection and data analysis, striving for an improvement of such data. Unfortunately, as pointed out by Kabat-Zinn (1990), a striving attitude could undermine the cultivation of mindfulness. Therefore, a mindfulness app should collect and display data in ways that are consistent with its purpose, i.e. helping users achieve a state that requires time to be cultivated. For example, the app could record the time users spent dissolving their thoughts and offer them the possibility to assess their level of decentering no more than weekly, e.g. by using the EQ questionnaire (Fresco et al., 2007). Interestingly, as pointed out by (Plaza et al., 2013), none of the current mindfulness apps provides assessment features based on recognized questionnaires. Then, the app could analyze the collected data, for example by correlating time of use and changes in the level of decentering from week to week or month to month, to assess if decentering increased when users practiced more with the app. In describing the results to users, the app could follow the qualitative method described in (Bentley et al., 2013), which employs natural language to present significant connections between the data collected by a mobile app, e.g. users' weight, sleep hours, weather and mood. Results of the study described in (Bentley et al., 2013) pointed out that this method helped users in increasing self-understanding that led to focused behavior changes. In our specific case, once enough data is collected and the result of a new statistical analysis is available, the app can present users a message in the form: "You were more mindful last week, when you used more the app" or "Your level of decentering increased this month", which could help them gain a better understanding on how they are performing the learning and practice of mindfulness.

Lastly, another suggestion that emerged consists in the possibility for users to share the process to dissolve a worry together with other users. This could be an interesting feature to implement, as it allows leveraging the principle of social facilitation, i.e. the fact that people are more likely to perform a target behavior if they can discern via technology that others are performing the behavior along with them (Fogg, 2002), and thus persuade people to use more the app. However, as learning mindfulness involves the training of attention (Shapiro et al., 2006), this feature introduces a risk of possible distraction to users when practicing and should be carefully evaluated.

2.5 Quantitative In Situ Long-Term Study

The purpose of the qualitative study described in Section 2.4 was to thoroughly investigate users' experience with AEON. In particular, we were interested to gain a deep understanding of users' perceptions in using the app for ameliorating worry, as well as discovering possible different patterns of use and design opportunities. Results point out that several participants experienced decentering from their worries when using the app, suggesting that the prolonged use of AEON might be beneficial for people.

With the study described in this section, we aim at quantitatively investigating the effectiveness of AEON when used by people in their everyday contexts for a prolonged period of time, as well as extend the study sample. For such purposes, we included into the app a questionnaire to assess mindfulness and employed the research in the large methodology (see Section 1.2) to carry out a 4-week study. Interestingly, no interactive practice has been evaluated in participants' natural settings over a prolonged period of time. Indeed, the review of the literature on computer-supported mindfulness included in Section 2.1.2 highlights two different situations. On one side, there are a few studies that showed the efficacy of interactive practices in helping naive meditators achieve mindfulness, but they consist of evaluations carried out in a lab or in a similar artificial setting that assess only short-term effects. On the other side, there are studies conducted over a prolonged period of time in everyday settings, but they analyze traditional practices taught without computers, see (Keng et al., 2011) for examples, or using computers only to provide teaching materials at a distance.

The research methodology we employ for this evaluation has been previously used in mindfulness studies. Indeed, Killingsworth and Gilbert (2010) used research in the large to collect happiness and mindfulness reports from a large number of users, while Runyan et al. (2013) employed it to deliver a behavioral assessment and intervention aimed at increasing self-awareness. However, the mobile apps used in such studies were only aimed at prompting specific questions to participants and did not provide any support to mindfulness practice.

To the best of our knowledge, no study employed research in the large to evaluate a mobile mindfulness app. This research approach would allow the obtained results to have a greater ecological validity than those of the evaluations carried out so far (Miller, 2012).

Participants could take part in this study by downloading AEON from Apple's App Store as well as Google Play and accepting the study conditions that were shown after the first launch of the app. To measure the level of participant's mindfulness, we included into the app a mindfulness questionnaire that assessed decentering. Participants were asked to answer the mindfulness questionnaire three times: (i) after acceptance of the study conditions (START), (ii) two weeks after (2WEEKS), and (iii) four weeks after (4WEEKS) the acceptance of the conditions. Finally, at the end of the four weeks, the app proposed to fill out an optional qualitative questionnaire.

Since AEON was designed with naive meditators in mind, we also included a short initial questionnaire to distinguish naive meditators from people with experience with meditation (in the following, *experienced meditators*). This distinction allows us to investigate possible differences between the two categories of participants.

2.5.1 Hypothesis

Since our previous evaluation of AEON pointed out that the app can help people achieve decentering during a short usage session (see Section 2.3), we hypothesize that the prolonged use of the app could help them cultivating such mindful state and improving it over time. In particular, we expect that participants' level of decentering will grow after two and then after four weeks of app use.

The study of effects on experienced meditators is instead more exploratory in nature. Indeed, experienced meditators might already show high levels of decentering at the beginning of the study, and using the mobile app might not necessarily be beneficial to them, also because, if they are already well trained in mentally practicing distancing from thoughts, a technological adjunct that introduces to the practice might be more hindrance than help.

2.5.2 Materials and Apparatus

For this study, we ported the app to the Unity development environment (Unity Technologies, 2013) to make it multi-platform (Android and iOS) as in the study described in Section 2.4. This choice allowed us to (i) extended the potential population of participants, as the two platforms might have different types of users (Cramer et al., 2010), and (ii) provide an identical visual appearance and behavior of the app on the two operating systems.

Figure 2.16 shows the "My Thoughts" screen of the Unity version of the app. The behavior of this screen is the same as the one described in Section 2.2.2 and in Section 2.4.1.1 Moreover, we provided users with a quick tutorial shown the first time they launched the app (see Figure 2.17a and 2.17b).

We provided users with a quick tutorial also the first time they entered the "Practice" screen (see Figure 2.18a), whose appearance and behavior is the same as the one described in Section 2.2.2, and the first time they dissolved a thought, to illustrate the swipe actions (see Figure 2.18b).

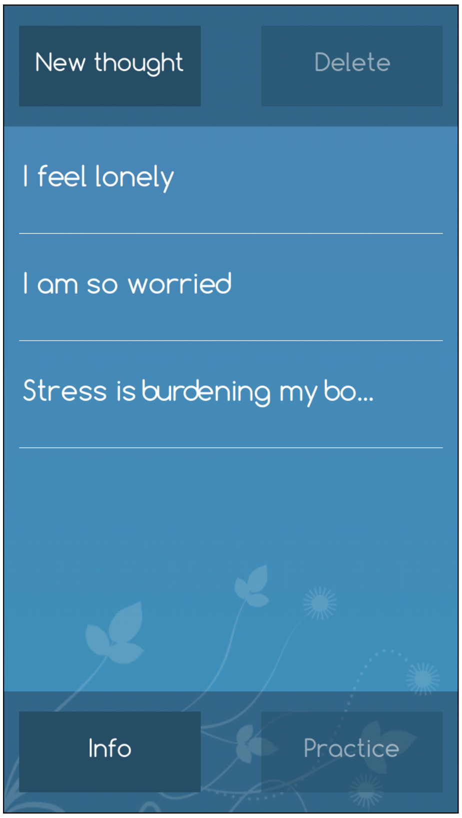
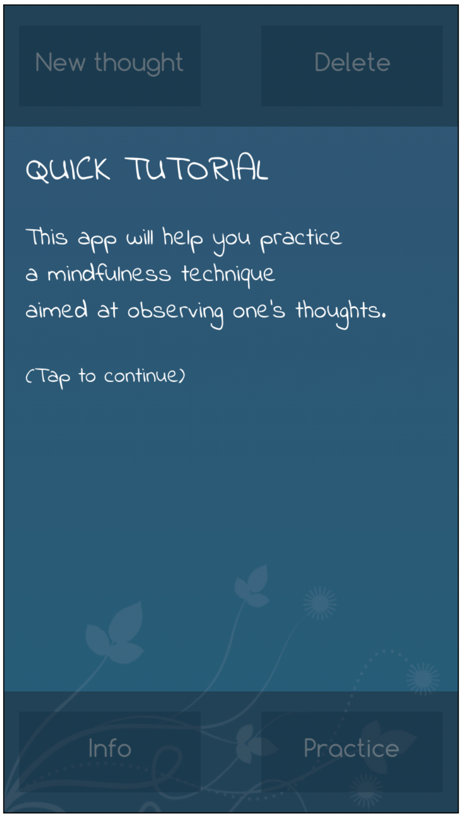
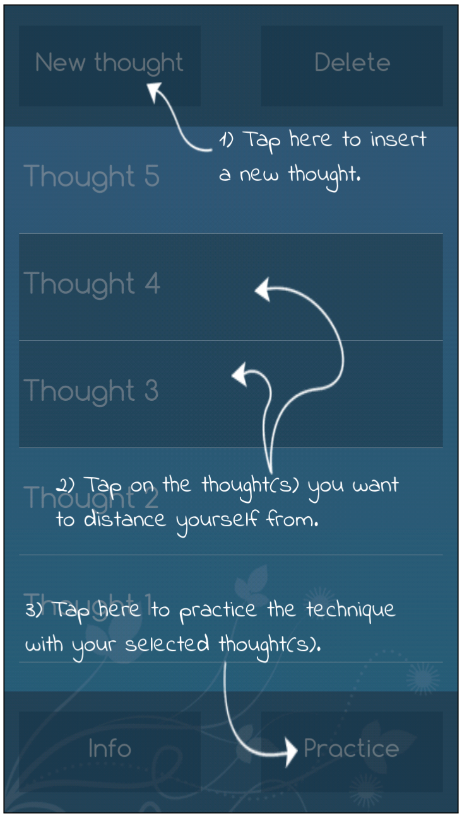


Figure 2.16: “My Thoughts” screen.



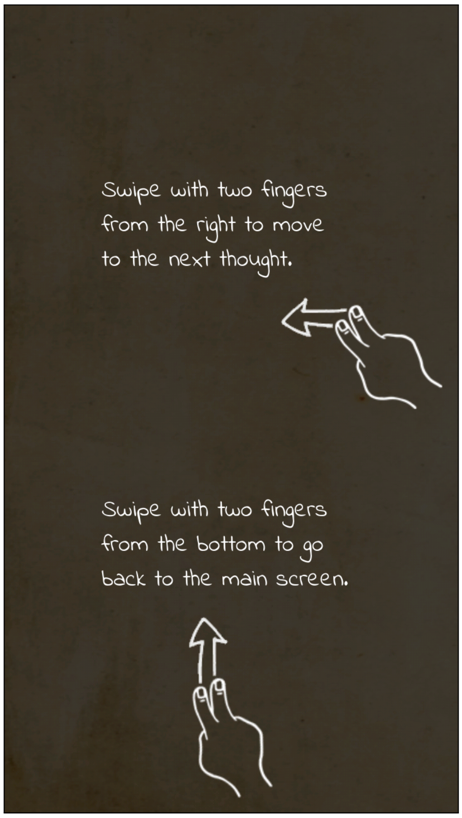
(a)



(b)

Figure 2.17: Tutorial of the “My Thoughts” screen. First page (a); Second page (b).

(a)



(b)

Figure 2.18: Quick tutorials for the “Practice” screen (a) and for the swipe actions (b).

2.5.3 Measures

2.5.3.1 Demographics and Usage Data

A short initial questionnaire included into the app (see Figure 2.20) was used to collect participants' age and gender, and to identify naive meditators. To this purpose, we followed the criteria by Lau et al. (2006), who considered as naive meditators those participants having no experience at all or less than 8 weeks of experience of daily practice with mindfulness or any form of meditation (including yoga, tai chi, and qi-gong). Thus, we asked participants if: (i) they had ever attended a course on or were practicing any form of meditation, e.g. mindfulness, yoga, tai-chi, qi-gong; (ii) they had ever practiced or were practicing daily meditation techniques for at least eight weeks. We considered as naive meditators those participants who answered (i) negatively both questions, or (ii) affirmatively only the first one. On the contrary, participants who answered affirmatively both questions were considered experienced meditators.

During the study period, AEON logged the total amount of time participants spent in the "Practice" screen (hereinafter, *practice session*) and the total number of different thoughts they practiced on. The app collected also some device characteristics, i.e. the operating system, the language that participants set in the device and the screen resolution. Finally, the app assigned a unique and anonymous code to each participant, generating it from the device identifier.

All data was saved on the mobile device during the study and sent to a secure server when an Internet connection was available.

At the end of the study, the data collected by the app was used to derive the values of the following variables for each participant:

- total practice time: total amount of time (in seconds) the participant practiced distancing from thoughts;
- daily practice times: practice time during each day of the study. If a practice session spans two days, the practice time is associated to the day in which the session started;
- distribution of practice sessions throughout the day: how the total number of practice sessions in a day is distributed (in percentage) on each of the single hours. If a practice session spans two hours, the session is associated to the hour in which the participant started practicing;
- total days of practice: total number of days in which the participant practiced distancing from thoughts during the study period;
- total number of thoughts: total number of different thoughts on which the participant practiced distancing from thoughts during the study period.

2.5.3.2 Decentering

Decentering was measured with the 11-item Decentering Subscale of the Experience Questionnaire (DEQ) (Fresco et al., 2007), which was included into the app (see Figure 2.21). The subscale includes items such as "I am better able to accept myself as I am" or "I am not so easily carried away by my thoughts and feelings" - the full list of items is available in (Fresco et al., 2007). Items are rated on a 5-point scale ("never", "rarely", "sometimes", "often", "all the time") where "never" corresponds to 1 and "all the time" to 5. Ratings are summed to obtain an overall score that can range from 11 to 55. We measured internal reliability with Cronbach's alpha, $\alpha = .86$ (START), .88 (2WEEKS), .93 (4WEEKS).

2.5.3.3 Qualitative Feedback

The optional qualitative questionnaire presented at the end of the study contained six open-ended questions that asked participants about their experience with the app (see Table 2.7). The questionnaire was included into the app and participants could write their answers in the text area that was available below each the question (see Figure 2.22).

Table 2.7: Questions of the qualitative questionnaire.

N.	Question
1	What did you think while you were using the app?
2	How did you feel while you were using the app?
3	Did you notice anything new in your days or in yourself during the period of the study?
4	Did you relate to your worries or think about them differently during this period?
5	Now that you are familiar with the app, what do you think about it?
6	Is there anything that you would change or improve in the app?

2.5.4 Method and Procedure

The study was based on a within subjects design with the assessment point (START, 2WEEKS and 4WEEKS) as independent variable.

We released AEON on Google Play the 21st of February 2014 and on Apple's App Store on the 11th of March and participants with devices running at least Android 4.0.3 or iOS 7 could download it for free. On both app stores, the app was listed under the "Health & Fitness" category with the title "AEON Mindfulness app". In this way, participants could find it both by browsing the app category or by searching for apps using "mindfulness" as keyword.

The language used by the app menus and tutorial was English, but participants were able to enter their thoughts for the interactive practice in their preferred language. To encourage participants in using the app and answering the mindfulness questionnaire during the study period, we defined an incentives mechanism, following the general suggestions of previous studies that employed a research in the large paradigm, e.g. (Ferreira et al., 2012; Henze et al., 2011a; Miluzzo et al., 2010). In the following, we describe the incentives mechanism and the whole experimental procedure. When launched for the first time, the app informed participants that by using it they would participate in a study. Purpose of the study, data collected and anonymization measures were briefly explained. Additional information on the app and the terms and conditions of use were available to participants through two optional screens. Participants were then asked to use the app for a period of four weeks. As an incentive, the app informed participants that at the end of the study it was going to: (i) remain free to use, and (ii) unlock new "cool" features (see Figure 2.19).

When participants gave their consent to participate in the study, the app asked them to fill (i) the initial short questionnaire depicted in Figure 2.20 and (ii) the mindfulness questionnaire (see Figure 2.21).

The app exploited the notification system of the smartphone to ask participants to fill out again the mindfulness questionnaire two and four weeks after they filled it the first time. Each time, participants were remembered that by answering the questionnaire they were going to receive new app features at the end of the study.

After receiving the notification, participants could answer the questionnaire within three days. This was done to give participants the opportunity to answer the questionnaire at a convenient time when they were not in a rush and, at the same time, to ensure that they answered within the same limited number of days. If they did not answer within three days, they were excluded from the study. In this case, the app stopped collecting data, remained free to use but did not unlock new features.

After completing the third mindfulness questionnaire at the end of the study, participants received a proposal from the app to fill out the optional qualitative questionnaire (see Figure 2.22) and were informed that an extra feature was going be unlocked if they answered also that questionnaire.

Finally, the app thanked participants for their participation in the study and illustrated them the new features that were unlocked, i.e. the possibility to choose among a set of different backgrounds for the "Practice" screen and the possibility to see a line graph of their three measurements of mindfulness taken

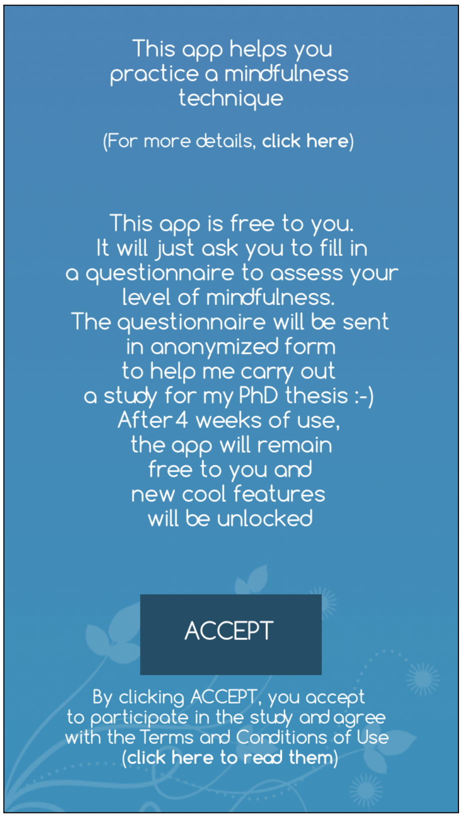
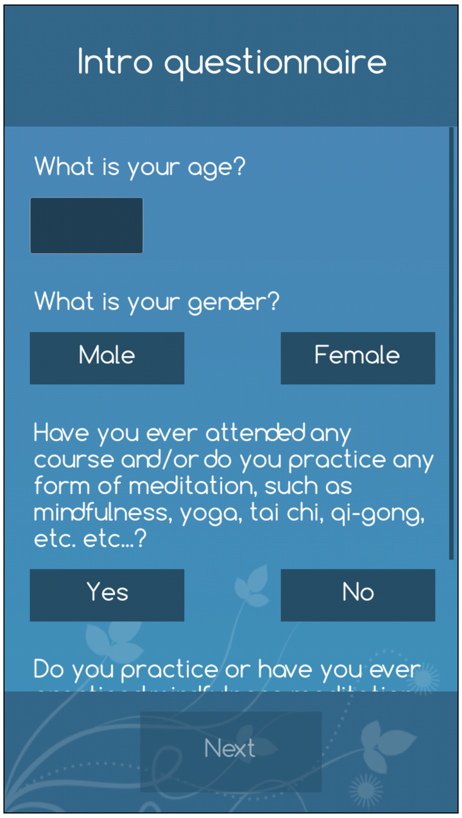


Figure 2.19: Screen shown by the app after its first launch.

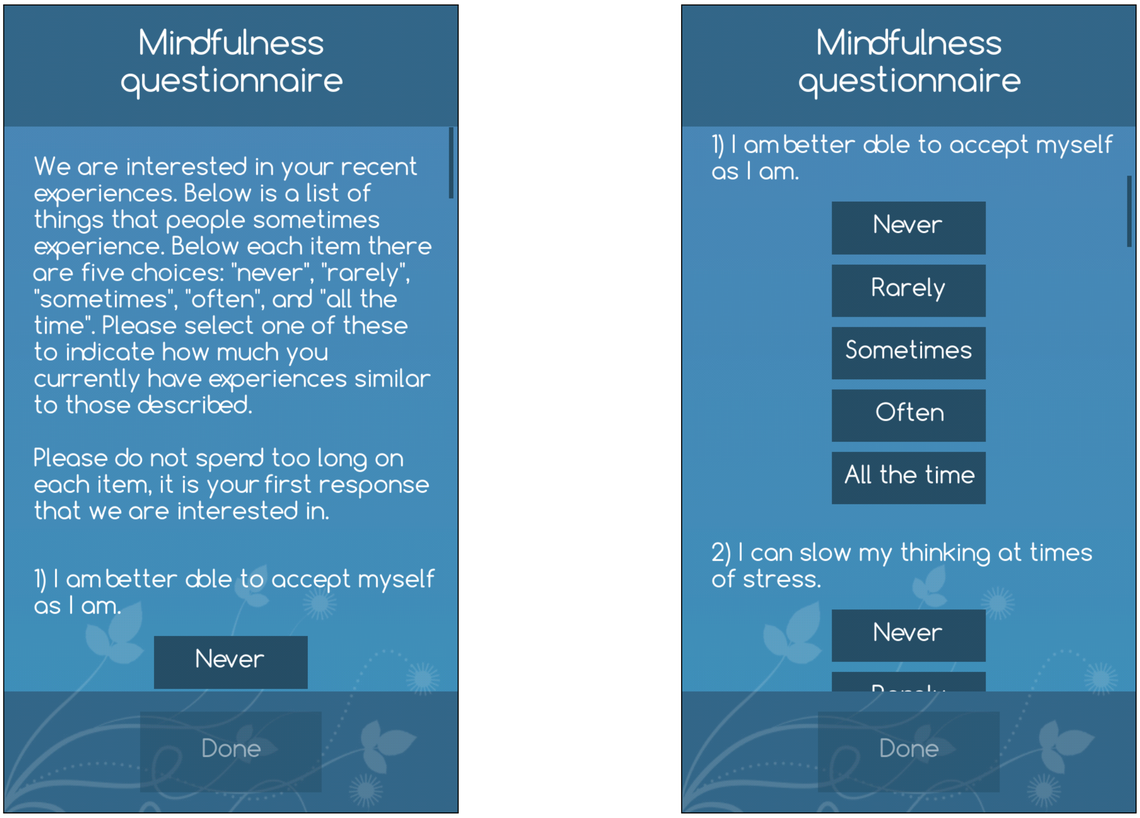
during the study. Participants who answered the qualitative questionnaire unlocked also an extra feature (water sound effects during the interactive practice, that can be turned on or off).



(a)

(b)

Figure 2.20: Initial questionnaire: demographic questions (a); questions to identify naive and experienced meditators (b).



(a) (b)

Figure 2.21: Mindfulness questionnaire: instructions (a); the first two items (b).

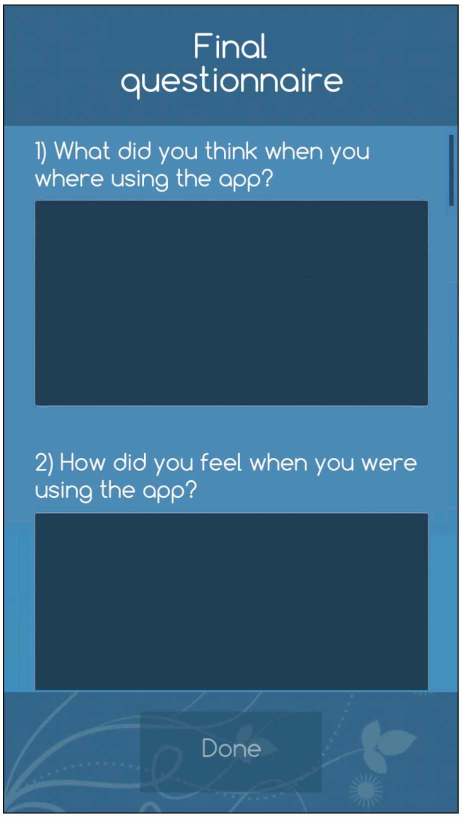


Figure 2.22: Screen shown by the app after its first launch.

2.5.5 Results

2.5.5.1 Demographics and Usage Data

Table 2.8 shows the number of users who completed each of the following steps with the app between the 21st of February (11th of March for the iOS version) and the 8th of October 2014: (i) download, i.e. downloading and installing of the app, (ii) accept conditions, i.e. clicking on the “Accept” button in the initial screen (Figure 2.19), (iii) initial questionnaire, i.e. completion of the questionnaire in Figure 2.20, (iv) mindfulness START, i.e. completion of the mindfulness questionnaire (Figure 2.21) at the beginning of the study, (v) mindfulness 2WEEKS, i.e. completion of the mindfulness questionnaire after two weeks, (vi) mindfulness 4WEEKS, i.e. completion of the mindfulness questionnaire after four weeks, and (vii) qualitative questionnaire, i.e. completion of the optional qualitative questionnaire (Figure 2.22). For each step, the table indicates on how many Android and iOS devices it was completed. From step (iii) on, it provides also the number of naive meditators (NM) and experienced meditators (EM). Finally, from step (iv) on, it also highlights on how many devices the interactive practice was used (Practice used) and on how many it was never used during the study (Practice never used).

Table 2.8: Number of participants for each step of the study.

Completion of step	Total devices	Practice used	Practice never used
(i) Download	3979 (2791 Android, 1188 iOS)[a]		
(ii) Accept conditions	2997 (2132 Android, 865 iOS)		
(iii) Initial questionnaire	2891 (2065 Android, 826 iOS; 2549 NM, 342 EM)		
(iv) Mindfulness START	2817 (2009 Android, 808 iOS; 2485 NM, 332 EM)	1540 (1127 Android, 413 iOS; 1374 NM, 166 EM)	1277 (882 Android, 395 iOS; 1111 NM, 166 EM)
(v) Mindfulness 2WEEKS	386 (255 Android, 131 iOS; 326 NM, 60 EM)	348 (231 Android, 117 iOS; 295 NM, 53 EM)	38 (24 Android, 14 iOS; 31 NM, 7 EM)
(vi) Mindfulness 4WEEKS	147 (96 Android, 51 iOS; 128 NM, 19 EM)	136 (89 Android, 47 iOS; 120 NM, 16 EM)	11 (7 Android, 4 iOS; 8 NM, 3 EM)
(vii) Qualitative questionnaire	66 (46 Android, 20 iOS; 54 NM, 12 EM)	62 (43 Android, 19 iOS; 52 NM, 10 EM)	4 (3 Android, 1 iOS; 2 NM, 2 EM)

[a] The count of iOS devices can be lower than the actual number of devices in which the app has been installed. Indeed, from iTunes Connect we could only get the number of installs by user and not the number of devices in which the same user installed the app.

The sample of our study was thus formed by the 136 participants who answered all three mindfulness questionnaires and practiced distancing from thoughts with the app. Based on participants' self-reported data, there were 39 male and 97 female respondent in the sample, and there were 120 naive meditators and 16 experienced meditators. Figure 2.23 shows the distribution of participants' self-reported age ($M = 37.85, SD = 11.40$).

Table 2.9 reports data about participants' device language. The most common language was English (69.1% of devices), followed by German (6.6%) and Swedish (5.1%).

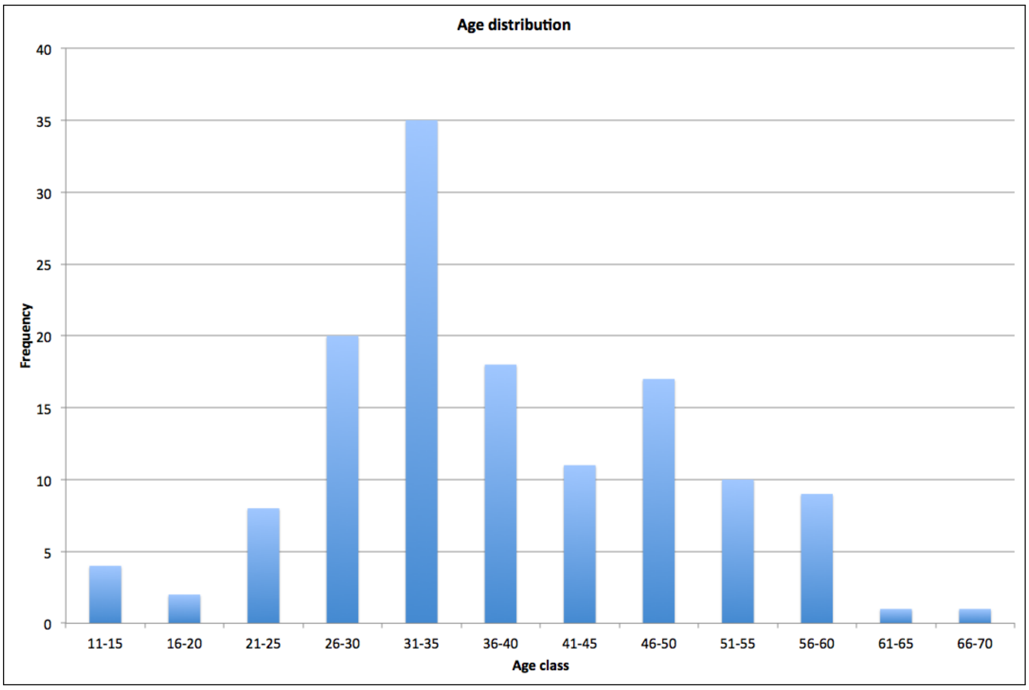


Figure 2.23: Distribution of participants' self-reported age.

Table 2.9: Participants' device languages and their frequency.

Language	Frequency	Percentage
Arabic	1	0.7%
Chinese	1	0.7%
Czech	1	0.7%
Danish	2	1.5%
Dutch	6	4.4%
English	94	69.1%
French	2	1.5%
German	9	6.6%
Hungarian	1	0.7%
Italian	4	2.9%
Portuguese	2	1.5%
Russian	1	0.7%
Slovenian	1	0.7%
Spanish	2	1.5%
Swedish	7	5.1%
Unknown	2	1.5%

Figure 2.24 illustrated data about participants' device resolution. The most common resolution was 1080×1920 (19.1% of devices), followed by 720×1280 (13.2% of devices) and 480×800 (12.5% of devices).

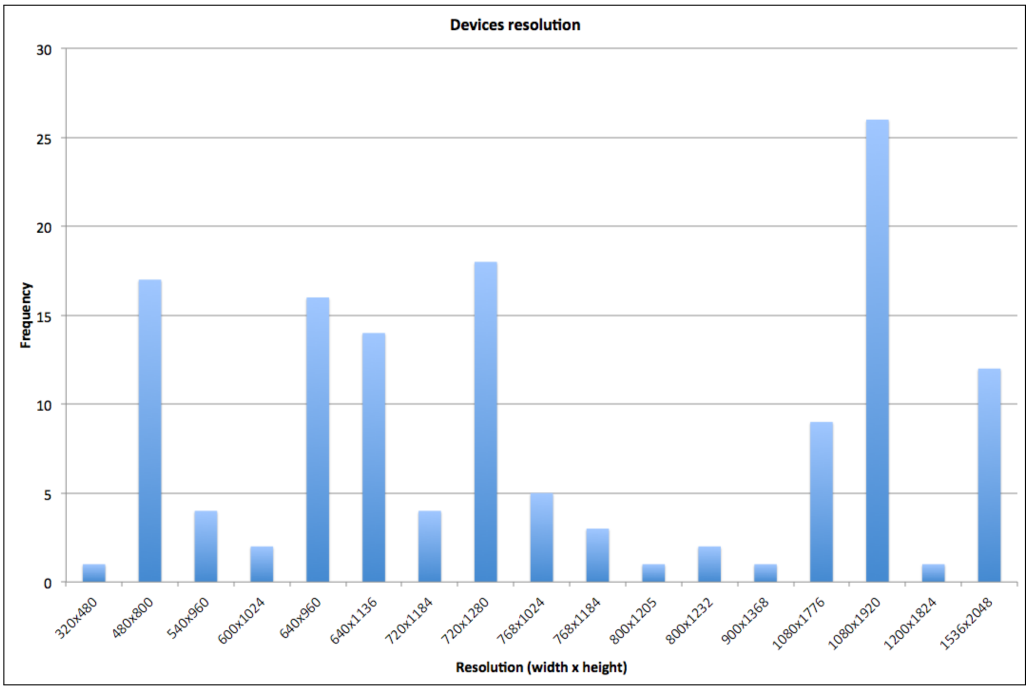


Figure 2.24: Participants' device resolution.

Figure 2.25 shows when the user practiced during the day, considering the practice sessions of all participants. The highest percentages of sessions occurred during evening hours (10.2% of sessions start during 21:00-21:59, while 10.8% start during 22:00-22:59) and 44.9% of sessions start during 18:00 to 00:59.

Figure 2.26 shows the number of participants who practiced (active participants) for each day of the four weeks of the study. The number of active participants shows a quick drop after the first 4 days, displays a peak on the 15th day, and then drops again to low values until the end of the study. The mean of active participants' practice time during each day of the study ranged from 77.14 to 278.08 seconds $(M = 163.92, SD = 53.68)$.

Figure 2.27 shows for how many days participants practiced with the app through a frequency distribution. Participants who used the application for most of the days are rare: the distribution reaches the highest values between 1 and 6 days of practice.

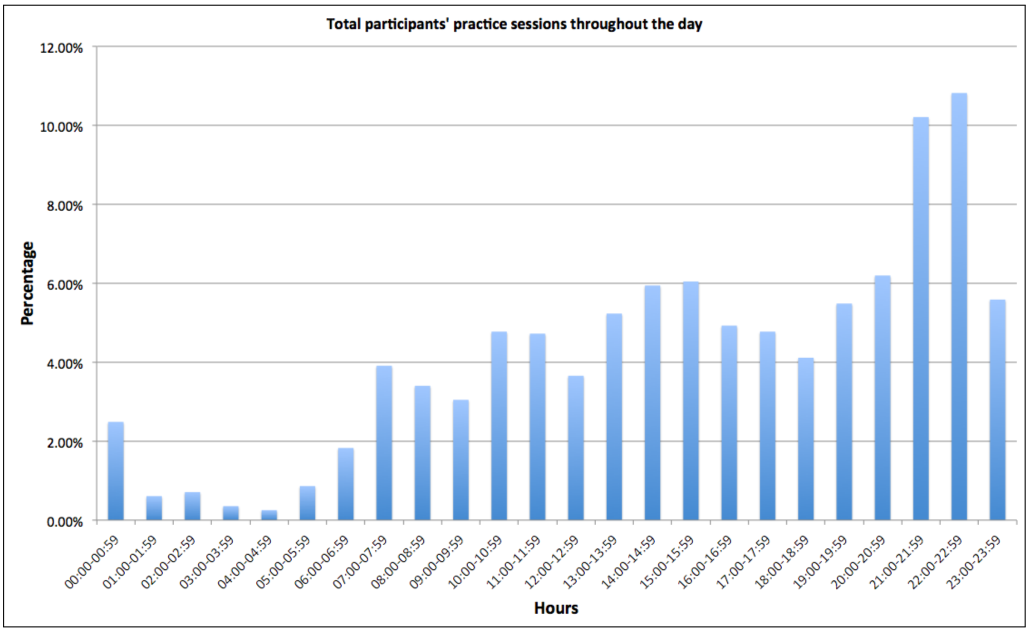


Figure 2.25: Percentage of practice sessions throughout the day, considering all participants.

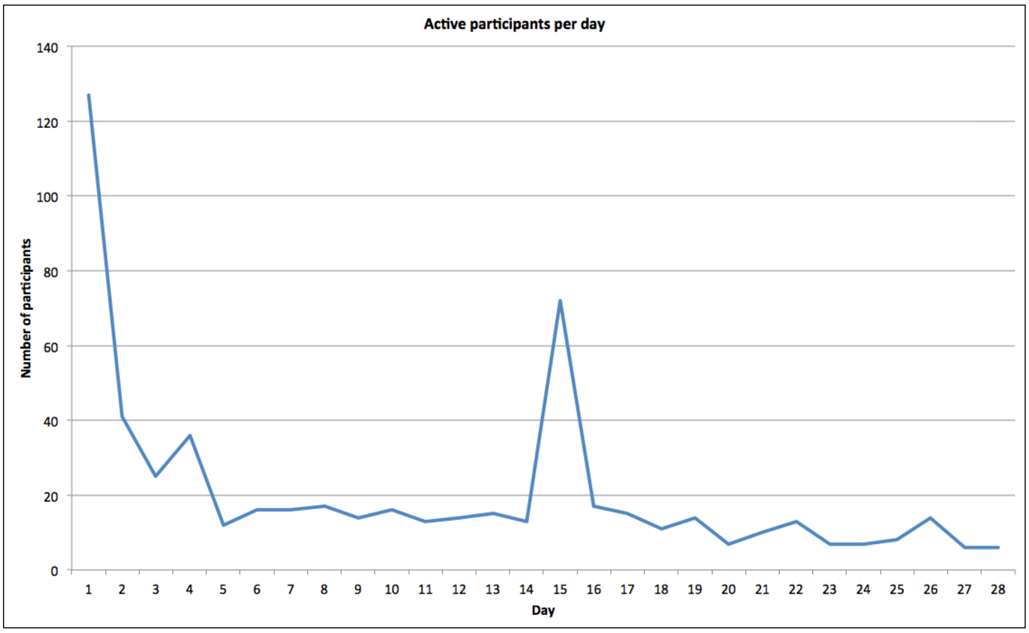


Figure 2.26: Active participants per day.

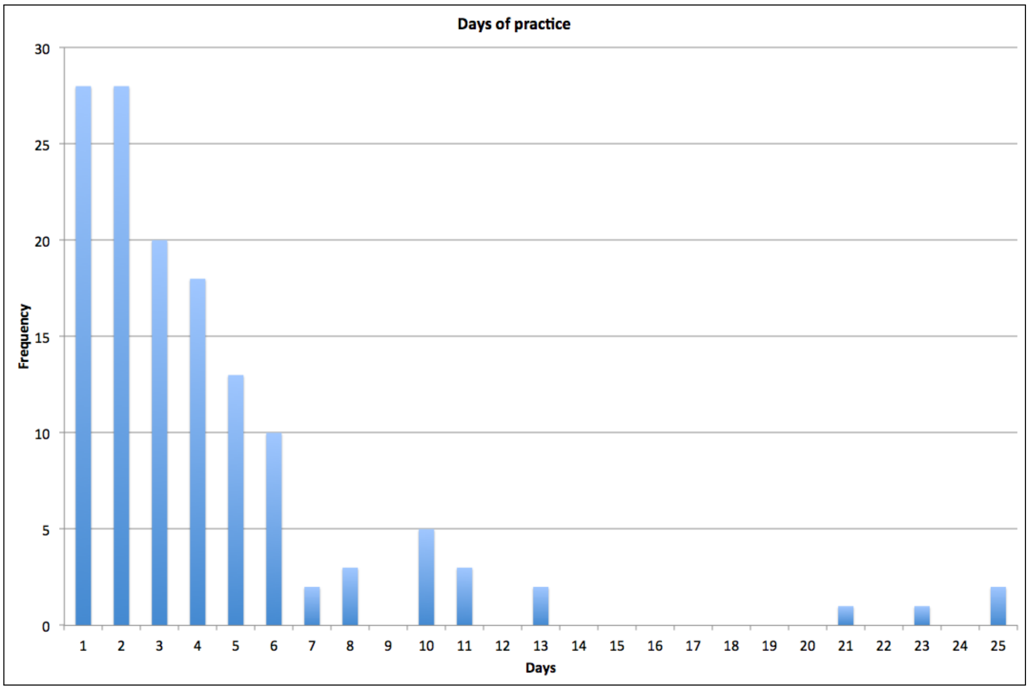


Figure 2.27: Frequency distribution of participants' days of practice.

Table 2.10 shows participants' total number of thoughts. The most frequent case is the 1-5 range.

Table 2.10: Total number of thoughts and their frequency.

Number of thoughts	Number of participants	Percentage
1-5	97	71.3%
6-10	21	15.4%
11-15	6	4.4%
16-20	4	2.9%
21-25	2	1.5%
26-30	2	1.5%
31-35	2	1.5%
41-45	1	0.7%
96-100	1	0.7%

2.5.5.2 Decentering

Figure 2.28 shows the mean level of decentering of naive and experienced meditators on the three assessment points. The scores were analyzed with a two-way mixed design ANOVA with Greenhouse-Geisser correction. The within-subject variable was assessment point (START, 2WEEKS, 4 WEEKS) and the between-subject variable was experience with meditation (naive meditator, experienced meditator). The analysis pointed out a significant main effect for both experience with meditation ($F(1,134) = 9.10, p < .01, \eta_G^2 = .05$), with experienced meditators achieving higher levels of decentering than naive ones, and for assessment point ($F(2,268) = 12.80, p < .001, \eta_G^2 = .02$), and the interaction did not attain significance ($F(2,268) = 1.64, p = .20$)). The main effect of assessment point was then investigated by carrying out t-test pairwise comparisons with Bonferroni correction. For naive meditators, the post-hoc analysis revealed a significant difference ($p < .001$) between START ($M = 31.14, SD = 5.98$) and 2WEEKS ($M = 33.94, SD = 6.26$), and a significant difference ($p < .001$) between START and 4WEEKS ($M = 35.67, SD = 7.23$): the level of decentering of naive meditators was higher after two as well as four weeks with respect to the beginning of the study. A significant difference ($p < .01$) was also found between 2WEEKS and 4WEEKS: the level of decentering of naive meditators was higher at the end of the study with respect to the 2WEEKS assessment. For experienced meditators, the post-hoc analysis found instead no significant differences in any pair of conditions.

The level of decentering increased from START to 2WEEKS for 61.8% of participants (64.2% of naive meditators, 43.8% of experienced meditators), from START to 4WEEKS for 70.6% of participants (71.7% of naive meditators, 62.5% of experienced meditators), and from 2WEEKS to 4WEEKS for 58.8% of participants (57.5% of naive meditators, 68.8% of experienced meditators).

We tested if the changes in the level of decentering correlated with days of practice, practice time and practice sessions, finding no significant correlations.

2.5.5.3 Qualitative Feedback

Among the 136 participants who practiced distancing from thoughts with the app, 62 answered also the optional qualitative questionnaire. The qualitative questionnaires from three participants had to be discarded, because two of them contained only random characters, while the answers in another one were in Swedish. In the following, we briefly summarize the answers to each item of the questionnaire provided by these 59 participants (50 naive meditators, 9 experienced meditators).

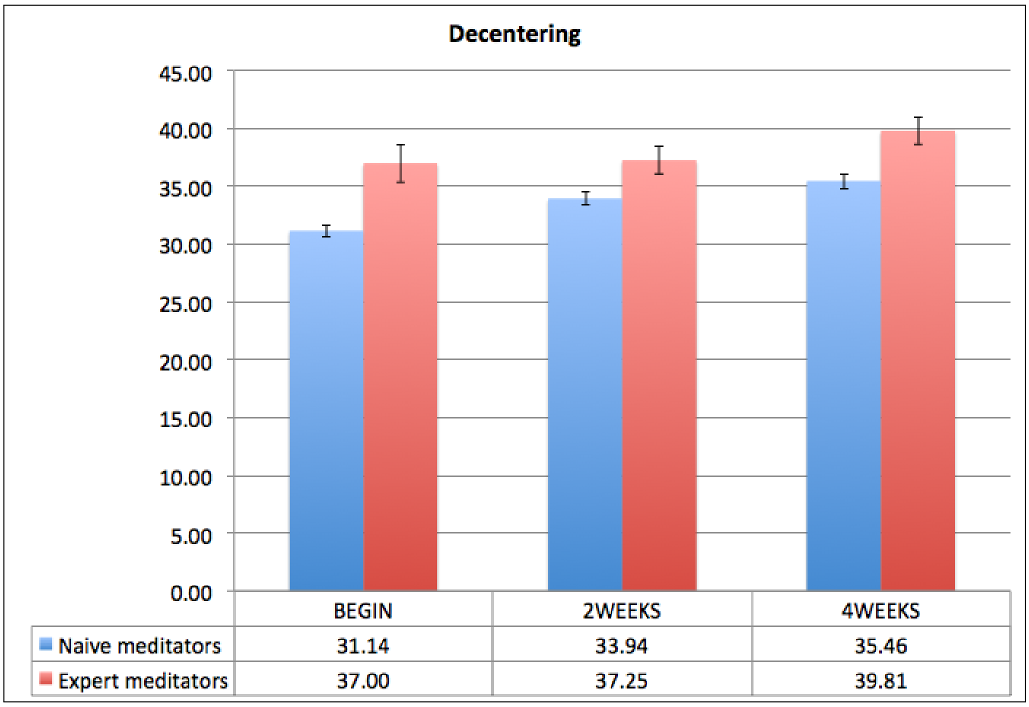


Figure 2.28: Mean level of decentering (capped bars indicate $\pm 1SE$).

First question

Almost half of the participants (47.5%) who answered the qualitative questionnaire pointed out that, when using the app, they thought it was helpful. In particular, 28.8% of participants thought that the app was useful to distance themselves from their thoughts or worries or see them as external objects. Interestingly, one participant remarked that the app helped him realize that he was not his thoughts.

A group of participants (20.3%) pointed also out that they liked the app and enjoyed using it. Another group (13.6%) mentioned instead confusion about how to use the app properly or skepticism about its usefulness.

Finally, 11.9% of the respondents said that they focused on their thoughts or worries when using the app, 6.8% said that they did not think about anything special, and one said he tried not to think at all.

Second question

The usage of AEON elicited positive feelings in the majority of participants (84.7%) who answered the qualitative questionnaire, with relaxation and peacefulness as the most mentioned positive feelings (49.2%), followed by a sense of well-being (23.7%).

A few respondents (5.1%) pointed out that they did not experience any particular feeling when using the app, while 6.8% of participants reported also in this question that they were confused about how to use the app or skeptical about its effectiveness.

Finally, for 5.1% of respondents, using AEON to face their thoughts sometimes elicited negative feelings such as sadness.

Third question

The majority (76.3%) of participants who answered the qualitative questionnaire pointed out that they experienced something new during the period of the study. In particular, a group of respondents (15.3%) said that they were more mindful or aware of themselves, another group (23.7%) reported that they were better able to handle or distance themselves from their thoughts, including negative ones, and one participant said both things. Other participants referred to a sense of calmness (10.2%) or well-being (8.5%). However, 13.6% of respondents were not sure whether such new feelings were due to the use of AEON or external factors.

Finally, 23.7% of respondents said that they did not experience anything new during the period of the study. For three participants, this could be due to the fact that they barely used the app.

Fourth question

The majority (81.4%) of participants who answered the qualitative questionnaire pointed out that during the period of the study they related to their thoughts, worries or feelings differently. In particular, they stated that the app was helpful for (i) seeing their worries as more external objects (25.4%), (ii) not reacting in response to their thoughts (10.2%), and (iii) distancing from their thoughts (8.5%). For 18.6% of respondents, these different perspectives toward their thoughts, worries or feelings were experienced only slightly or only sometimes.

Finally, 16.9% of respondents did not relate to their thoughts, worries or feelings differently during the study period.

Fifth question

The majority (66.1%) of participants provided positive opinions about the app. In particular, 40.7% of respondents said they liked the app and 28.8% said it was useful. Other respondents (8.5%) explicitly mentioned they were going to keep using the app in the future.

The negative opinions (11.9% of respondents) mentioned instead that the app was not easy or clear to use, or that it was too simple to be useful.

Sixth question

Participants who answered the qualitative questionnaire provided some suggestions for improvement. The most frequent one, provided by 11.9% of respondents, is the possibility to receive more instructions to properly practice distancing from thoughts or advice on the thoughts to enter into the app, such as a list of common thoughts that people might have. Another suggestion, given by 10.2% of respondents, was to extend the app with more interactive practices to support other mindfulness exercises in addition to distancing from thoughts. For 5.1% of respondents, the interactive practice should also offer background music or ambient sounds.

Finally, other suggestions concerned the possibility to receive reminders to use the app (6.8%) or usage statistics (5.1%).

2.5.6 Discussion

2.5.6.1 Demographics and Usage Data

The data collected by the app point out that 2997 out of the 3979 users who downloaded the app accepted to take part into the study. This loss of potential participants (24.7%) could be explained by the fact that users are not accustomed to participate in a study by downloading and using research apps (Ferreira et al., 2012) and, in our case, they had no option to opt-out from the study but quitting the app after reading the "Accept" screen. The fact that these participants downloaded and run the app suggests that they were interested in trying it. However, when informed that trying the app required to participate in a study and that their questionnaire and usage data was going to be sent to the researchers, they changed their mind. The percentage of users who accepted the conditions (75.3%) is in line with a previous study that employed research in the large and adopted the same mechanism to ask for user consent (Henze et al., 2011a), obtaining an acceptance rate of 81.3%. Other studies report varying percentages; acceptance was 87.6% in (Henze et al., 2011b), while it was 25.8% in (McMillan et al., 2010). The low acceptance rate of (McMillan et al., 2010) could be due to the fact that their app collected also users' location data, suggesting that the more potential participants feel researchers are able to access their personal data, the more they are likely to change their mind about trying the app.

Most users who accepted the conditions of the study answered also the initial questionnaire and the first mindfulness questionnaire. The few who did not answer such questionnaires (106 users abandoned the initial questionnaire, and 74 abandoned the first mindfulness questionnaire) could have become concerned about disclosing such personal data or not interested in assessing their level of mindfulness, finding it tedious to complete the corresponding 11-items questionnaire.

Focusing our attention on the 2997 participants who accepted the conditions to take part in the study, there are four things interesting to note. First, 41.3% of them answered the first mindfulness questionnaire, but never used the interactive practice and did not answer the second mindfulness questionnaire. Such users could have changed their mind soon after answering the first mindfulness questionnaire or expected a different content from the app.

Second, only 12.9% answered also the mindfulness questionnaire after two weeks, and only 4.9% answered also the same questionnaire after four weeks. Although this percentage is small, it is in line with previous research in the large conducted over a 4-weeks period (Sahami Shirazi et al., 2011), in which only 5.0% of participants answered the questionnaire at the end of the study. Evidence of the fact that research apps are used only for a short period of time, and thus longer studies see large dropout percentages, emerges also from other studies that employed research in the large, e.g. (Ferreira et al., 2012; Henze et al., 2011a; Miluzzo et al., 2010). As suggested by Miluzzo et al. (2010), this could due to the fact that users are not accustomed to use research apps and need to perceive they are receiving clear benefits from using them, otherwise they use such apps rarely or for a short period of time. Moreover, according to recent market research (Consumer Health Information Corporation, 2011), 26% of health apps in on-line stores are downloaded and used only once, and 74% of health app users drop out by the 10th app use.

In our specific case, those who practiced with the app but dropped out might have expected to obtain immediate benefits in their everyday life. Unfortunately, mindfulness and decentering require time and regular practice to be developed and enhanced (Kabat-Zinn, 1990, 2003).

Third, restricting attention to participants who answered the second and third mindfulness questionnaire, respectively 38 (3.1%) and 11 (0.9%) of those participants never practiced with the app. Such participants could have been willing to practice but never found the time to do it. Alternatively, they could have been caught up in the general tendency of people to escape or avoid contact with their internal experience, such as thoughts or worries (Kabat-Zinn, 1990, 2005). This could have led them to procrastinate the practice of distancing from thoughts, and once they reached the end of the study, they could have been only curious to see the new app features we promised to unlock.

Finally, only 2.2% of participants who accepted the conditions to take part in the study answered also the qualitative questionnaire. This result confirms the fact that with research in the large it could be difficult to obtain qualitative feedback from users (Ferreira et al., 2012; Henze et al., 2011a). On a positive side, most qualitative questionnaires we received from participants who practiced with the app (i.e., 59 out of 62) were carefully answered and useful to the research.

The previously cited considerations on the brief use of health and research apps are confirmed by the usage data collected in our study. Indeed, the majority of participants used the interactive practice for a total number of days that ranged from 1 to 6 (see Figure 2.27), and the number of active participants per day shows a quick drop after the first 4 days (see Figure 2.26). The sudden peak in the number of active participants toward the 15th day of the study is likely due to the notification on the 14th day with which the app asked participants to answer the second mindfulness questionnaire, which could have attracted attention towards the app, prompting them to use it. This suggests that notifications might contribute to make the app more used by some participants, as remarked also by some respondents to the qualitative questionnaire.

The analysis of when participants' practiced during the day pointed out that a high number of sessions (44.9%) occurred during evening hours, from 18:00 to 00:59 (see Figure 2.25). According to mindfulness teachers, e.g. (Gunaratana, 2002; Kabat-Zinn, 1990), a good time to meditate is during moments of relative peace and quiet, as it requires a sustained attention. Thus, although the app gave no instructions about when to use it, many participants could have naturally found that the evening was a peaceful and quiet moment for them to practice thought distancing with the app. Moreover, this result is in line with the qualitative long-term study of AEON described in Section 2.4.

The collected data about device language (Table 2.9) highlights that AEON was used by participants on devices with 16 different languages. Although one cannot know for sure that the language participants set in their smartphone is their mother language, this result suggests anyway that AEON was used by people from different cultural contexts, thus increasing the external validity of the study (Henze and Pielot, 2013).

Finally, the app was used by devices with 17 different screen resolutions. Also this factor relates to the ecological validity of the study (Cramer et al., 2010): unlike a laboratory evaluation in which all participants use the app with the same device that is not theirs, in this research in the large the app was used in more naturalistic settings that included the participants' own mobile device.

2.5.6.2 Decentering

The analysis of mindfulness questionnaire data points out that participants' level of decentering increased during the study, and that the level of decentering was higher for experienced meditators than for naive meditators.

The difference between naive and experienced meditators is quite predictable: experienced meditators are likely to have already cultivated a good level of decentering before using the app. Indeed, as pointed out by Shapiro et al. (2006), if an objective shift in perspective toward the internal and external experience normally occurs in individuals' development and growth across the lifespan, i.e. a basic level of decentering, a regular mindfulness practice continues and accelerates this shift, thus resulting in higher levels of decentering. Our result is also consistent with previous studies that compared individuals with different meditation experience, e.g. (Lykins and Baer, 2009).

The analysis of decentering over time in our study showed that for naive meditators the level of decentering significantly increased after two and four weeks of app use with respect to the beginning of the study, and also between two and four weeks. A possible explanation of this result is the fact that AEON provides users with an external visualization of their thoughts and of their disappearing. Such visualization could have helped naive meditators develop the ability to see their thoughts as external and temporary objects rather than inherent aspects of the self or as reflections of reality. This perspective on the internal experience, also referred to as *reperceiving* or *metacognitive awareness*, is fundamental to cultivate a decentering state (Shapiro et al., 2006; Teasdale et al., 2002; Wells, 2006).

For experienced meditators, the post-hoc analysis did not point out any statistically significant differences in level of decentering across the three assessment points. With their previous practice of mindfulness techniques, experienced meditators could have already developed the ability of distancing from thoughts. Moreover, if their previous practice was regular, they could also have developed mindfulness at the trait level (Carmody and Baer, 2008), i.e. the capacity of being mindful in their everyday life (Thompson and Waltz, 2007). As any trait-like quality, trait mindfulness tends to be stable over time (Brown et al., 2007). However, it is interesting to note that the trend in the level of decentering over time shown in Figure 2.28 is positive also for experienced meditators. Therefore, the pessimistic scenario of a detrimental effect on experienced meditators that we had considered as a possibility in Section 2.5.1 was not found to occur.

Overall, these results suggest that the use of AEON can help people, and in particular naive meditators, increase their level of decentering. Moreover, considering that the majority of participants did not practice with the app very frequently, these results might suggest that also a short use of the app can be beneficial to users. Even practicing for a few sessions could have been sufficient for evoking in participants a reflection on the transitory nature of their thoughts. This seems to emerge from the answers provided by respondents to the qualitative questionnaire. Indeed, most of them pointed out that by using AEON they related to their thoughts differently. In particular, some of them explicitly pointed out that the app helped them distance themselves from their thoughts or feelings, including the negative ones. Some participants highlighted also that the app was useful to see their worries as external objects, while others said that AEON helped them not to react in response to their thoughts. For some participants, the app was also useful to be more mindful or aware of their thoughts or in general.

However, we cannot exclude that other factors, different from the interactive practice, contributed to the obtained result. In particular, we should consider that the version of AEON in the study did not provide users only with the interactive practice (as the original AEON did), but included also the mindfulness questionnaire they had to answer three times during the 4 weeks to assess the decentering variable. One could hypothesize that the questionnaire items might also have encouraged in participants a reflection on the nature of their thoughts and contributed to change their perspective on them. To explore this consideration, we analyzed data of the 11 participants (8 naive and 3 experienced meditators) that had to be excluded from the study because they never practiced, Although the small number of participants in the two groups does not allow us to carry out the statistical analyses performed in the study, the trend for these participants looks positive (see Figure 2.29). This suggests that the possible role of the questionnaire is worth investigating further, for example by assigning some users to a version of the app that only allows them to answer the mindfulness questionnaire during the first 4 weeks and makes the interactive practice available only at the end of the fourth week. In any case, the extension of the original AEON app with a mindfulness questionnaire feature could be an option of interest to users.

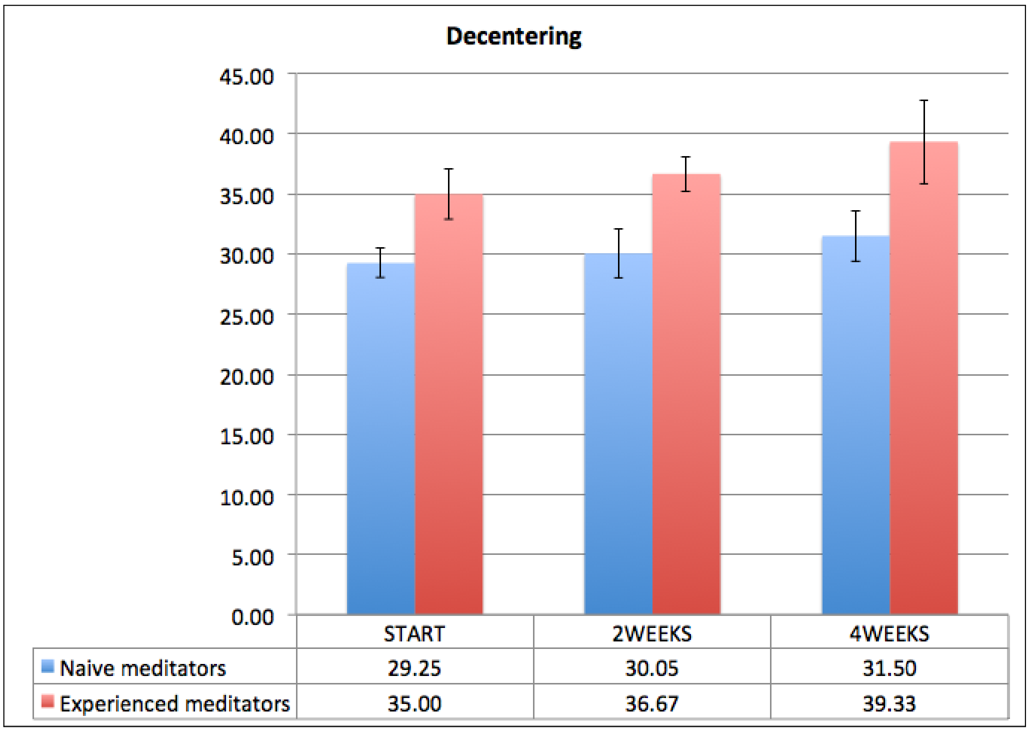


Figure 2.29: Mean level of decentering of the 11 (8 naive, 3 experienced) excluded participants (capped bars indicate $\pm 1 SE$).

Another limitation of the current study, shared with other research in the large studies, is the lack of control on how participants use the app, which can possibly threaten internal validity of the study. However, according to Henze and Pielot (2013) the larger the sample of the study, the more individual differences and contextual factors could be factored out.

2.5.6.3 Qualitative Feedback

In addition to the effectiveness of the app (see previous Section), the feedback provided by participants points out that for the majority of them using the app elicited positive feelings, with relaxation and well-being as the most common ones. Other than the mindfulness practice, a possible factor that could have contributed to this result is the fact that passive or active interactions with nature or natural elements on a computer, e.g. looking at images of nature on a computer screen or exploring a natural 3D environment, can have a restorative effect on people, including stress reduction, relaxation and an overall restoration in energy and well-being, see e.g. (Bates and Marquit, 2011; Berman et al., 2008; Valtchanov et al., 2010). In our case, the interactive practice lets users directly interact with a simulated natural element and see the effect of their actions on it. Another factor that could have contributed to this result is that participants directly acted on the visualization of their thoughts and made them virtually disappear, which could have provided a relieving effect, as pointed out by a few participants.

These considerations could also explain the fact that a high number of participants found the app beautiful. Indeed, the visual stimuli offered by the water simulation could have been perceived as aesthetically pleasant.

3

Design and Evaluation of Mobile Health Apps for Neurological Assessment

In the second case study, we focus on the *Personal Health and Wellness* category (see Section 1.1.2). Among all the possible health behaviors and conditions in this category, we considered a neurological aspect of health, i.e. motor control, which is the process by which people use their neuromuscular system to coordinate and perform motor skills. It is crucial for regulating motion, balance, stability, coordination, and our interaction with the world (Rosenbaum, 2009; Wise and Shadmehr, 2002). Studies in the literature have shown that motor performance changes across the life span, i.e. it becomes better from childhood to young adulthood and then worsen with aging (Leversen et al., 2012; Seidler et al., 2010).

Traditionally, neurologists have used paper-and-pencil tests, e.g. the Trail Making (Reitan, 1958) and the Spiral Drawing (Bain and Findley, 1993) test, or special hardware, e.g. a button or a musical keyboard, to assess patients' motor skills or impairments. Recently, the advent of touchscreen-based mobile devices such as smartphones and tablets offers neurologists new, cheaper and more precise methods to carry out motor tests. In addition, mobile devices could allow their users to easily perform these motor tests anywhere without the need of going to a hospital or lab. However, the only two approaches based on mobile devices that has been proposed in the literature, i.e. (Surangsrirat and Thanawattano, 2012; Westin et al., 2010b), specifically target Parkinson patients and require participants to carry out the motor tests with a stylus pen. Thus, their use is restricted to clinical settings, a specific medical condition, and to stylus-operated mobile devices. On the other hand, implementing motor tests into a mobile app and exploiting the possibilities offered by research in the large make it possible to turn millions of existing mobile devices into assessment and data collection tools.

For this case study, we developed a mobile app (MotorBrain) for the two major mobile platforms (Android and iOS) that implements computerized versions of the Trail Making and Finger-Tapping Tests and offers users a preliminary assessment of their motor performance. For these reasons, according to the classification based on the PPM model (see Section 1.1.1), the app is an *enabling* and *reinforcing app*. In particular, we designed the interface of MotorBrain to make it usable by different kinds of patients as well as healthy individuals and let them interact more naturally and directly with the tests by using the index finger. Moreover, we employed gamification to turn the motor tests into simple games with the aim of attracting a large number of users. Indeed, by exploiting research in the large, the app can send the acquired motor skills data to a remote server, allowing us to build a normative database of individual motor skills data, a goal that is currently difficult and costly to accomplish. Such data can be used for a more detailed analysis of the aging of the population's motor performance. In particular, the analysis can be useful for investigating the physiological aspects that are involved in motor control aging, which have been identified only recently (Cicerale et al., 2014). Finally, the collected data offers neurologists the possibility to study how individuals learn and improve a new motor skill over time. Such information can be used in the diagnosis of movement disorders and rehabilitation process of a pathological motor function. In this way, our proposed app significantly extends the population that can be assessed, makes it much easier for users to perform motor tests even without the presence of a clinician, increases the number of opportunities and the contexts in which they can assess their motor skills, and supports additional purposes other than individual assessment, in particular neurological studies and motor functions rehabilitation.

This chapter is organized as follow: Section 3.1 briefly surveys the methods traditionally used for assessing motor performance and the approaches that employed gamification for large-scale data collection. Section 3.2 presents the first prototype we developed of MotorBrain and Section 3.3 describes the qualitative lab study that we carried out to evaluate users' experience with the app. Section 3.4 illustrates the second prototype of MotorBrain that we developed taking into account the results of the qualitative study. Finally, Section 3.5 reports the quantitative pilot study we conducted to assess whether the data collected by the app can be meaningful for studying human motor performance.

3.1 Related Work

3.1.1 Motor Performance Assessments

Paper-and-pencil tests, such as the Trail Making (Reitan, 1958) and the Spiral Drawing (Bain and Findley, 1993) test, are a traditional method that neurologists use to assess patients' motor skills or impairments. The Trail Making Test requires subjects to connect several targets in a predefined order without taking off the pencil from the sheet. Subjects have to be as fast and accurate as possible in finding and connecting the targets. The Spiral Drawing Test requires instead individuals to draw a spiral from the center with increasingly larger radius. The spiral is then compared to a set of previously rated spirals displaying varying degrees of motor impairments.

Other methods rely instead on special hardware to record users movements. For example, the Finger Tapping Test (Collyer et al., 1994) is a motor test in which subjects have to tap with a finger over a stimulus, e.g. a button or a musical keyboard, following a predefined pace. Subjects have to match the rate of tapping responses to the rate of the pacer stimulus.

The advent of technologies such as digitizing tablets or touchscreens offered neurologists new, cheaper and more precise methods to carry out the above mentioned and other motor tests. Indeed, by connecting these devices to a computer, data about patients' motor skills can be automatically acquired and analyzed, see e.g. (Hubel et al., 2013; Miralles et al., 2006; Pullman, 1998; Rabuffetti et al., 2002; Saunders-Pullman et al., 2008; Tippett and Sergio, 2006). Now, these possibilities can be further extended by possibly exploiting touchscreen-based mobile devices such as smartphones and tablets, which might allow their users to easily perform motor tests anywhere without the need of going to a hospital or lab. The computational capabilities of current smartphones would also allow to process the collected motor skills data and to provide users with an assessment of their motor performance.

A tablet version of the Spiral Drawing Test has been proposed by Surangsrirat and Thanawattano (Surangsrirat and Thanawattano, 2012) on the Android platform, while the Java-based app for Windows Mobile PDAs developed by Westin et al. (Westin et al., 2010b) included also the Finger Tapping Test. However, both mobile apps specifically target Parkinson patients. Moreover, they require participants to carry out the motor tests with a stylus pen. Thus, their use is restricted to clinical settings, a specific medical condition, and to stylus-operated mobile devices.

3.1.2 Gamification for Large-Scale Data Collection

Gamification refers to "the use of game design elements in non-game contexts" (Deterding et al., 2011), and in recent years it has been used in several scientific domains for collecting large amount of data.

The first examples of such use of gamification may be traced back to the so-called "Games With A Purpose" (GWAP), i.e. "games in which the players perform a useful computation as a side effect of enjoyable game play" (von Ahn and Dabbish, 2008). Basically, these games are on-line crowdsourcing systems, i.e. systems that enlist a crowd of users to help solve a problem defined by the system owners (Doan et al., 2011). They use gamification to turn simple tasks into games in order to attract a large number of users. Thus, by letting users play and advance through levels as in entertainment games, they collect data to help accomplish tasks that computers are unable to perform, or that would be much more difficult, time consuming, and in some cases nearly impossible for a lone person or for a small group of individuals to perform (von Ahn and Dabbish, 2008).

The first GWAP was the ESPGame (von Ahn and Dabbish, 2004), i.e. a web-based computer game in which pairs of players challenge each other in guessing the label that explains the content of an image. In this way, by offering an enjoyable experience to users, the ESPGame was able to collect useful data to assign meaningful labels to a large number of images, a task impossible to be carried out automatically by a computer. Other web-based GWAP have been proposed for different purposes in the field of artificial intelligence, e.g. for sound tagging (von Ahn and Dabbish, 2008), building databases of common-consensus knowledge (von Ahn et al., 2006; Orkin and Roy, 2007; Speer et al., 2009) or building and refining ontologies (Siorpaes and Hepp, 2008).

In addition to the above mentioned tasks, GWAP have been successfully employed also for solving difficult scientific problems in the fields of biology and medicine (Good and Su, 2011, 2013; Schrope, 2013). In these contexts, GWAP have also been referred to as “Games With A Scientific Purpose” (GWASP) (Good and Su, 2011). For example, data collected by GWASP was found to be useful for predicting protein structures, e.g. FoldIt (Cooper et al., 2010), improving the accuracy of the multiple sequence alignment problem, e.g. Phylo (Kawrykow et al., 2012), or counting malaria parasites in thick blood smears, e.g. MalariaSpot (Luengo-Oroz et al., 2012).

Finally, GWASP have been used in the field of cognitive science. For example, Dufau et al. (2011) employed a mobile game for a psycholinguistics task. They point out that the app was useful to collect normative data on reading skills across countries, which could be used to assist in the diagnosis and remediation of reading impairments. In a very recent study, Stafford and Dewar (2014) describe a web-based game, i.e. Axon, which tests rapid perception, decision-making and response actions. They point out that the large dataset collected by the game allowed them to test theories of human skill acquisition.

3.2 First Prototype of MotorBrain

The first prototype of MotorBrain contains a total of seven gamified tests that we designed by collaborating with an expert neurophysiologist of the Department of Neurology of the university hospital of Udine. The tests are accessed from a main menu (Figure 3.1). Users can play a game by first tapping on the corresponding menu item and then by choosing the hand with which to play. Users can return to the main menu after each game session.

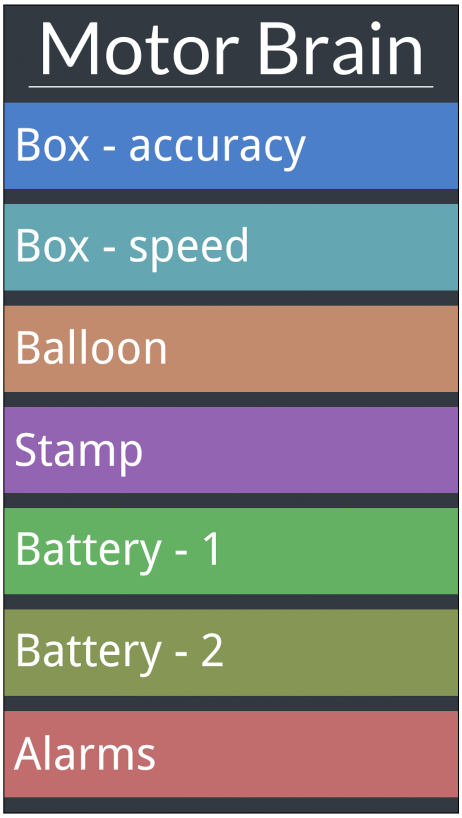


Figure 3.1: Main menu screen of the first prototype of MotorBrain.

3.2.1 The Games

In each gamified test, the screen is divided vertically in two equal parts: the lower one contains the controls that allow users to play the game, while the upper one shows game animations. For each test, the application checks the pixel density of the device display and adapts the resolution of the controls to play the game to make sure they are displayed with the same size across different smartphone models.

In each test, users can play with the index finger of their right or left hand (hereinafter, *playing finger*). In the following, we describe in detail how each motor test has been gamified. The games referred to as BOXA, BOXS, BALL and STAMP are gamifications of Trail Making Tests, while the games BATTERY1, BATTERY2 and ALARMS are gamifications of Finger-Tapping Tests:

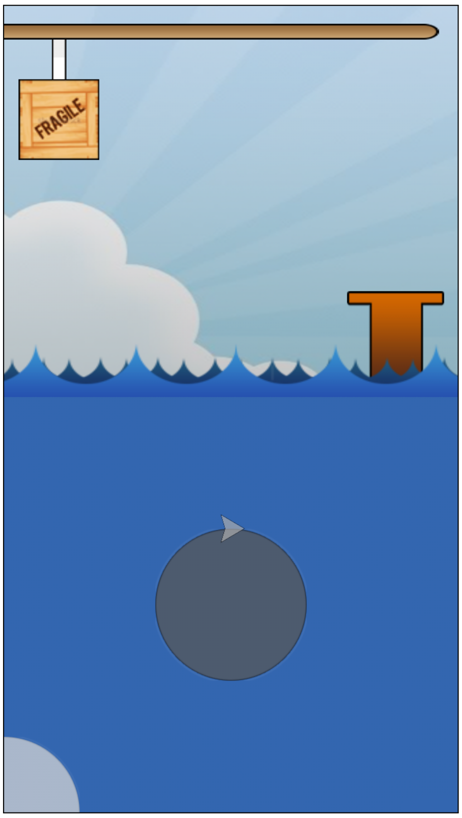


Figure 3.2: BOXA game.

- Box - Accuracy (BOXA). The lower part of the screen contains a circle of 2 cm in diameter, while the upper part initially shows a box on the left attached to a crane suspended over water and a target platform on the right (see Figure 3.2). To play the game, users have to move their playing finger over the entire perimeter of the circle (clockwise or counterclockwise) once and without ever lifting it. The direction of movement is indicated by the game based on whether users are playing with their right or left hand respectively. When the player moves the playing finger over the perimeter of the circle (within a 0.25 mm threshold on both side of the perimeter) in the right direction, the crane shifts the box to the right and a pulley sound is played. Otherwise, if the player moves the finger outside the perimeter or moves it in the wrong direction, the crane stops moving. The game ends when the player has covered a distance equal to the full perimeter of the circle, considering his/her movements over any location of the playing area, or lift the playing finger for more than 0.15 s. In both cases, the box falls down. Thus, the goal of the user is to be as accurate as possible in moving the playing finger over the perimeter of the circle to have the box fall over the base.

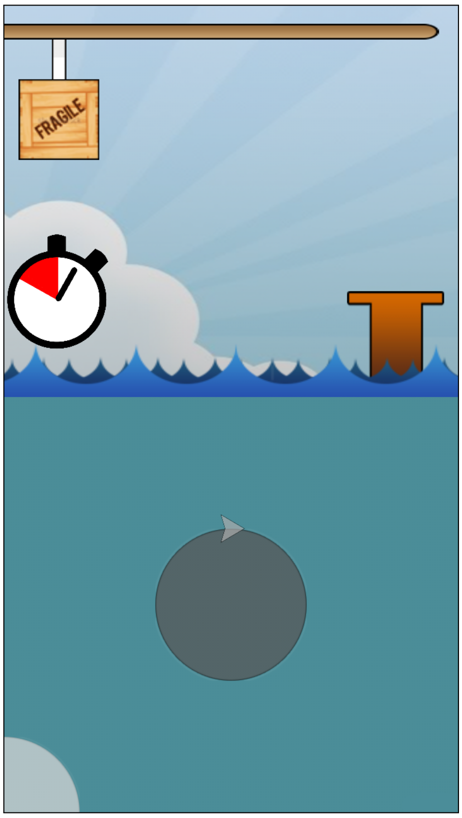


Figure 3.3: BOXS game.

- Box - Speed (BOXS). The graphics of this game are the same as the BOXA game, but the way the crane is controlled by the player is different. Players have to follow the entire perimeter of the circle with their playing finger (within a 0.50 mm threshold on both side of the perimeter) as many times as possible over a 30-second interval to move the crane (a minimum of 40 turns over the perimeter is necessary to reach the platform). The remaining time is indicated by a chronometer shown in the left part of the screen (see Figure 3.3) and four time-ticking sounds are played during the last 5 seconds. When time expires, a gong sound is played and the box falls, ending the game. As in BOXA, the box falls also when players lift the playing finger for more than 0.15 s.

- Balloon (BALL). The lower part of the screen contains a 4.5 cm sloping line (see Figure 3.4) that goes from the bottom-left to the middle-right border of the screen (if the user is playing with the right hand) or from the bottom-right to the middle-left border (if the user is playing with the left hand). The upper part of the screen shows instead a deflated balloon attached to a pump. The balloon inflates only when players move their playing finger over the line (within a 0.25 mm threshold on both sides of the line), from the bottom end to the upper end and back for 10 times. Each time they reach one of the endpoints an inflating sound is played. When players have covered a distance equal to 10 times the length of the line, considering their movements over any location of the playing area, or lift the playing finger for more than 0.15 s, an animation shows either the balloon flying out of the screen (if it has been sufficiently inflated) or deflating, and a related sound is played. In both cases, the game ends. The goal of the user is to be as accurate as possible in moving the finger over the line to let the balloon fly. Moreover, (s)he is also told at the beginning of the game to be fast.

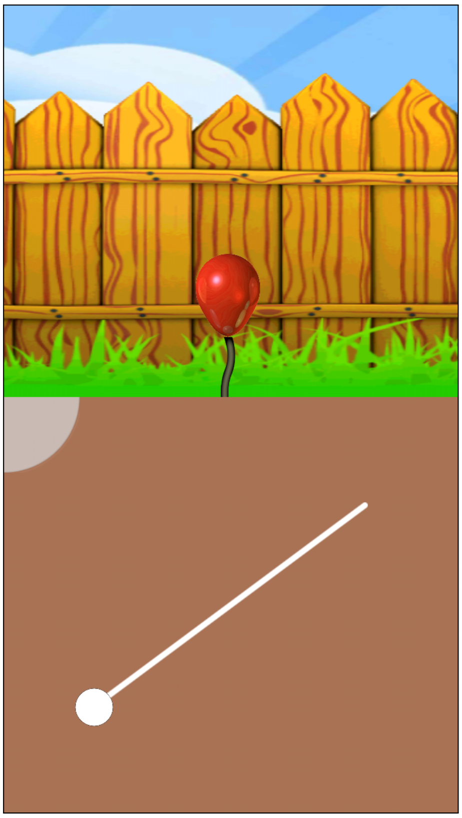


Figure 3.4: BALL game.

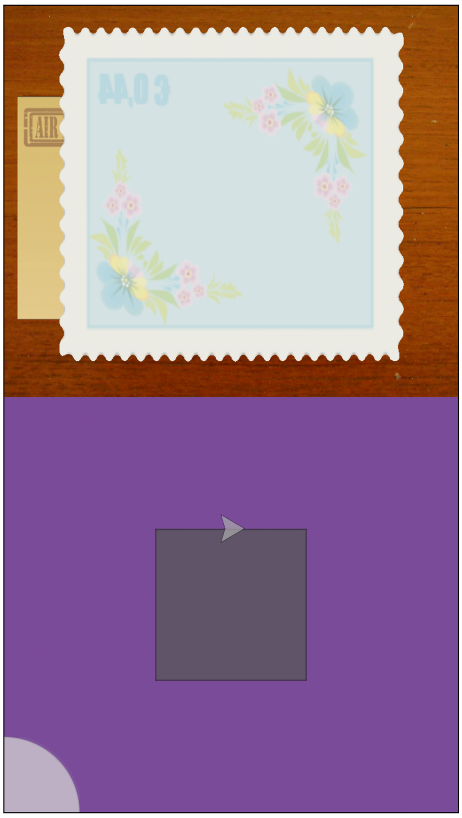


Figure 3.5: STAMP game.

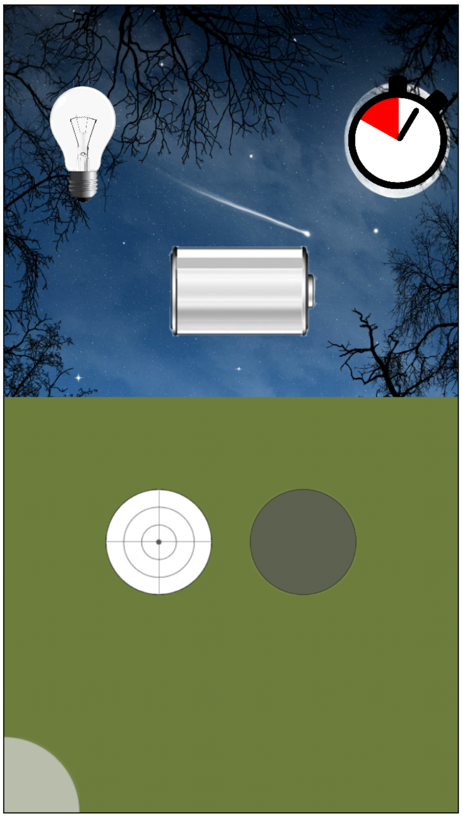
- Stamp (STAMP). The lower part of the screen contains a square with 2 cm sides, while the upper part shows the back of a postage stamp and an envelope placed over a table in the background (see Figure 3.5). To play the game, users have to move their playing finger over the perimeter of the square (within a 0.25 mm threshold on both side of the perimeter) once, without ever lifting it. As in BOXA, the direction of movement depends on the hand with which the user is playing. Moving the playing finger on the perimeter of the square in the right direction has the effect of applying glue on the back part of the stamp and a related sound effect is played. When players have covered a distance equal to the full perimeter of the square, considering their movements over any location of the playing area, or lift the playing finger for more than 0.15 s, an animation shows the stamp rolling over and leaning on the envelope. Then, if enough glue has been applied, the animation shows a postmark being printed on the stamp, which is in this way attached to the envelope. Otherwise, a blow of wind makes the stamp fly away and go out of screen. In both cases, related sounds are played and the game ends. Thus, the goal of the user is to be as accurate as possible in moving the playing finger over the square to attach the stamp to the envelope.

- Battery - 1 (BATTERY1). The upper part of the screen shows a battery and a light bulb, while the lower part shows two round buttons of 1.4 cm in diameter, which are aligned vertically (see Figure 3.6). Only one button is enabled at a time, highlighted by a black viewfinder on a white background. Otherwise, a dark brown background is shown. Players have to tap with the playing finger on the highlighted button. As a feedback, a red circle is shown where the player has tapped on the screen and a clicking sound is played if (s)he has tapped the button. Tapping a highlighted button moves the highlight to the other button and, at the same time, increases the charge of the battery by a small amount that progressively lights up the bulb. Thus, the goal of the player is to tap on as many highlighted buttons as possible in order to fully charge the battery and completely light up the bulb

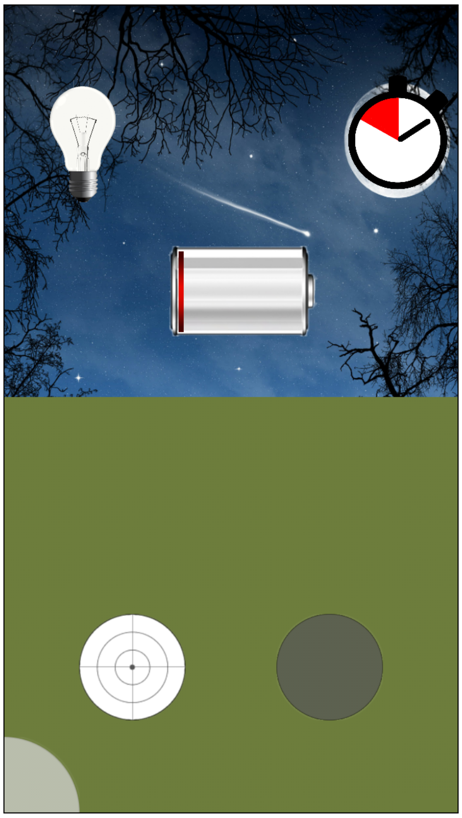
Figure 3.6: BATTERY1 game.

over a 20 seconds interval (a minimum of 38 correct taps is necessary to fully charge the battery). As in BOXS, remaining time is indicated by a chronometer shown in the upper part of the screen. Moreover, players are also told at the beginning of the game to accurately tap on the center of the highlighted button.

- Battery - 2 (BATTERY2). The upper part of the screen is the same as BATTERY1, but the lower part alternatively displays two pairs of buttons, one in a higher and one in a lower position (see Figures 3.7a and 3.7b). All buttons are displayed as those in BATTERY1, and only one at a time is highlighted. Unlike BATTERY1, tapping a highlighted button in one pair moves the highlight to a randomly selected button in the other pair. Unlike the other games, this one requires users to play with two fingers. Indeed, if users are playing with their right hand, they have to tap with the index finger on the highlighted button that appears on the left, while they have to tap with the middle finger on the highlighted button that appears on the right (or vice versa, if they are playing with the left hand). As in BATTERY1, players' goal is to tap on as many highlighted buttons as possible in 20 seconds and to be accurate at the same time (a minimum of 25 correct taps is necessary to fully charge the battery).



(a)



(b)

Figure 3.7: BATTERY2 game. (a) Higher pair of buttons; (b) Lower pair of buttons.

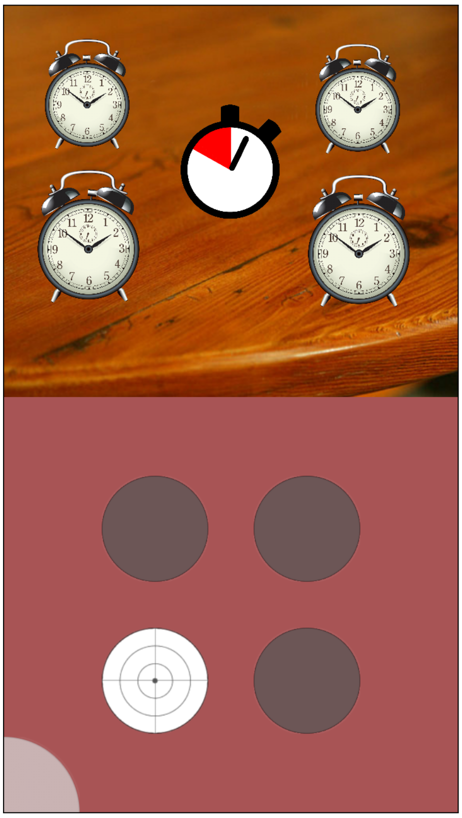
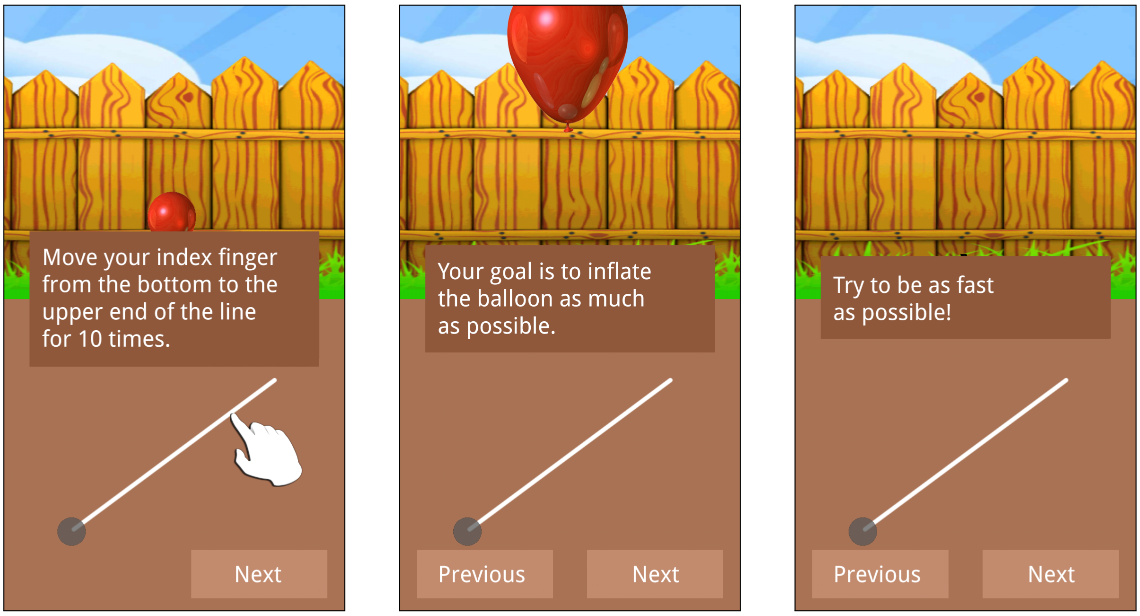


Figure 3.8: ALARMS game.

- Alarms (ALARMS). The upper part of the screen shows four alarm clocks placed over a table (see Figure 3.8), while the lower part contains four buttons, one for each alarm clock, displayed as in BATTERY1 and BATTERY2. Only one button at a time can be highlighted and tapping a highlighted button moves the highlight to another (randomly selected) one. When a button is highlighted, the corresponding alarm clock visibly shakes and a ring sound is played. Thus, the goal of the user is to tap on as many highlighted buttons as possible with the playing finger to turn off the highest number of alarm clocks over a 20 seconds interval. Moreover, players are also told to be accurate by tapping on the center of each highlighted button.

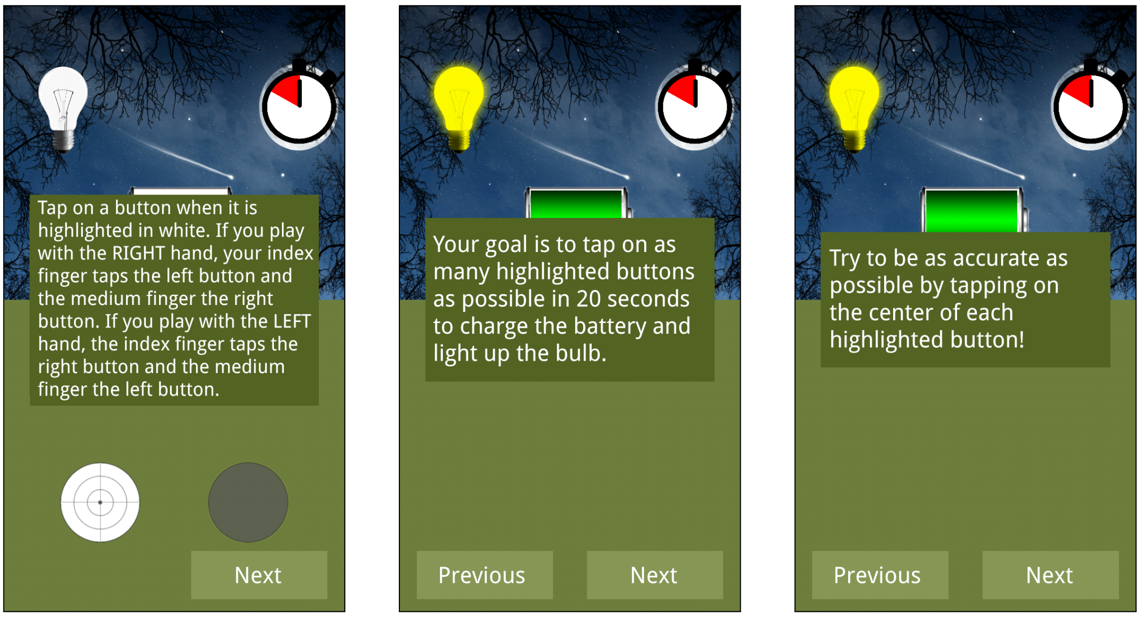
3.2.2 Game Tutorials

When players choose a game in the main menu for the first time, they are presented with a tutorial that explains the game goals and the actions required to play. Each tutorial is organized in three screens (see two full examples in Figures 3.9 and 3.10). Users can freely navigate back and forth through the screens. They can exit the tutorial by pressing the "Next" button on the third screen.



(a) (b) (c)

Figure 3.9: Tutorial of the BALL game. (a) First screen; (b) Second screen; (c) Third screen.



(a) (b) (c)

Figure 3.10: Tutorial of the BATTERY2 game. (a) First screen; (b) Second screen; (c) Third screen.

At the end of any tutorial, the “Hand choice” screen is showed, and the user has to select the playing hand or can start the tutorial again (see Figure 3.11). This screen is always shown by the app when the user chooses a game from the main menu, also when the tutorial is not automatically shown because the user has already played before that game.

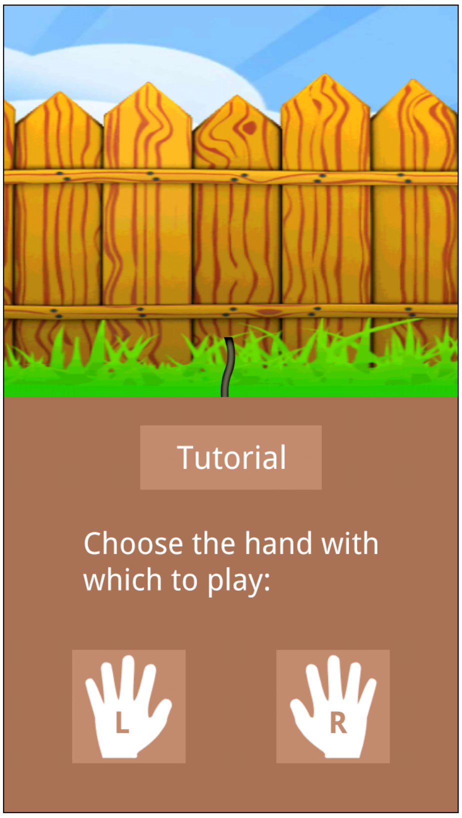
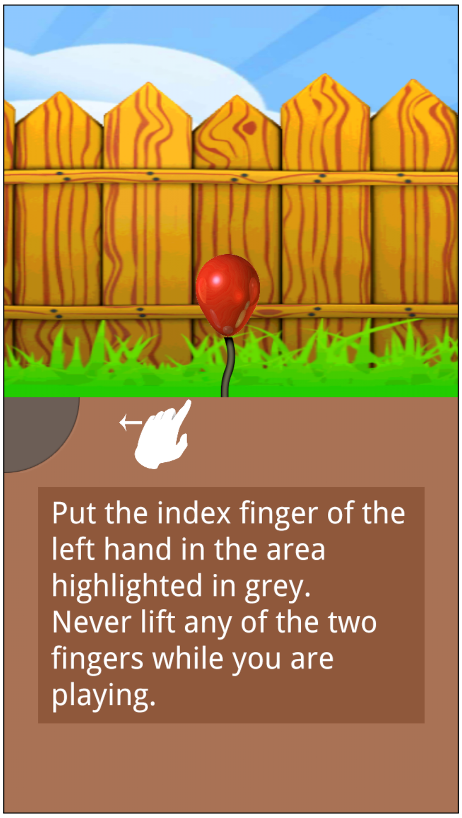


Figure 3.11: “Hand choice” screen in the BALL game.

After players have chosen their playing hand, the app shows a screen that asks them to place the index finger of the other hand (hereinafter, *non-playing finger* and *non-playing hand*) in the bottom-left corner of the screen (if the user is playing with the right hand) or in the bottom-right corner (if the user is playing with the left hand). A colored semicircle is displayed in these areas (hereinafter, *corner areas*) while an arrow points to them (see Figure 3.12 for two examples). The position of the corner areas is the same for all games except BALL, in which they are placed close to the middle-right or middle-left border of the screen (see Figure 3.4 and Figure 3.12a).



(a)

(b)

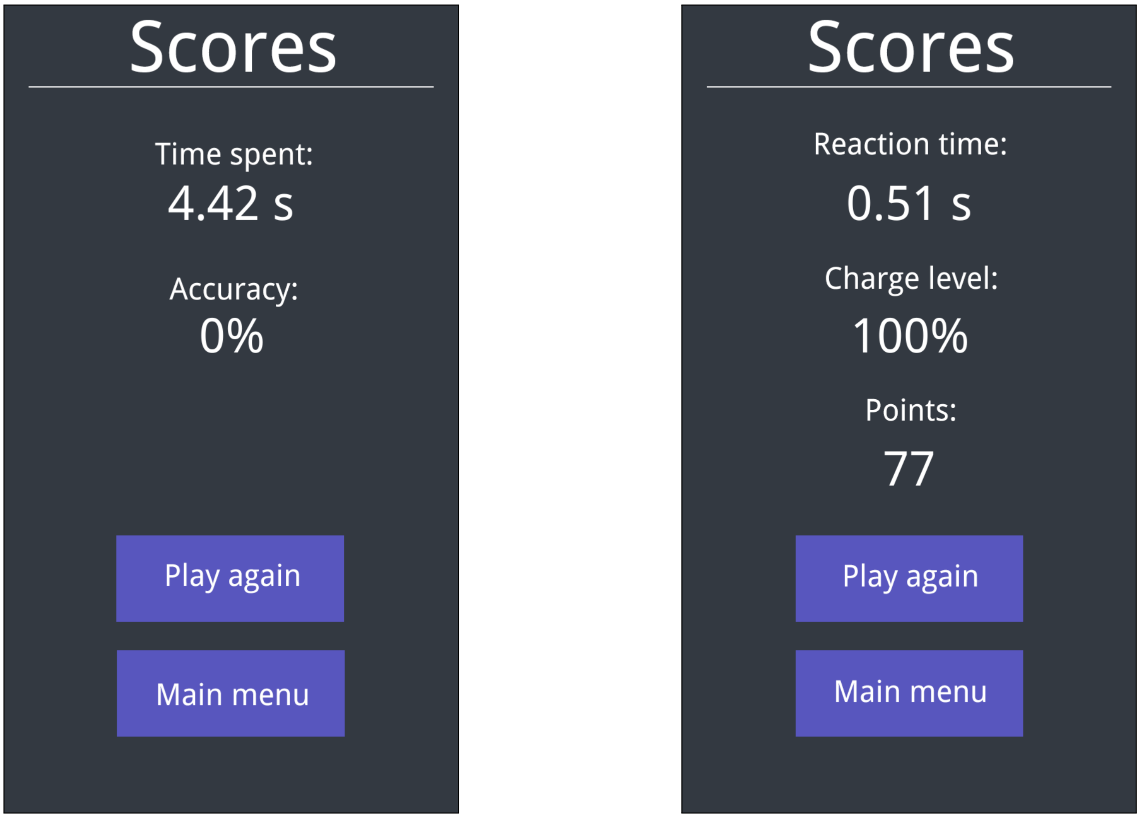
Figure 3.12: Asking players to place the non-playing finger in the corner area for the games (a) BALL and (b) BATTERY2.

Only after the user has placed the non-playing finger in the corresponding corner area, the game controls appear in the lower part of the screen, e.g. the square in STAMP or the buttons in ALARMS, and the timer in BOXS, BATTERY1, BATTERY2 and ALARMS starts. Players have to keep the non-playing finger in the corner area throughout the game. On the contrary, if they lift or move away the non-playing finger from the corner area for more than 1 second, the game ends.

The reason for asking users to place the non-playing finger in the corner area is to prevent the collection of unreliable data. Indeed, since the app will be used by people in their everyday environments, and it is not possible to control how they will actually interact with it (Henze et al., 2011a), this game requirement encourages them to adopt a more stable and controlled posture to play, e.g. placing the phone over a table. Moreover, since some of the proposed gamified motor tests (BATTERY1, BATTERY2 and ALARMS) could be more easily played by using the index fingers of both hands at the same time, this requirement can also prevent users from cheating to obtain a higher score.

3.2.3 Game Scores

When a game ends, the app shows the “Score” screen and two buttons, one to play again and one to go back to the main menu (see Figure 3.13 for two examples).



(a) (b)

Figure 3.13: “Score” screen of the games (a) BALL and (b) BATTERY2.

The displayed scores are the following:

- time spent (BOXA, BALL, STAMP): number of seconds the user spent to play the game;
- accuracy (BOXA, BALL, STAMP): distance covered by the user with the playing finger over the perimeter of the circle in BOXA, over the line in BALL or over the perimeter of the square in STAMP. This score is expressed in percentage with respect to the target distance, i.e. length of perimeter of the circle in BOXA, 10 times the length of the line in BALL and length of perimeter of the square in STAMP;
- number of turns (BOXS): total number of times the user followed the full perimeter of the circle in the BOXS game. This number can be at most 40;
- reaction time (BATTERY1, BATTERY2, ALARMS): mean of the reaction times, i.e. time elapsed between the appearance of a highlighted button and tapping. This score is expressed in seconds;
- charge level (BATTERY1, BATTERY2): total level of charge of the battery in percentage, which corresponds to the number of correct taps made by the user with respect to a maximum of 38 or 25 taps in the BATTERY1 and BATTERY2 games respectively;
- points (BATTERY1, BATTERY2, ALARMS): total points obtained by the player when tapping on highlighted buttons: 5 points are given for a tap on the center area of the viewfinder, 2 points on the

middle area or 1 point on the remaining button area. No penalties are applied when users tap outside the highlighted button area.

All raw data required to determine users' score, i.e. time-stamped pairs of coordinates, are automatically recorded by the app and sent to a server in a secure way at the end of each game.

The app was developed using the Unity game engine (Unity Technologies, 2013) because it supports development for the major mobile platforms, such as Android and iOS, increasing access to the games by a wide sample of users.

The aim of MotorBrain is to collect a large amount of data for building a database of human motor performance. To this purpose, we employed gamification for making traditional motor tests attractive and pleasant to execute for users. However, given the novelty of applying gamification in this context, we first wanted to assess the feasibility of our proposed approach. Then, we wanted to investigate whether the data collected by the app can be effectively used for describing the aging of the population's motor performance. To this purposes, we carried out two evaluations of the app before releasing it in app stores. In particular, we first conducted a qualitative lab study for investigating users' experience with the app. Then, we developed a second prototype of the app considering the problems that emerged from such evaluation. Finally, we carried out a quantitative pilot study for analyzing the data collected by the app.

The following sections describe: (i) the qualitative pilot study with the first prototype of MotorBrain (Section 3.3), (ii) the second prototype of the app that we developed (Section 3.4), and (iii) the quantitative pilot study with such prototype (Section 3.5).

3.3 Qualitative Lab Study

The goal of the qualitative lab study was to evaluate users' experience with MotorBrain. Participants were asked to play MotorBrain in a laboratory setting. Then, they were asked to answer a set of open-ended questions in writing.

3.3.1 Participants

The evaluation involved a sample of 13 participants (11 M, 2 F) recruited among undergraduate human-computer interaction students in return for course credit. The age of participants ranged from 21 to 26 ($M = 22.08, SD = 1.50$). Twelve participants were right-handed and one of them was left-handed. On a self-report scale ranging from 1 (low familiarity) to 7 (high familiarity), participants reported to be very familiar with mobile devices ($M = 5.69, SD = 1.25$).

3.3.2 Materials and Apparatus

MotorBrain was run on a Samsung Galaxy S III with Android 4.3, featuring a 4.8", 720×1280 pixel touch screen. During usage, the device was in portrait mode and placed over a mat on a 72 cm-high table. Participants were seated on a 44 cm-high chair in front of the table.

Before using the app, users filled a short demographic questionnaire (age, sex, dominant hand and familiarity with mobile devices).

The post-play questionnaire for the qualitative study contained:

- for each game, a brief summary and a picture of the game, followed by an open-ended question that asked participants to freely express their opinions on that game. As a partial list of possible non-mandatory examples, the question mentioned highlighting its strengths and weaknesses, if it was easy or difficult to play with it, if the animation in the upper part of the screen was distracting and what they thought about placing the non-playing finger in the corner area;

- two final open-ended questions which asked participants to point out any suggestion for improvement (for specific games or for the app in general) and to express any other comment on those topics not covered by the previous questions.

3.3.3 Procedure

Participants were individually taken to a quiet room, briefed about the nature of the experiment, and asked to fill out the demographic questionnaire.

Then they were asked to play with each game six times: three times with their dominant hand and three times with their non-dominant hand. The order of presentation of the games was varied for each participant.

To provide participants with ample time to fully write all their thoughts, a 10-days limit was given to return the questionnaires. Finally, they were thanked for their participation. Overall, carrying out the procedure took about 35 min per participant.

3.3.4 Data Analysis

All participants returned the qualitative questionnaire. We employed thematic analysis to identify and organize common and salient themes that emerged from participants' responses. To do so, we followed the procedure described in (Braun and Clarke, 2006), adopting an inductive approach as no assumptions were made prior to analysis on the themes that might emerge. For the analysis, the two authors of this thesis first read and re-read the answers to familiarize with the data. Then, we coded emerging features within the dataset and collated data relevant to each code. As a successive step, we combined all the identified codes into potential themes, gathering all data relevant to each potential theme. When a theme was particularly large or complex, it was divided into one or more sub-themes. Then, we checked if the themes and sub-themes were meaningful in relation to the coded extracts and to the entire data set. Finally, we refined each theme and sub-theme, assigning them clear definitions and names.

We reviewed themes and sub-themes following an iterative process that ended when we reached agreement. Then, to assess their validity, we involved an external independent coder. To this purpose, we followed (DeCuir-Gunby et al., 2011) to create a codebook in which the codes were the themes and sub-themes pointed out by the thematic analysis. For each code, we provided a label, a full definition, i.e. an extensive definition that provides inclusion and exclusion criteria, and an example extracted from the data set. The coder was briefly introduced to the app and was told that he could assign two or more codes to the same fragment of text. Moreover, as the coding process was carried out with a Computer Assisted Qualitative Data Analysis Software (CAQDAS), the independent coder was instructed to use a specific tool, i.e. Coding Analysis Toolkit (CAT) (Texifter, 2007), an open-source web-based CADQAS. In total, there were two coders in the system, as we counted as one.

3.3.5 Results

We analyzed the coded dataset using Fleiss' kappa (Fleiss, 1971) to assess the level of agreement between coders. Results pointed out an overall kappa coefficient of .65, which indicates substantial agreement (Landis and Koch, 1977).

In the following, we present the results of the thematic analysis, organizing them into the four identified topic areas: (i) "Graphics", which contains the themes that concern participants' comments on the graphics or animations of the games, (ii) "Game features", which contains the themes related to participants' experience with the games, (iii) "Ergonomic aspects", which contains the themes that describe participants' comments on the actions and posture required to play, and (iv) "Suggestions", which contains the themes that concern participants' suggestions for improvement.

In the following sections, we describe each theme providing sample extracts from the interviews. In the extracts, the parts in square brackets are words or phrases added to clarify the meaning of the sentence.

3.3.5.1 Graphics

Table 3.1 summarizes the themes that belong to the “Graphics” topic area. For each theme, the table indicates the games to which the theme applies.

Table 3.1: Themes of the topic area “Graphics”. The last column indicates the games to which each theme applies.

Theme	Sub-theme	Description	Games
Beauty and intuitiveness of the graphics		Remarks about the beauty and intuitiveness of the graphics and animations of the games.	
	Beautiful and intuitive graphics	The graphics and animations of a specific game are beautiful and intuitive. In some cases, they help players to concentrate during the game.	BOXA, BOXS, BALL, STAMP, BATTERY1, BATTERY2, ALARMS
	Not beautiful or not intuitive graphics	The graphics and animations of a specific game are not beautiful or not intuitive.	BOXA, BOXS, BALL, STAMP, BATTERY1, BATTERY2, ALARMS
Chronometer aspects		Remarks about the chronometer included in some of the games.	
	Useful chronometer	The chronometer is useful to check the available time.	BOXS
	Confusing chronometer	The chronometer is not clear and sometimes confusing.	BOXS, BATTERY1, BATTERY2, ALARMS
Animation aspects		Remarks about the animations displayed in the upper part of the screen.	
	Useful animation	The animation in the upper part of the screen of a specific game is useful to the purpose of the game.	BOXA, BOXS, BALL, STAMP
	Animation does not attract attention	The animation in the upper part of the screen of a specific game does not attract participants’ attention when they are playing.	BOXA, BOXS, BATTERY1, BATTERY2, ALARMS
	Distracting animation	The animation in the upper part of the screen of a specific game distracts participants when they are playing.	BOXA, BOXS, BALL, STAMP, BATTERY1, BATTERY2, ALARMS

The majority of participants provided comments on the graphics of all the games. Nine of them pointed out positive aspects, which referred to their simplicity, intuitiveness, and the fact that it helped them to remain concentrated during the game. On the contrary, six participants reported that the graphics of a specific game were not beautiful and that it could be improved.
Participants considered also to the animations in the upper part of the screen. In particular, for five of them, the animation of the BOXA, BOXS, BALL and STAMP games was beautiful and a useful reference for achieving the goal required by the game, for example:

A glance at the box could be useful to be sure of having placed it above the target platform. Thus, looking at the animation helps to reach the goal. - P10 - BOXA

Similarly:

Thanks to the animation, the user can immediately figure out if he has reached the goal, without waiting for the screen with the scores [. . .] - P4 - BALL

On the contrary, seven participants pointed out that the animation of the games was a source of distraction that worsened their performance. For example:

My peripheral vision was continuously distracted by the battery. - P3 - BATTERY1, BATTERY2

Similarly:

I was often distracted by the inflating balloon, because I was interested to see at which point of the game I was. For this reason, I looked away from what was my goal and as a consequence I did not move the finger accurately over the line. - P6 - BALL

For eight participants, the animation of a specific game did not attract attention. For example:

In this case, as the game was a test of speed, the box did not distract me because I could just concentrate on the circle and move the finger very fast around it. - P11 - BOXS

In particular, some of these participants explicitly referred to the scarce usefulness of the upper part of the screen, as P2:

My concentration was totally focused on the lower part of the screen to obtain a better accuracy, making the animation unnecessary. - ALARMS

Finally, three participants provided comments on the chronometer used by some games to indicate remaining time. For two of them, i.e. P4 and P8, the chronometer was very useful to indicate the passing of time and the last available seconds in the BOXS game. However, P8 highlighted also that the first times he played with this game it was unclear to him whether the chronometer hand would make another turn at the end of the first one. Instead, P10 pointed out that he did not understand the purpose of the red part of the chronometer, which indicates the last available seconds (Figure 3.3):

It was not clear whether the red zone in the chronometer was an alert for the time that was passing, or a zone in which the chronometer hand would never go. In this last case, looking at the clock diverts one's attention from the buttons, causing loss of precious time and concentration. - BOXS, BATTERY1, BATTERY2

This participant highlighted also that in the ALARMS game it was difficult to clearly see the chronometer, as it is placed among the four alarm clocks.

3.3.5.2 Game Features

Table 3.2 summarizes the themes and sub-themes that belong to the topic area "Game features". For each theme and sub-theme, the table indicates the games to which they apply.

Table 3.2: Themes and sub-themes of the topic area “Game features”. The last column indicates the games to which each theme or sub-theme applies.

Theme	Sub-theme	Description	Games
Easy and intuitive game		A specific game is intuitive and it is easy to reach its goal.	BOXA, BOXS, BALL, STAMP, BATTERY1, BATTERY2, ALARMS
	Very clear tutorial	A specific game is easy and intuitive thanks to its introductory tutorial.	BOXA, BOXS, BALL, STAMP, BATTERY1, BATTERY2, ALARMS
Unclear game behavior and purpose		The behavior and the purpose of a specific game are unclear.	BOXA, BALL, BATERY1, BATTERY2
	Unclear tutorial	A specific game has unclear behavior and purpose due to its introductory tutorial, which does not give correct or most appropriate information.	BOXA, BOXS, BALL, STAMP, BATTERY1, BATTERY2, ALARMS
Difficult game		Participants find difficult to reach the goal of a specific game.	BOXA, BOXS, BALL, STAMP, BATTERY1, BATTERY2, ALARMS
	Difficult to combine accuracy and speed	A specific game is difficult due to the fact that it requires participants to be both accurate and fast.	BOXS, BALL, BATTERY1, BATTERY2, ALARMS
Compelling and entertaining game		Participants find playing with a specific game compelling and entertaining.	BOXS, BATTERY1, BATTERY2, ALARMS
Boring game		Participants find playing with a specific game boring and not enjoyable.	BOXA, BOXS, BALL, STAMP, BATTERY2

Table 3.2: Themes and sub-themes of the topic area "Game features". The last column indicates the games to which each theme or sub-theme applies.

Theme	**Sub-theme**	**Description**	**Games**
	Too simple game	A specific game is boring because it is too simple to accomplish its goal.	BOXA, BOXS, BALL, STAMP, BATTERY1, BATTERY2, ALARMS
	Repetitive game	A specific game is boring because it requires repetitive actions to play.	BOXA, BOXS, BALL, STAMP, BATTERY1
Scores		Aspects regarding the games' scores.	
	Consistent scores	The score of a specific game is consistent with players' performance.	BOXA, BOXS, BALL, STAMP, BATTERY1, BATTERY2, ALARMS
	Inconsistent scores	The score of a specific game is inconsistent with players' performance.	BOXA, BOXS, BALL, STAMP, BATTERY1, BATTERY2, ALARMS
Sounds		Aspects regarding the games' sounds.	
	Useful sounds	The sounds of a specific game are useful to help participants achieving the goal.	BOXA, ALARMS
	Sounds elicit negative feelings	The sounds of a specific game elicit negative feelings in participants.	BATTERY2, ALARMS
No appropriate feedback		A specific game does not provide appropriate feedback on participants' performance or errors.	BOXA, BOXS, BALL, STAMP, BATTERY1, BATTERY2, ALARMS
Possibility of cheating		Participants point out the possibility of cheating with the BATTERY2 game.	BATTERY2

Participants provided opposite opinions on several game features, i.e. intuitiveness and difficulty of the games, engagement, scores and sounds. Considering intuitiveness, nine participants pointed out that the games were intuitive to play and that they found it easy to accomplish their goal. Four of these participants revealed that the easiness of the games was due to their introductory tutorial that clearly explained the game instructions.

On the contrary, 10 participants highlighted that the purpose and the behavior of specific games were unclear and that it was difficult to achieve the goal of the game. For example, considering the purpose of the games, P8 said:

I had difficulties to understand what to do in this game [. . .] I found the required movement not intuitive. - BALL

Considering the overall behavior of the games, P3 pointed out:

Initially, due to the past experience with the BOXS game, I did not think that I had to move the finger over the perimeter only once to reach the goal. I haven't understood when the box falls yet. - BOXA

In particular, for eight of these 10 participants a specific game has unclear behavior and purpose due to its introductory tutorial, which does not give correct or most appropriate information. For example (see Figure 3.9 for the BALL tutorial):

I found it difficult to play the game. Indeed, by following the explanations of the tutorial, I continued to fail in reaching the goal. I did not understand why, despite the six times I played. It would be necessary to revise the tutorial. - P12 - BALL

Similarly (see Figure 10 for the BATTERY2 tutorial):

The tutorial did not explain in a very clear way how to play. - P1 - BATTERY2

Considering the difficulty to achieve the goal of the games, five participants mentioned the actions required to play, for example:

I found it difficult to keep the finger over the line correctly. Most of the times, this stopped the pump before the balloon was fully inflated. - P5 - BALL

Some of these participants referred instead to the size of the controls with which they could interact with the game:

The circle was too small [2 cm of diameter] considering the overall screen, making it more difficult to reach the goal of the game. - P2 - BOXA, BOXS

Similarly, P11 pointed out:

I found it difficult to accurately tap over a button that has the size [1.4 cm of diameter] of my fingertip. - BATTERY1, BATTERY2, ALARMS

Moreover, five of these seven participants also reported that the difficulty of the games BOXS, BALL, BATTERY1, BATTERY2 and ALARMS was due to the fact that they had to be fast and accurate at the same time, as the former requirement affected the latter. For example:

This game is more difficult than BOXA. The fact that also the speed is important affects a lot the accuracy. It is not easy to move the finger over the circle accurately and quickly at the same time, especially with my non-dominant hand. P10 - BOXS

Similarly:

I tried to be as fast as possible in tapping the buttons without trying to be accurate at the same time, because I thought it was an impossible goal. Indeed, although I slowed down a little bit and tried to tap exactly on the center of the button, I still had difficulties in obtaining a good accuracy. P2 - BATTERY1, BATTERY2, ALARMS

The comments related to engagement reveal that all participants found the games BOXS, BATTERY1, BATTERY2 and ALARMS, fun and entertaining. Their main reasons were the games' temporal limit or, in the case of BATTERY2 and ALARMS, the randomness with which the buttons appeared.

On the contrary, eight participants said that the game was boring, for example P11:

I did not enjoy this game. I would find a more realistic situation to increase the interest toward the game. - BATTERY2

For six of these eight participants, boredom was due to the fact that it was too simple to achieve the goal and that the games did not provide challenges with an increasing level of difficulty. For example, P9 said:

The level of the challenge does not seem particularly compelling and inflating the balloon seems more a "job" than a game. Moreover, the action required is not so complex. - BALL

Similarly:

I find the game boring because I always win and the level of difficulty does not increase. - P3 - BOXA

For seven of these eight participants, the games (except BATTERY2 and ALARMS) were boring because they required repetitive actions to play and the game that suffered most from this problem was BATTERY1. For example, P8 pointed out:

Very boring and predictable. The appearance pattern of the buttons is so simple that it took me 2-3 seconds to realize that after tapping on the left button, the right one appeared.

Considering the aspects related to the scores, three participants highlighted that they were consistent with their performance with the game, while seven participants had an opposite opinion. More specifically, four of these seven participants said that the number of turns displayed in the "Score" screen of the BOXS game was always 40 (this was the number of turns necessary to make the crane reach the target platform). One participant (P3) pointed also out that he was disappointed because the errors he made in the games BATTERY1 and ALARMS were not taken into account for the final scores (if the player tapped outside the highlighted button no points were subtracted to the score). Finally, P13 highlighted instead that, as he tried each game more times, he found it difficult to memorize all the scores and to compare the performance between the two hands. Thus, he suggested to display the scores inside each game and to include in the app a screen that summarizes all the scores for each game.

Taking into account participants' comments on the sounds, for three of them the sounds in BOXA and ALARMS were helpful to concentrate during the game, while for other three participants the sounds in BATTERY2 and ALARMS elicited anxiety. For one of these participants, the sounds had a negative impact on the accuracy:

The alarm clock sound elicited in me a bit of anxiety. Thus I did not try to accurately tap on the button, but I tried to turn off the alarm clock as fast as possible, without looking at the center of the button. - P4 - ALARMS

Finally, other gaming aspects that emerged from participants' comments were related to the feedback provided by the games, as for four of participants it did not accurately reflect their performance or their errors, and, especially for BATTERY2, the possibility to cheat, as four participants pointed out the possibility to use only one finger. Two of these participants considered this a weakness.

3.3.5.3 Ergonomic Aspects

Table 3.3 summarizes the themes that belong to the topic area "Ergonomic aspects". For each theme, the table indicates the games to which they apply.

Table 3.3: Themes of the topic area "Ergonomic aspects". The last column indicates the games to which each theme applies.

Theme	Sub-theme	Description	Games
Corner area		Aspects regarding the corner area.	
	Useful corner area	Placing the non-playing finger in the corner area is a useful action to avoid cheating or moving the smartphone when playing.	BOXA, BOXS, BALL, BATERY, BATTERY2, ALARMS
	Uncomfortable corner area	Placing the non-playing finger in the corner area can be a hindrance to play the game and uncomfortable when the smartphone is not on a table.	BOXA, BOXS, BALL, STAMP, BATTERY1, BATTERY2, ALARMS

Table 3.3: Themes of the topic area "Ergonomic aspects". The last column indicates the games to which each theme applies.

Theme	Sub-theme	Description	Games
	Useless corner area	Participants do not understand the purpose of placing the non-playing finger in the corner area.	BOXA, BOXS, BALL, STAMP
Playing actions		Aspects regarding the actions required to play the games.	
	The playing actions highlight differences between the dominant and non-dominant hand	The actions required to play the games are useful to notice differences between the dominant and non-dominant hand.	BOXA, BOXS, BALL, STAMP, BATTERY1, BATTERY2, ALARMS
	Appropriate playing actions	The actions required to play the games are appropriate to reach the goal.	BOXA, BOXS, BALL, BATTERY1, BATTERY2
	Problematic playing actions	The actions required to play the games could cause problems or are inappropriate to reach the goal.	BOXS, STAMP
	Unsynchronized playing actions	The actions required to play the game are not well synchronized with the animation in the upper part of the screen.	BOXA, BOXS

Almost all participants provided comments on the placement of the non-playing finger in the corner area. In particular, seven participants pointed out that it is a useful action to avoid cheating, for example:

[. . .] it is not a hindrance to play, rather, it prevents cheating. - P12 - BOXA

P13 highlighted that it is useful to prevent the smartphone to move:

[. . .] it allows [the smartphone] to have more stability. - P13 - BOXS

Interestingly, P11 said that this action was useful for not wasting seconds in the games that have a timer, i.e. BOXS, BATTERY1, BATTERY2, ALARMS (in these games, the timer starts when players put the non-playing finger in the corner area).

However, participants provided also negative opinions. Indeed, for 10 of them placing of the non-playing finger in the corner area was uncomfortable. The main reason was that it forces to place the smartphone over a table in order to play, as pointed out by P7:

As in BOXA, the requirement of using both hands to play is a problem, as it forces the user to put the smartphone on a table, which is not always available. - BOXS

Other participants referred instead to the fact that it was a cause of distraction during the game, i.e. for P4, or a hindrance to the movements of the other hand, i.e. for P5 and P6.

Finally, three participants pointed out their skepticism on this action, as they did not understand its motivation in the games BOXA, BOXS, BALL and STAMP, for example:

I'm still not convinced of the usefulness of placing the index finger in the corner [of the screen]. - P3 - STAMP

Participants provided comments also on the gestures required to play the games. Interestingly, two participants said that these gestures were useful for highlighting the differences between their dominant and non-dominant hand. For example, P1 pointed out:

This was the funniest game. It combines business with pleasure, as it involves the user and at the same time it allows to assess the different performance of the right and left hand. Ð ALARMS

For five participants, these gestures were appropriate to reach the goals of the games. For example, P13 said:

The direction of the movement changes with the hand. This feature makes it possible to carry out the task optimally. - BOXS

Considering the BATTERY1 game, P10 said:

The alternation of the allowed the fingers to move very fast, as it was possible to predict and anticipate the next movement. This helps a lot in performing the task quickly.

However, three participants provided contrary opinions, in particular for the gestures required to play the BOXS game. Indeed, they pointed out that moving the finger quickly around the circle was troublesome due to the friction of their fingertip with the screen, or because they moved the smartphone. One of these participants, i.e. P9, pointed also out that the touchscreen is not appropriate for the accuracy task of the STAMP game.

Finally, P1 highlighted that the gestures in the BOXA and BOXS games were unsynchronized with the animation in the upper part of the screen.

3.3.5.4 Suggestions

Table 3.4 summarizes the themes that belong to the topic area “Suggestions”. For each theme, the table indicates the games to which they apply.

Table 3.4: Themes of the topic area “Suggestions”. The last column indicates the games to which each theme applies.

Theme	Description	Games
New challenges	To make the games more compelling, participants suggest to include new challenges.	BOXA, BOXS, BALL, STAMP, BATTERY1, BATTERY2, ALARMS
Increase playing area	Participants suggest to increase the area of the screen devoted to play.	BOXA, BOXS, BALL, STAMP, BATTERY1, BATTERY2, ALARMS
Improve graphics	Participants suggest to improve the graphics of the games or of some of their elements.	BOXA, BOXS, BALL, STAMP, BATTERY1, BATTERY2, ALARMS

Table 3.4: Themes of the topic area "Suggestions". The last column indicates the games to which each theme applies.

Theme	Description	Games
Change gestures	Participants suggest to change the gestures required to play.	BOXA, BOXS, BALL, STAMP, BATTERY1, BATTERY2, ALARMS
Other	Other suggestions.	BOXA, BOXS, BALL, STAMP, BATTERY1, BATTERY2, ALARMS

Participants provided different suggestions to improve the games. Seven participants suggested the introduction of new challenges into the games, possibly with increasing levels of difficulty. For example, P7 suggested introducing the timer also in the BALL game together with the requirement of inflating as much balloons as possible when advancing in the game levels, while P12 suggested making the buttons appear faster in the BATTERY2 and ALARMS games. P4 and P9 suggested moving the position of the buttons inside the playing area in these games and in BATTERY1. Finally, P9 and P3 suggested introducing a leaderboard into the app.

Another suggestion, given by five participants, concerned the extension of the area devoted to play, i.e. the one in the lower part of the screen. One of them suggested making it bigger than the upper one that is a source of distraction, especially for the BOXA, BOXS and STAMP games. The other four participants suggested instead to have a unique area on the screen, by making the upper one bigger. Indeed, they would have preferred to interact directly with the animations. For example, P11 said:

I would have preferred moving the finger directly over the post stamp. - STAMP

Similarly, P4 said:

The ALARMS game would be nicer if it allowed tapping directly on the ringing alarm clock instead of using the four buttons. - ALARMS

P7 provided this suggestion for the BATTERY1 and BATTERY2 games.

Four participants suggested changing the gestures required to play and three of them suggested playing with the thumbs. Instead, P11 suggested using both hands at the same time. In his opinion, this solution could improve users' performance and game enjoyment.

Other suggestions regarded the graphics of the game, as four participants suggested changing the background images or colors. One of these participant suggested also changing the animation of the BALL game, making the balloon breaking instead of deflating. Finally, two participants (P10 and P11) suggested organizing the main menu in a different way, for example placing similar games together. P11 suggested also using calming music in the games that require users to be accurate, and a more lively music in the games that require users to be fast.

3.3.6 Discussion and Design Opportunities

3.3.6.1 Graphics

Although participants provided positive comments on the pleasantness of the graphics in general and the animations in the upper part of the screen, they highlighted also negative aspects. In particular, for seven participants the animations were a source of distraction, due to their too vivid colors, size and movements, which affected the accuracy with which they carried out the required tasks. Considering the main goal of our application, i.e. the collection of accurate motor skills data to be used in scientific neurological studies, we gave more relevance to such opinions provided by users. More specifically, we have realized that animations of interactive objects on which the user can act only indirectly (or not at all) can be problematic. Indeed, while such objects may serve important aesthetic and entertainment functions, they can interfere with the main goal of the application. In general, when applying gamification to collect motor skills data, designers should thus carefully evaluate the introduction in the experience of such interactive objects that support only indirect interaction. Extending the part devoted to direct user interaction to the entire screen could instead help users be better focused on the actions they have to perform, which could improve their accuracy in this kind of tests, and thus also the quality of the data collected (in the following, *Design Opportunity 1*).

The analysis pointed out also opposite opinions regarding the chronometer used by some games to indicate remaining time. Indeed, while for two participants it was very useful, for other two participants it was not clear and sometimes confusing. In general, when applying gamification to motor tests that require to be completed within a predefined amount of time, it is crucial to adopt clear and unambiguous methods to indicate the available time, such as a bar whose length decreases to show remaining time (*Design Opportunity 2*).

3.3.6.2 Game Features

Despite the fact that participants found some games very simple and intuitive to play, also thanks to the help given by their introductory tutorials, other results revealed that for seven participants some games were difficult to play and that for 10 of them behavior and purpose of some games were unclear. This last finding suggests that some of the game mechanics we employed to turn neurological motor tests into games, e.g. tapping on buttons to charge a battery or moving the finger over a line to inflate a balloon, were not the most appropriate to offer clear and understandable games' behaviors and goals. Instead, they could have altered users' perception of the motor tests' goals, making them appear as not intuitive. In general, when applying gamification for motor skills data acquisition, the gamification strategy should be more focused on the motor performance evaluated by the tests, such as accuracy and/or reaction time (*Design Opportunity 3*).

In addition to these considerations, another factor that could have contributed to make the behavior and goals of the games unclear is the fact that for eight participants, the tutorials did not provide the most appropriate information about the actions they had to perform or the goals of the game. Since the tutorials actually explained all the game details across their three screens, a possible explanation of this finding could be that participants read them too quickly (as some of them admitted to have done) and thus skipped the information necessary to understand each game. Moreover, some of these participants pointed out that in the BOXA and STAMP games they thought they had to perform the required movement more than once. This could be due to the fact that the tutorial of each game included an animation loop which repeatedly showed the movements required by the game. In general, this result suggests that, in the context of motor skills data acquisition, simple and short tutorials should better focus the player on the specific aspects (such as speed and/or accuracy) on which user's performance will be evaluated by the gamified motor test. Short words describing those specific aspects could be written prominently in the user interface of the tutorial and then kept in the game as reminders (*Design Opportunity 4*).

Although participants found the games that involve a temporal limit compelling and entertaining, the analysis pointed also out that eight participants found these and other games boring. In particular, they said it was too simple to achieve the goal of the game or that the gameplay became repetitive after initial learning. This could be due to the fact that the games did not provide challenges with an increasing level of difficulty. In general, levels are a powerful tool to engage users and without them players may lose interest in the game because they have no measurable sense of progress (Lindley et al., 2008; Zichermann and Cunningham, 2011).

However, it must be noted that the neurological motor tests that were gamified by the games do not involve a sense of progression per se. In general, when gamification is applied to turn motor skill tasks into games, another organization of the games and their decomposition into simpler ones could be taken into consideration. For example, as the BOXS game could be seen as a more difficult version of the BOXA game, the overall experience could be organized in such a way that BOXS could be a more advanced level of BOXA that could be played only after successfully completing BOXA. Considering instead the BALL game that requires users to be accurate and fast at the same time, the game could be decomposed into simpler levels: players could be asked to be first accurate only, then, if they successfully complete the accuracy level, they could be asked to be fast only, and, after completing the speed level they could be asked to be fast and accurate at the same time. This decomposition and different organization of the games could give users a sense of progression and thus increase their involvement with the games (*Design Opportunity 5*).

The analysis revealed also that seven participants believed that the final scores did not reflect their performance with the specific game. In particular, they perceived some of the scores as meaningless, as they reported always the same result when the game was completed, e.g. 40 turns in the BOXS score, or 100% battery charge in the BATTERY1 and BATTERY2 games. This can be due to the fact that the score given to the player is not representative enough of the motor skill performance obtained, but is more an indication of percentage level of game completion. Moreover, one can also consider that scores were confined to each game. Thus, by not providing a global score system, users did not receive feedback on the global progress of their game skills (Werbach and Hunter, 2012). In general, when gamification is applied to turn motor skill tests into games, clear scoring of user's motor skills should be provided inside each game as well as in the context of a global scoring system that aggregates scores of the individual levels into a whole picture. A visualization of the global scoring system could help users gain a better perception of their progression (*Design Opportunity 6*).

3.3.6.3 Ergonomic Aspects

Although for seven participants placing the non-playing finger in the corner area was useful to avoid cheating or moving the smartphone when playing, results pointed out also that it was an uncomfortable requirement for 10 participants. The main reported reason was that this forces to place the smartphone over a table in order to play, which is unpractical in many everyday situations. Moreover, for some of them, it was a cause of distraction during the game, as the non-playing hand was in their visual field, or a hindrance, as it limited the movements of the playing hand. Finally, three participants pointed out that they did not understand its motivation. In addition, from our direct observation of the participants playing, it came out that eight of them did not understand this requirement the first time they played a game: they placed instead the playing finger in the corner area and thus were not able to play the game.

As described in Section 3.2.2, this requirement was introduced as a mechanism to constrain all users to adopt the same type of posture while playing and to prevent them from cheating by using both hands (which would make the collected motor data unreliable). However, the fact that for participants it was an uncomfortable requirement indicates that restricting too much the way users can interact with the games could not reflect the way they typically want to play. In the worst case, participants could be temped to adopt other postures to play, e.g. by placing the thumb of the non-playing hand in the corner area and holding the device with the other four fingers, that could make the device not completely stable and thus lower the quality of data collected.

In general, this highlights a tradeoff between extending the opportunities and the contexts in which users can play and collecting high-quality motor skills data. The solution that received the above mentioned users' criticism gave priority to data quality. A more flexible solution could be considered, by letting users keep the smartphone with the entire non-playing hand (this will make it unlikely to use that hand as a second hand for playing) while analyzing in real-time device accelerometer data to check whether users are handling it inside a given range of acceptable pitch and yaw angles. The game could start only when users are inside the acceptable range and the device could detect if they change significantly their position while playing, discarding data from unreliable game sessions. Overall, this solution could let users play also if they do not keep the phone on a table, possibly obtaining more game session recordings, while at the same time checking if possible changes in user's position during gameplay could make data unreliable (*Design Opportunity 7*).

The analysis also pointed out that for three participants the actions required by the BOXS game were troublesome, due to the friction with the screen of the fingertips or because they moved the smartphone. A possible solution to reduce the friction could be to limit the speed needed to carry out the task, by not requiring a minimum number of turns over a time interval (*Design Opportunity 8*). In addition, letting users handle the smartphone with their non-playing hand as we proposed above may also help in limiting the movements of the device when the users are playing and thus contribute to solve the issues faced by the three participants.

3.4 Second Prototype of MotorBrain

The study described in the previous section pointed out that participants perceived the gamified motor tests as pleasant, simple and entertaining to play. However, results revealed also negative aspects. We focused on the most important problems that emerged from the study, since they point out issues that could interfere with the main purpose of MotorBrain, i.e. the collection of useful motor skill data. Moreover, together with the domain expert, we preliminary evaluated the data collected by the app to assess weather it can be meaningful to evaluate users' motor skills. As a result, we decided to discard the motor test gamified by the BATTERY2 game, since the collected data contained a lot of noise. This could be probably due to the fact that most users used the two fingers required to play, i.e. index and medium, at the same time instead of one at a time, as we observed.

We then focused on the remaining six games and developed a second prototype of MotorBrain considering the design opportunities we outline in Section 3.3.6. In particular, we removed those interactive objects that supported only indirect interaction and extended the part devoted to user interaction to the entire screen (*Design Opportunity 1*). Given the greater amount of space that was available in this way, we increased the size of the graphic elements on which the user interacts to perform the motor tests. In particular, we modified the BALL game by adding other lines to form a path.

Moreover, we revised the gamification strategy for better challenging users to reach their best level of accuracy and/or speed and/or reaction time over three trials (*Design Opportunity 3*) and provide users with a global display of their performance after completing all trials of a test (*Design Opportunity 6*). We removed also the minimum number of actions required to reach the goal in BOXS and BATTERY (*Design Opportunity 8*), while we added the timer in the motor test gamified by the BALL game. To make all games clearer to understand, we revised the games' tutorials, highlighted their goals in the user interface (*Design Opportunity 4*) and changed the method to indicate the available time in those tests that have a temporal limit, i.e. BOXS, BATTERY1 and ALARMS, (*Design Opportunity 2*).

Finally, partially following *Design Opportunity 7*, we removed the constraint of placing the index finger of the non-playing hand in the corner area, thus allowing users to adopt a more flexible posture to play, i.e. by keeping the smartphone with the non-playing hand instead of being forced to place it on a table.

The second prototype of MotorBrain contains six motor tests, each one assessing one or two among the following variables: (i) accuracy, (ii) speed, (iii) reaction time. Tests are chosen from a main menu together with the hand for which the assessment is carried out (Figure 3.14). Since we removed most of the game elements from the tests, we changed also their name. The gamified motor tests that were referred to as BOXA, BOXS, BALL, STAMP, BATTERY1 and ALARMS are now referred to as CIRCLE-A, CIRCLE-S, BOX, PATH, TAPPING2 and TAPPING4.

Each gamified motor test requires three repetitions of a task to produce an assessment. Once users have completed the three trials, the app displays an assessment of motor performance concerning the test and the hand used.

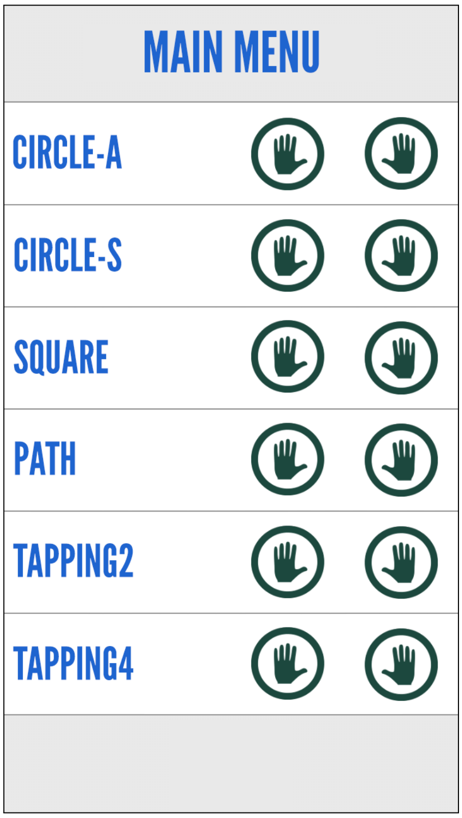


Figure 3.14: Main menu screen of the second prototype of MotorBrain.

3.4.1 Screen Organization

In all motor tests, the screen is organized in three parts: (i) the upper one indicates the motor performance variable(s) that will be assessed, (ii) the central one (hereinafter, *task area*) contains the graphic elements on which the user will perform the task, and (iii) the lower one indicates the number of completed trials for the current test. For each test, the application checks the pixel density and the resolution of the device display and scales the graphical elements on which to perform the task to make sure they are displayed with the same size regardless of the mobile device used. Users are asked to perform each test with the index finger of the assessed hand.

3.4.2 The Gamified Motor Tests

In the following, we describe in detail each test (the assessed variables are reported between brackets):

- CIRCLE-A (accuracy): The task area contains a colored ring with a thickness of 0.5 cm, and a diameter of 4 cm for the outer circle (Figure 3.15). Users have to move their index finger inside the ring, following it in all its length once and without ever lifting the finger. The starting point and the direction of movement depends on the chosen hand. For the right (left) hand, the starting point is located 30° on the left (right) with respect to the center of the ring and the movement must be performed in the clockwise (counterclockwise) direction. When the user moves the finger inside the ring in the right direction for the chosen hand, a white trace is shown as feedback. Otherwise, if the user moves the finger outside the ring or moves it in the wrong direction, a grey trace is shown. A trial ends when the player has covered a distance equal to the mean between the perimeter of the outer and inner circle of the ring, considering his/her movements over any location of the task area, or lifted the finger from the screen for more than 0.15 s.

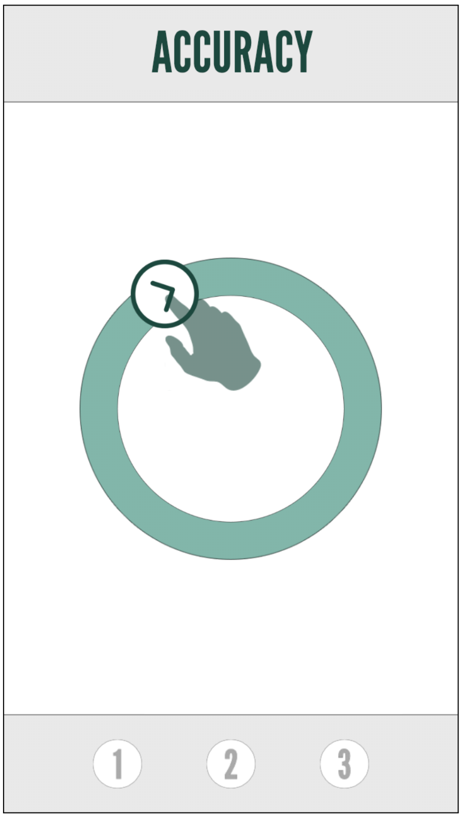


Figure 3.15: CIRCLE-A game.

- CIRCLE-S (accuracy, speed): The graphic elements of this test are the same of CIRCLE-A, but the thickness of the ring is larger, i.e. 0.7cm. Users have to move the index finger inside the ring following its entire length as many times as possible over a 10-second interval. The timer starts when the user touches the starting point and time remaining is indicated by a decreasing bar showed in the upper part of the task area (Figure 3.16). A trial ends when time expires or when users lift the finger for more than 0.15 s.

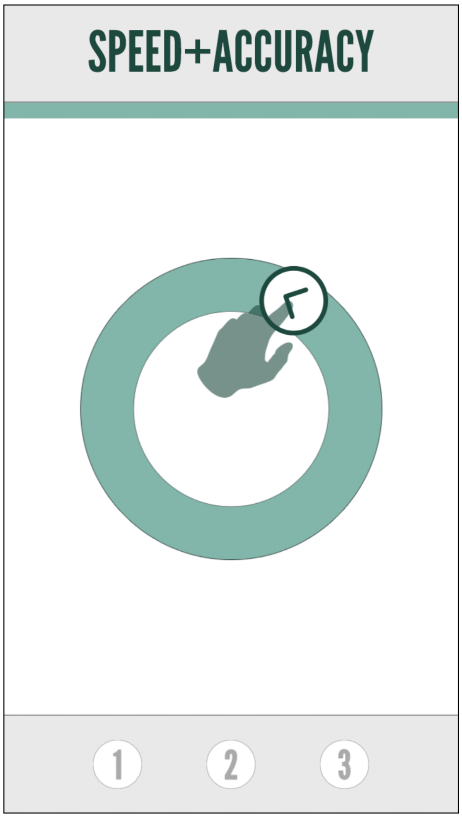


Figure 3.16: CIRCLE-S game.

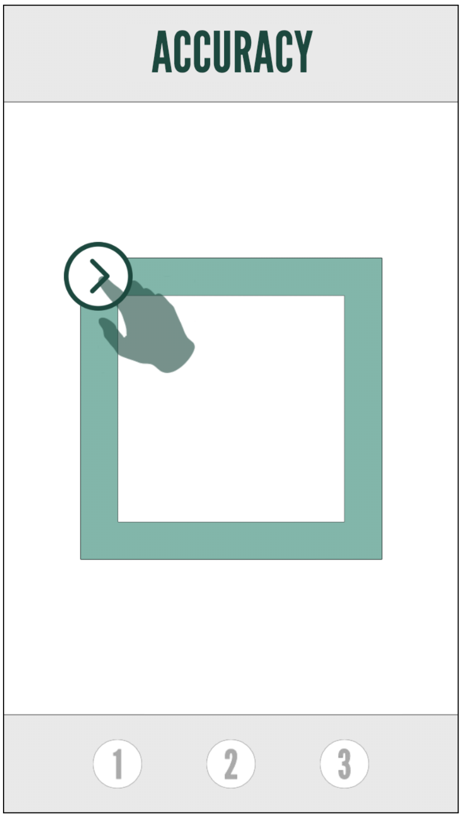


Figure 3.17: SQUARE game.

- SQUARE (accuracy): The task area contains a colored frame with a thickness of 0.5 cm, and outer square with 4 cm sides (Figure 3.18). Users have to move their index finger inside the frame, following its entire length once and without ever lifting the finger. As in CIRCLE-A, the starting point and the direction of movement depend on the chosen hand, i.e. upper left vertex of the frame and clockwise direction for the right hand and upper right vertex of the frame and counterclockwise direction for the left hand. Moreover, a white or grey trace is shown as feedback. A trial ends when users lift the finger for more than 0.15 s or when they have covered a distance equal to the mean between the perimeter of the outer and inner square, considering their movements over any location of the task area.

- PATH (accuracy, speed): The task area contains a colored path formed by four connected sloping lines that have a thickness of 0.5 cm (Figure 22). The length of each line is 3.72 cm and the angle between a line and the next one is 19.8°. The path starts from the upper-left part of the task area (if the user is performing the test with the right hand) or from the upper-right part (left hand). Users have to move the index finger from the start to the end of the path as fast as possible over a 5-second interval. The timer starts when the user touches the starting point and time remaining is displayed as in CIRCLE-S. A trial ends when time expires, when users lift the finger for more than 0.15 s or when they have covered a distance equal to the path, considering their movements over any location of the task area.

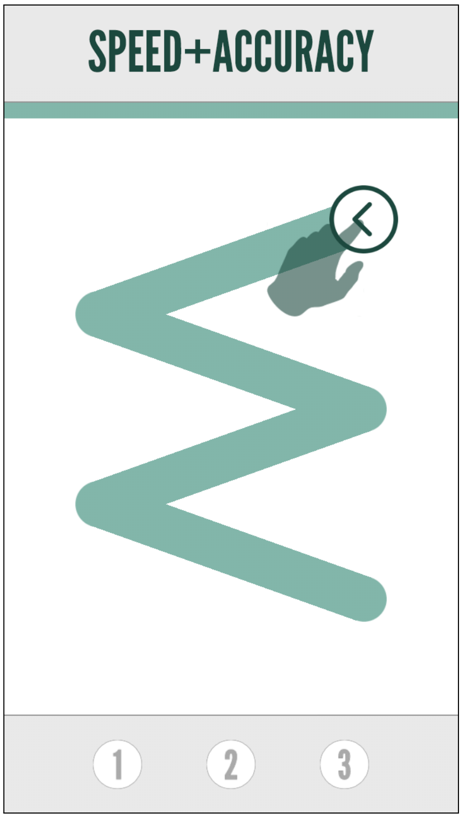


Figure 3.18: PATH game.

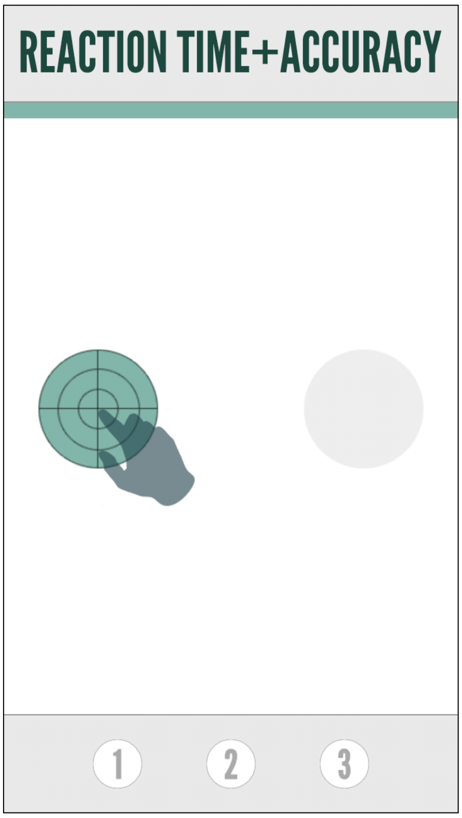


Figure 3.19: TAPPING2 game.

- TAPPING2 (accuracy, reaction time): The task area contains two round buttons, 1.6 cm in diameter, vertically aligned (Figure 3.19). Only one button is enabled at any time, and is highlighted by a black viewfinder on a colored background. The viewfinder consists of three concentric circles that divide the button area in three parts: central, middle and external. Otherwise, when a button is not highlighted, it is displayed in light grey. Users have to tap the highlighted button (within a 0.2 cm margin) with the index finger. A black circle is shown as a feedback where the player has tapped on the screen. Tapping an highlighted button moves the highlight to the other button and users have to tap on as many highlighted buttons as possible over a 15 seconds interval. The button that will be initially highlighted depends on the hand chosen by the user to perform the test: left button for the right hand and right button for the left hand. The timer starts when the user tap on the first highlighted button and time remaining is indicated as in CIRCLE-S. Users are told at the beginning of the test to accurately tap on the center of the highlighted button.

- TAPPING4 (accuracy, reaction time): The task area shows four buttons displayed as in TAPPING2 (Figure 24). At any time, only one button is highlighted and tapping an highlighted button moves the highlight to another (randomly selected) one. Unlike TAPPING2, also the initial highlighted button is randomly selected, regardless of the chosen hand. Users have to tap on as many highlighted buttons as possible over a 15 seconds interval. Moreover, users are told at the beginning of the test to accurately tap the center of each highlighted button.

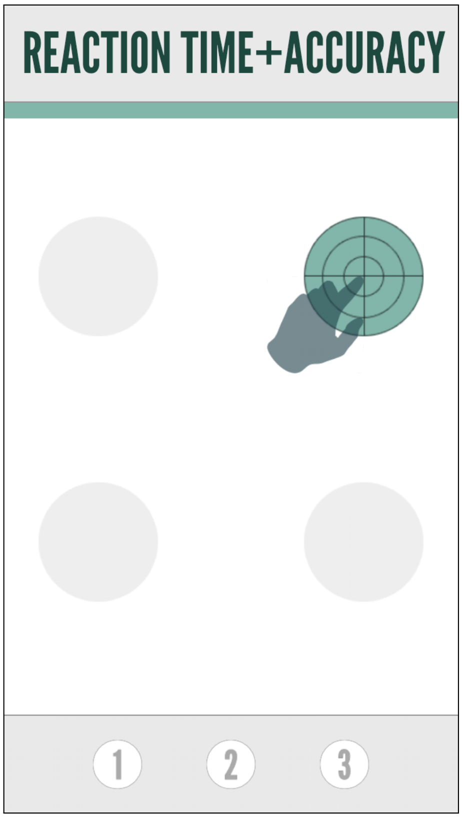
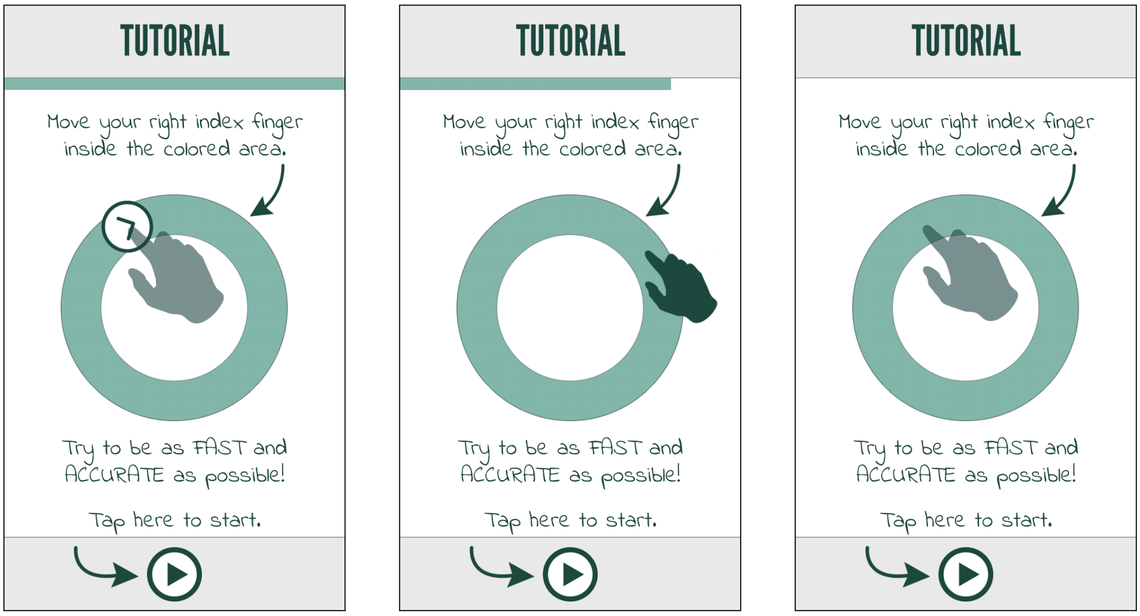


Figure 3.20: TAPPING4 game.

3.4.3 Test Tutorials

When users choose a motor test in the main menu, they are presented with a tutorial that explains, through short texts and an animation, the actions required to perform the task and the variables that will be assessed (see Figure 3.21 for an example).



(a) (b) (c)

Figure 3.21: Tutorial of the CIRCLE-S test. (a) Start of the animation, illustrating with a fade-in hand the action required to start the test; (b) the animation shows the hand performing the movement required to carry out the task until the available time (indicated by the green bar) expires (c) end of the animation with fading-out of the hand.

3.4.4 Assessed Variables and Data Collection

When a motor test ends, the app shows the "Assessment" screen and two buttons, one to perform the test again and one to go back to the main menu (see Figure 3.22).

Considering single trials, the assessed variables are defined as follows:

- accuracy (CIRCLE-A, CIRCLE-S, SQUARE, PATH, TAPPING2, TAPPING4): distance covered by the user with the index finger over the colored ring in CIRCLE-A and CIRCLE-S, over the colored frame in SQUARE and over the colored path in PATH, expressed in percentage with respect to the target distance, i.e. length of the mean between the perimeter of the outer and inner circle of the ring in CIRCLE-A and CIRCLE-S, length of the mean between the perimeter of the outer and inner square of the frame in SQUARE and length of the path in PATH.

  In TAPPING2 and TAPPING4, accuracy is calculated by adding 5 points for each tap on the center area of the viewfinder, 2 points for each tap on the middle area and 1 point for each tap on the external area;

- reaction time (TAPPING2, TAPPING4): mean of the reaction times (in seconds), i.e. time elapsed between the appearance of a highlighted button and tapping, during the trial;

- speed (PATH, CIRCLE-S): number of seconds the user spent to perform the trial in PATH; number of times the user was able to follow the ring entirely in CIRCLE-S.

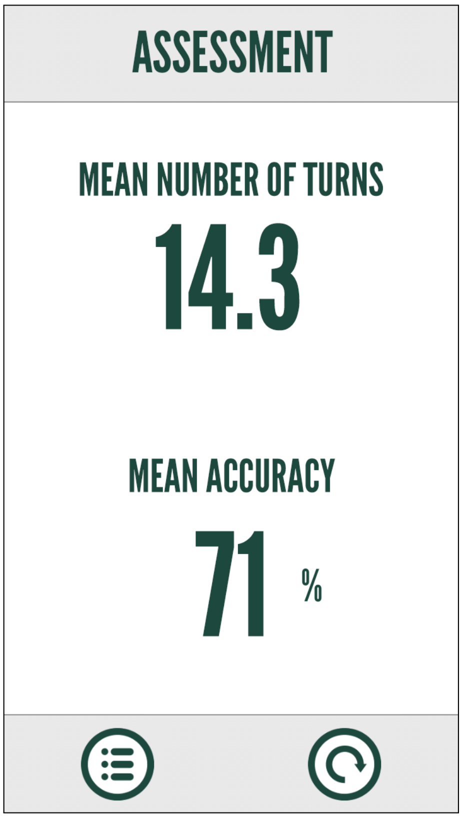


Figure 3.22: Assessment of user's performance displayed after completing the three trials of the CIRCLE-S test.

For each test, the displayed assessments are the means of the variables assessed in each trial. For TAPPING2 and TAPPING4, mean accuracy is expressed in percentage with respect to the ideal situation in which the user always taps on the center area of each highlighted button in each trial.

All data required to determine these assessments is automatically recorded by the app with an event-driven sampling: each time the user touches the screen or changes finger position on the screen, the coordinates in pixels and the timestamp of the event are automatically saved. In the CIRCLE-S test, the sampling is reduced by one event out of two: since the user has to perform the required movements several times, the app collected a huge amount of data but only half of it was found necessary to calculate the displayed assessment.

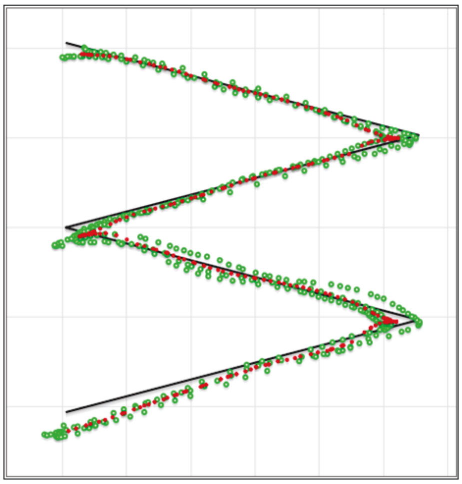
The app collects also some device specifications, such as screen resolution and density, and the coordinates in pixels of the graphical elements in the task area. All this data is then sent to a server in a secure way at the end of each test. This information is necessary to reconstruct and analyze the test trial off-line. In particular, for each trial, we calculate the following variables:

- Movement Time (MT) (CIRCLE-A, CIRCLE-S, SQUARE, PATH): number of seconds the user spent to perform a trial in CIRCLE-A, SQUARE and PATH; number of times the user was able to follow the ring entirely in CIRCLE-S;

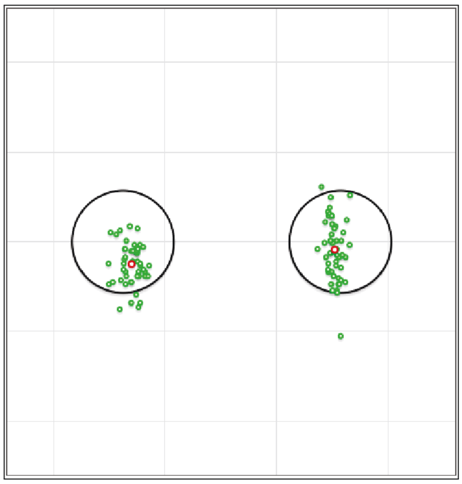
- Shape Accuracy (SA) (CIRCLE-A, CIRCLE-S, SQUARE, PATH, TAPPING2, TAPPING4): for CIRCE-A, CIRCLE-S, SQUARE and PATH, this variable is the same as *accuracy* defined before but not expressed in percentage. For TAPPING2 and TAPPING4, this variable is the total number of points obtained by the user;

- Reaction Time (RT) (TAPPING2, TAPPING4): this variable is the same as *reaction time* defined before.

Then, considering all the three trials of a test, a bidimensional graph (Movement Graph) is generated off-line to show all user's trajectories of movement inside the test area (CIRCLE-A, CIRCLE-S, SQUARE, PATH), as well as their means. For TAPPING2 and TAPPING4 this graph contains all the points where the user tapped on the screen (see Figure 3.23 for two examples).



(a)



(b)

Figure 3.23: Movement Graph for the PATH (a) and TAPPING2 tests (b). The black shapes represent the graphic elements on which the user performed the task, i.e. the path in PATH and the two buttons in TAPPING2. The green dots indicate all user movements (PATH) or where the user tapped (TAPPING2) on the screen, considering all three trials of each motor test. Finally, the red dots indicate the mean of the movements (PATH) or of the tap positions (TAPPING2) over the three trials.

Moreover, all trials are grouped according to the motor test they belong and the preferred hand (right-handed and left-handed) of the user who performed it. Then, according to users' age, they are further divided into two age groups, i.e. a *Young group* (participants who are less than 50 years old), and an *Old group* (participants who are more than 50 years old. For each age group, the following variables (performance assessments) are calculated, considering separately all trials performed with the right and left hand (DX and SX):

- means of MT (MMT_DX, MMT_SX) for the CIRCLE-A, CIRCLE-S, SQUARE, and PATH tests;
- means of SA (MSA_DX, MSA_SX) for the CIRCLE-A, CIRCLE-S, SQUARE, PATH, TAPPING2, and TAPPING4 tests;
- means of RT (MRT_DX, MRT_SX) for the TAPPING2 and TAPPING4 tests.

Similar measures of motor performance have been employed in other literature studies, e.g. (Raw et al., 2012; Welsh et al., 2007).

3.5 Quantitative Pilot Study

The goal of the quantitative pilot study was to evaluate whether the data collected by the second prototype of MotorBrain can be effectively used for describing the aging of human motor performance.

3.5.1 Participants

Sixty healthy participants (29 M, 31 F) were recruited through direct contact. All of them were right handed and they neither reported movement disorders nor suffered from headache or migraine during the study. Both the Young and the Old group consisted of 30 participants. The age of the Young group (13 M, 17 F) ranged from 18 to 49 ($M = 27.83, SD = 10.05$), while the age of the Old group (16 M, 14 F) ranged from 51 to 74 ($M = 59.43, SD = 4.90$).

All participants were asked to perform all tests with the index finger of their dominant hand, while keeping the smartphone with the other hand. The order of the tests was counterbalanced.

3.5.2 Results

Table 3.5 reports the means of performance assessment variables. The table indicates also the results of the tests we employed to assess the statistical significance of differences between the two groups. In particular, we employed a *t-test* when the assessed variables followed a normal distribution. Otherwise, we employed a *Wilcoxon signed-rank test*.

Table 3.5: Means of the performance assessment variables for each motor test and age group. The table indicates also the results of the tests we employed to assess the statistical significance of differences between the two groups.

Test	**Variable**	**Young Group**	**Old Group**	**Significance**
CIRCLE-A				
	MMT_DX	$NT = 94, M = 2.71, SD = 1.31$	$NT = 122, M = 2.82, SD = 1.674$	$U = 5504, p = .61$
	MSA_DX	$NT = 94, M = 0.93, SD = 0.134$	$NT = 122, M = 0.89, SD = 0.15$	$U = 6893, p < .05, r = .17$
CIRCLE-S				
	MMT_DX	$NT = 98, M = 15.03, SD = 6.46$	$NT = 111, M = 11.99, SD = 4.80$	$T(207) = 3.82, p < .001, \eta^2 = .07$
	MSA_DX	$NT = 98, M = 0.86, SD = 0.12$	$NT = 111, M = 0.87, SD = 0.11$	$U = 5215.5, p = .61$
SQUARE				
	MMT_DX	$NT = 96, M = 3.58, SD = 1.37$	$NT = 111, M = 3.39, SD = 1.35$	$U = 5872, p = .21$
	MSA_DX	$NT = 96, M = 0.93, SD = 0.10$	$NT = 111, M = 0.89, SD = 0.13$	$U = 6522.5, p < .01, r = .19$
PATH				
	MMT_DX	$NT = 89, M = 2.52, SD = 0.90$	$NT = 117, M = 2.43, SD = 0.78$	$U = 5573, p = .39$
	MSA_DX	$NT = 89, M = 0.89, SD = 0.13$	$NT = 117, M = 0.90, SD = 0.08$	$U = 5460, p = .55$
TAPPING2				
	MRT_DX	$NT = 90, M = 0.36, SD = 0.17$	$NT = 108, M = 0.43, SD = 0.23$	$U = 4061, p < .05, r = .14$
	MSA_DX	$NT = 90, M = 122.98, SD = 36.19$	$NT = 108, M = 131.56, SD = 55.19$	$U = 4794.5, p = .87$

NT: Number of trials.

Table 3.5: Means of the performance assessment variables for each motor test and age group. The table indicates also the results of the tests we employed to assess the statistical significance of differences between the two groups.

Test	Variable	Young Group	Old Group	Significance
TAPPING4				
	MRT_DX	$NT = 83, M = 0.62, SD = 0.11$	$NT = 111, M = 0.69, SD = 0.12$	$U = 2605, p < .001, r = .37$
	MSA_DX	$NT = 83, M = 85.83, SD = 16.73$	$NT = 111, M = 88.36, SD = 22.45$	$T(192) = -0.90, p = .37$

NT: Number of trials.

3.5.3 Discussion

The analysis of the assessment variables for participants' dominant hand points out that accuracy was lower for the Old group than the Young group on those tests that required only accuracy, i.e. CIRCLE-A and SQUARE. This result could be due to the decline in sensorimotor control and functioning that comes with aging (Seidler et al., 2010), such as increased variability of movement (Contreras-Vidal et al., 1998). No differences were instead found for movement time in CIRCLE-A and SQUARE: this can be explained by the fact that participants were not asked to be fast as done by other tests.

The analysis highlighted also that in those tests that required participants to react and move fast, the Old group performed worse than the Young one. Indeed, participants in the Old group did fewer turns in the CIRCLE-S test, while in the TAPPING2 and TAPPING4 tests their reaction time was higher than the Young group. Interestingly, no differences were found in terms of accuracy between the two groups in the three above mentioned tests. The increase in movement and reaction time can be seen as a compensatory action for maintaining accuracy (Seidler et al., 2010). Moreover, other factors that can contribute to such result are reduced conduction velocity in nerve fibres, reduction of muscle elasticity, increased musculoeskeletal rigidity (Seidler et al., 2010; Welsh et al., 2007), and age-related changes in the central nervous system that affects muscles activation and coordination in the initial stages of a movement (Bautmans et al., 2011). that identified an age-related decline in movement time, reaction time or accuracy with similar tasks (Bautmans et al., 2011; Welsh et al., 2007; Raw et al., 2012).

PATH was the only test that did not point out any statistically significant difference between the two groups. However, in this case, the graph that can be generated by the raw data (see Figure 3.23) allows one to observe interesting characteristics of motor control. For example, inaccuracy of trajectories across the corners of the path can be useful for highlighting rigidity or slowness when initiating a new movement. Such characteristics are useful in assessing of movement disorders caused by Parkinson's disease (Hoehn and Yahr, 1967), and have been used in (Agostino et al., 1992).

In general, the Movement Graph of all tests can be useful for outlining motor control impairments. For example, the graph of a subject on a single test trial can be quantitatively compared to predefined templates of the same test, or qualitatively rated by experts, to identify motor tremors or impairments. Similar analyses and assessments of a traced shape to identify motor disorders have been carried out in previous literature studies, e.g. (Surangsrirat and Thanawattano, 2012; Westin et al., 2010a).

4

Using Research in the Large with an On-Line Health App for Psychological Assessment

In the third case study, we focus again on the *Mental and Emotional Health* category (see Section 1.1.2), as in Chapter 2. This time, we considered the issue of problematic Internet use (PIU), i.e. a multidimensional syndrome that consists of cognitive, emotional, and behavioral symptoms that result in difficulties with managing one's offline life (Caplan, 2005). PIU is attracting increasing interest from the scientific community, due to the negative outcomes that it could lead to, such as work or school related problems (Young, 1998b) and social skill deficit (Caplan, 2005; Kraut et al., 1998).

For this case study, we developed an on-line desktop app (Psicoscrigno) that includes a scientifically evaluated assessment of PIU. According to the classification based on the PPM model (see Section 1.1.1), Psicoscrigno is a *predisposing* and *reinforcing app*. Then, we exploited the possibilities offered by research in the large (see Section 1.2) to make our app available to a high number of users as well as carry out a scientific experiment for better describing PIU in terms of personality traits. To this purposes, we published Psicoscrigno on Facebook and included into it also an assessment of time perspective (TP), i.e. "a relatively stable individual-differences process" by which individuals partition the flow of human experience into the distinct temporal categories (or frames) of past, present, and future (Zimbardo et al., 1997; Zimbardo and Boyd, 1999). To the best of our knowledge, no study so far has investigated the possible relationship between TP and PIU.

This chapter is organized as follows: Section 4.1 briefly surveys PIU and TP studies and the instruments we used to assess the two constructs. It also provides an overview on the available health apps that offer an assessment of PIU. Section 4.2 presents Psicoscrigno and its public release on Facebook. Then, Section 4.3 describes the correlational study we carried out (Chittaro and Vianello, 2013).

4.1 Related Work

4.1.1 Problematic Internet Use

In the literature, different names have been used to refer to ways of using the Internet that lead to negative outcomes, e.g. Internet addiction (Young, 1998b), Internet abuse (Morahan-Martin, 1999), and problematic Internet use (Caplan, 2002). The different terms reflect the different conceptualizations that have been given of PIU. According to the review by Morahan-Martin (2008), despite the different conceptualizations of PIU in the literature, there is a general agreement that PIU is defined in terms of the negative effects of Internet use, which can affect an individual's life. Moreover, researchers agree that PIU involves preoccupation with using the Internet, compulsive Internet use, subjective feelings of inability to limit use, and using the Internet to escape or alter negative moods.

As a result of different attempts to conceptualize the PIU construct, a number of measurement scales have been developed - see (Laconi et al., 2014) for a review. For Psicoscrigno and for our study we employ Caplan's Generalized Problematic Internet Use Scale 2 (GPIUS2) (Caplan, 2010), because it taps all the aspects mentioned above, which most researchers agree are related to PIU despite the different conceptualizations they adopt for PIU.

The GPIUS2 consists of five subscales that tap four constructs. The first subscale concerns Preference for Online Social Interaction (POSI), which is characterized by beliefs that one is safer, more efficacious, more confident, and more comfortable with online interpersonal interactions and relationship than with traditional face-to-face social activities. The second subscale taps Mood Regulation (MR), which emphasizes the motivation to use the Internet to alleviate distressing feelings. The third and fourth subscales, Cognitive Preoccupation (CP) and Compulsive Internet Use (CIU) focus on Deficient Self-Regulation (DSR). In particular, CP refers to obsessive thought patterns involving Internet use, while CIU reflects the compulsive nature of PIU. The fifth subscale taps Negative Outcomes (NO), which reflects the negative outcomes that are associated with a problematic use of the Internet. The GPIUS2 contains 15 items that ask respondents to rate their extent of agreement on an 8-point scale (1="definitely disagree", 8 = "definitely agree"). Each of the five subscales contains 3 items. The GPIUS2 scale can be used in two different ways: as a set of separate subscales or as an overall composite score which reflects the degree of PIU of an individual (Caplan, 2010). In our study, we consider both the composite score (GPIUS2CS) and the five subscales.

4.1.2 On-Line Health Apps for PIU Assessment

Currently, there are several on-line health apps that offer people an assessment of different mental health conditions, such as anxiety or depression, see e.g. Mental Health Screening Tools (Mental Health America, 2014) or Project Implicit Mental Health (Project Implicit, 2014). A number of such apps are also focused on problematic Internet use. In particular, some of them, i.e. Counselling Resource (CounsellingResource.com, 2014), HealthyPlace (HealthyPlace.com, 2014), Netaddiction (The Center for Internet Addiction Recovery, 2014), and Psychology Tools (Psychology-Tools.com, 2014), employ an electronic version of the Internet Addiction Test (IAT) (Young, 1998a) and report its reference. Such apps can also provide users with other information on PIU or indicate ranges for interpreting their obtained IAT's score. The reason why all of the above mentioned app employ the IAT can be explained by the fact that it was one of the firsts developed measures of PIU and to date it is the most widely used in the scientific literature - see (Laconi et al., 2014) for a recent review of PIU assessments.

However, other health apps do not employ any scientifically evaluated instrument for assessing PIU or do not report any reference for the questionnaire they employ. In particular, two of these apps state that the questionnaire they offer was developed by a psychological expert, but do not report any reference to its evaluation, e.g. Psych Central (Psych Central, 2014) and Interactive Quiz (Weinstein and bambooyou.com, 2014). Several other on-line apps do not mention any expert at all but include questionnaires created by their users, e.g. in AllTheTests.com (AllTheTests.com, 2014) or in GoToQuiz.com (GoToQuiz.com, 2014). Finally, assessments of PIU are also included into newspapers websites, e.g. TheGuardian.com (The Guardian.com, 2013), or created for fun purposes, e.g. in QuizMeme (serve-you.net, 2014).

Overall, among the high availability of on-line health apps that contain an assessment of PIU, several of them do not employ a valid and reliable measure of it. In these cases, such apps can provide people with unreliable information.

4.1.3 Time Perspective

Zimbardo and Boyd (1999) developed a scale to measure TP, the Zimbardo Time Perspective Inventory (ZTPI), which subdivides the past, present and future temporal frames into five subscales, or subframes. The Past Negative (PN) subscale reflects a generally negative, aversive view of the past, which may be due to actual experiences of unpleasant or traumatic events, negative reconstructions of benign events, or to a mix of both. The Past Positive (PP) subscale reflects a more sentimental and warm view of the past. The Present Hedonistic (PH) subscale reflects a hedonistic, risk-taking attitude towards time and life. It suggests an orientation towards present pleasure with little concern for future consequences. The Present Fatalistic (PF) subscale reflects a fatalistic, helpless, and hopeless attitude towards life. Finally, the Future (F) subscale reflects a general future orientation, suggesting that behavior is dominated by a striving for future goals and rewards. The ZTPI contains 56 items that ask respondents to indicate how characteristic a statement is of them on a 5-point Likert scale ranging from 1 (very uncharacteristic) to 5 (very characteristic). The PN subscale is composed by 10 items, PP by 9 items, PF by 9 items, PH by 15 items and F by 13 items. The score obtained on each subscale is independent of those obtained on the other subscales. A high score on a subscale reveals a temporal orientation of an individual on that temporal frame, while the overall time perspective of an individual is defined by the different results obtained on the five subscales.

The ZTPI has been used in the literature to investigate the relationships between TP and other personality concepts, e.g. some researchers have shown associations between TP and the Big Five personality traits, i.e. Neuroticism, Extraversion, Conscientiousness, Agreeableness, and Openness to experience (Costa and McCrae, 1992). The strongest association that emerged was between F and Conscientiousness (Adams and Nettle, 2009; Dunkel and Weber, 2010; Keough et al., 1999). Other studies have shown that a prevalence of a particular temporal frame has implications for various aspects of human behavior, see (Drake et al., 2008) for a review. In the following, we will focus on the relationships that have been found between TP and problematic human behaviors.

Zimbardo and Boyd (1999) found that a prevalence of PN is correlated with depression, anxiety and low self-esteem. Moreover, it has been shown that past negative oriented individuals have fewer close friends (Zimbardo and Boyd, 1999) and are more likely to be in alcohol and drug programmes (Klingeman, 2001) than people with a different time orientation. Present oriented (PH and PF) individuals have been shown to be less influenced by safe sex practices (Rothspan and Read, 1996) and more likely to engage in risky driving (Zimbardo et al., 1997), alcohol and drug misuse (Keough et al., 1999) and to suffer chronic homelessness (Epel et al., 1999). PF is also associated with aggression, anxiety, depression (Zimbardo and Boyd, 1999), and with avoidant procrastination, while PH is associated with arousal procrastination (Ferrari and Diazmorales, 2007). Prevalence in both PF and PN frames was shown to discriminate between severe suicidal ideators and nonideators among high school students (Laghi et al., 2009). As reported in (Drake et al., 2008), future orientation has been associated with less problematic behaviors. It has, however, been suggested that an overemphasis on future goals compromises spontaneity resulting in poor ability to switch off and enjoy the present (Boniwell and Zimbardo, 2004).

Considering the role that TP has in various problematic behaviors, it is worth investigating whether TP can be a predictor also for PIU.

4.2 Psicoscrigno and its Public Release

Psicoscrigno is an on-line health app that contains an electronic, Italian-translated version of the GPIUS2 and ZTPI. Considering its simplicity, we did not carry out user-centered studies to evaluate its design, as we did in the other two case studies - see Chapter 2 and Chapter 3. Instead, we made it directly available to people. In particular, we employed the research in the large paradigm (see Section 1.2) to make Psicoscrigno available to a large number of users and recruit participants for our study. To this purpose, we integrated the app with Facebook and included it in the Facebook app directory. Its availability was also announced on the web site of an Italian newspaper.

Users could answer the questionnaires by logging in to their Facebook profile and then using the application. They were preliminarily informed by the application that they would be allowed to answer each questionnaire only once (see Figure 4.1).

When the user chooses to answer a questionnaire, the app shows an additional page that includes information on the specific assessment, instructions for answering and reference of its scientific validation (Figure 4.2).

Figure 4.1: "Home" screen of Psicoscrigno. The app informs the user that he has already answered the ZTPI test.

Figure 4.2: Introduction page of the GPIUS2 test. The user can choose to start answering (above button) or go back to the "Home" screen (bottom button).

From this page, users can choose to start answering the questionnaire or go back to the "Home" screen of Psicoscrigno. If the user chooses to answer, the app shows the initial page of the questionnaire. This page asks also users to enter their gender and age (Figure 4.3).

Figure 4.3: First 5 questions of the GPIUS2 test. In this page, users also specified their age (Età) and gender (Sesso).

The application checks for possibly skipped answers on questionnaire items and initial demographic data, and asks users to enter them. Then, to thank users, after they complete questionnaires, the application provides them with the obtained scores and explanations of the scales (Figure 4.4).

Participants were also given the opportunity to publish a link to the application on or share the obtained scores on their profiles with a single click. Finally, it offered participants the standard Facebook "Invite friends" option (Figure 4.5).

Figure 4.4: "Results" page of the GPIUS2 test. From this page, the user can choose to (i) publish the score on his/her Facebook profile (first link), (ii) invite friends to install the app (second link), and (iii) go back to the "Home" screen (third link)

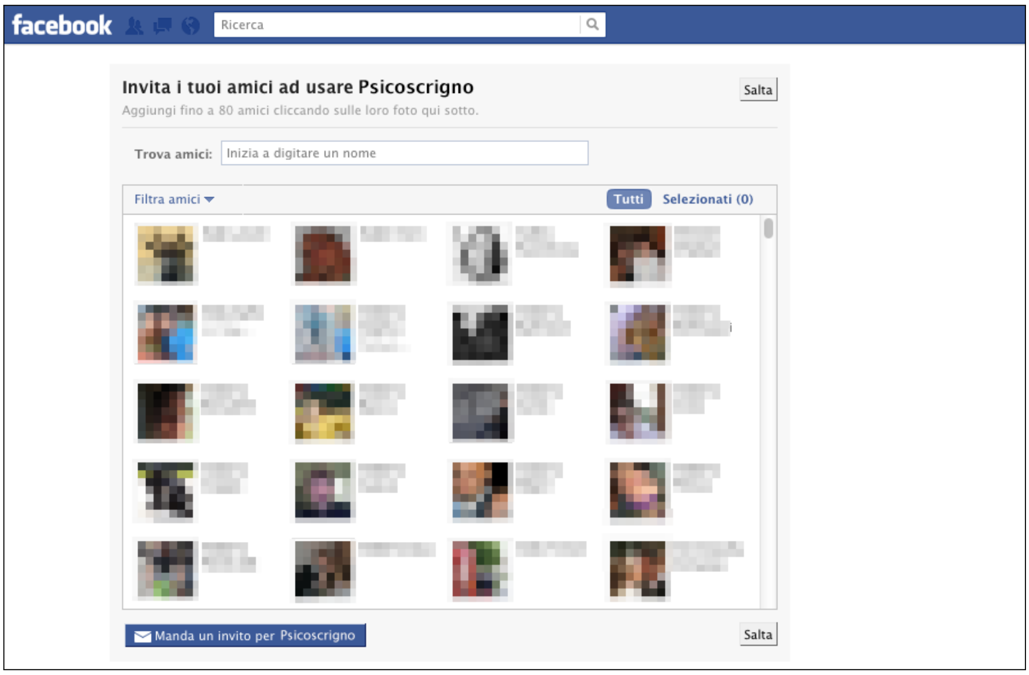


Figure 4.5: Facebook "Invite Friends" option: participants could conveniently suggest their Facebook friends to use the application (faces of users in this figure have been electronically blurred for privacy reasons).

4.3 Correlational Study With Research in the Large

4.3.1 Hypothesis

Given the dependence aspect that could be involved in PIU and the association of TP with problematic behaviors that have an addictive component, such as substance abuse and reduced offline social contacts, we hypothesize that TP predicts PIU. In particular, considering the role of PN and PF reported by the literature with respect to addiction and social contacts, we hypothesize that people who score high on PN or PF should report a high PIU.

4.3.2 Measures

4.3.2.1 Demographics and Usage Data

Before the installation, Psicsocrigno asked users for permission to have access to their basic profile information such as gender and age. Moreover, it collected also usage data, i.e. the answer to each question, the final scores, and the time required by participants to complete each questionnaire. Such data was stored in a secure database and in anonymous manner.

4.3.2.2 GPIUS2

Participants completed the electronic, Italian version of the GPIUS2, which had good internal reliability. Cronbach's alpha for GPIUS2CS was .89 (.91 in the English version (Caplan, 2010)), for POSI .87 (.82), for MR .81 (.86), for CP .83 (.86), for CIU .84 (.87) and for NO .86 (.83).

4.3.2.3 ZTPI

Participants completed the electronic, Italian version of the ZTPI, which had good internal reliability. Cronbach's alpha for PN was .82 in the Italian version (.82 in the English version (Zimbardo and Boyd, 1999)), for PP .71 (.80), for PF .71 (.74), for PH .78 (.79) and for F .66 (.77).

4.3.3 Results

4.3.3.1 Demographics and Usage Data

For 6 months since the day of the announcement, the app was used by 692 Facebook users: 459 participants chose to answer the ZTPI only, 78 participants the GPIUS only, and 155 participants answered both questionnaires. These 155 participants form the sample of our study.

We used questionnaire completion time data to identify cases that could need to be removed, because extremely quick completion times on on-line surveys suggest that respondents have not given the questions due consideration and thus can introduce lower quality data in the dataset while a long completion time might mean that respondents have been interrupted while filling out the questionnaire (Malhotra, 2008). To exclude these cases, we followed the approach of previous studies with online surveys (Yan and Tourangeau, 2008), analyzing the data for outliers and dropping the participants who were in the lower and upper one percentile of time completion values in at least one of the questionnaires. Based on this analysis, 6 participants were discarded. We did not take into consideration participants' connectivity speed, as each page of the questionnaires was very small (no larger than 59 Kb) and did not contain any image or video. In these cases, differences in connectivity speed should not produce meaningful differences in completion times (Malhotra, 2008).

Among the remaining 149 participants, the mean completion time for ZTPI was 519 seconds ($SD = 227.33$), while for GPIUS2 was 124 seconds ($SD = 53.79$). The mean of participants' self-reported age was 32.40 years ($SD = 11.80$).

Bases on participants' self-reported data, there were 79 male and 70 female respondents in the sample. For 21 participants, the Facebook user profile did not contain gender information and we used the gender they had stated in the questionnaires. For the remaining 128 participants, gender information was available in their Facebook profile as well in the questionnaires. For these participants we checked if there were any mismatches between the gender stated in their Facebook profile and the gender stated in the questionnaires and there was a mismatch for only one participant. In this single case of mismatch, we used the gender reported on the Facebook profile as the participant's gender. In general, Facebook encourages more truthful profiles than other online social networks (Lampe et al., 2007).

4.3.3.2 Relations Between the ZTPI Subscales and GPIUS2CS

Zero-Order Correlations

Data were first subjected to the Kolmogorov-Smirnov test of normality. The test revealed deviations from the normal distribution for GPIUS2CS and PP. The relationships between GPIUS2CS and the ZTPI subscales were thus assessed using Kendall's rank correlation coefficient. Results were then converted to Pearson's r following Walker (2003) (Table 4.1). Zero-order correlations showed that GPIUS2CS correlated moderately with PN and weakly with PF. Also, significant correlations between some ZTPI subscales were found: PN correlated weakly and negatively with PP; PN strongly and positively with PF; PH weakly and positively with PF; PH weakly and negatively with F; PF moderately and negatively with F.

We also tested if GPIUS2CS and the ZTPI subscales correlated with age and gender, which was dummy coded (0=men, 1=women), finding no significant correlations.

Table 4.1: Converted Pearson correlations (from rank order correlations) between ZTPI subscales and GPIUS2CS.

	PN	**PP**	**PH**	**PF**	**F**
GPIUS2CS ($M = 48.41, SD = 17.89$)	.37***	-.08	-.03	.26**	-.09
PN ($M = 3.22, SD = .69$)		-.22*	-.02	.54***	-.17
PP ($M = 3.38, SD = .57$)			.13	.06	.09
PH ($M = 3.46, SD = .47$)				.25**	-.25**
PF ($M = 2.69, SD = .56$)					-.37***
F ($M = 3.30, SD = .45$)					

Note: $n = 149$, * $p < .05$. ** $p < .01$. *** $p < .001$.

Partial Correlations

The zero-order correlations revealed that PN and PF are both correlated with GPIUS2CS. These correlations were of our primary interest, considering the aim of the study. Moreover, we found that PN correlated strongly and positively with PF. This correlation suggests that the use of the two subscales together when attempting to predict PIU might be possibly redundant and that the two subscales might not be two independent predictors. For this reason, we carried out two partial correlations, the first involving GPIUS2CS and PN partialing out PF, and the second involving GPIUS2CS and PF partialing out PN. The correlation between PN and GPIUS2CS remained significant after partialing out PF scores ($\tau = .19, r = .29, p < .001$).

Linear Regressions

In order to better investigate the effects of PN and PF on PIU we carried out two linear regressions. For both regressions, GPIUS2CS was the dependent variable, while PN was the independent variable in the first regression and PF was the independent variable in the second regression (see Table 4.2 for the regression parameters). Prior to analysis, GPIUS2CS was normalized using a square root transformation. Results indicated that both PN ($F(1,147) = 16.40, p < .001$) and PF ($F(1,147) = 8.78, p < .01$) are significant predictors of PIU. PN accounted for 9% of the GPIUS2CS variance while PF accounted for 5% of the GPIUS2CS variance.

Table 4.2: Regression parameters for GPIUS2CS.

Dependent variable	**Adjusted** R^2	**Predictor**	β	**Stand.** β	**t**	**p**
GPIUS2CS	.09	PN	.02	.32	4.05	$< .001$
GPIUS2CS	.05	PF	.02	.24	2.96	$< .01$

4.3.3.3 Relations Between PN, PF and the GPIUS2 Subscales

Zero-Order Correlations

To further understand the relationships of PN and PF with PIU, we assessed the possible associations with the GPIUS2 subscales using Kendall's rank correlation coefficient. The results were then converted to Pearson's r following Walker (2003) (Table 4.3). Zero-order correlations showed that PN is moderately correlated with POSI; PN and PF are moderately correlated with MR; PF is weakly correlated with CP; PN is weakly correlated with CIU and NO.

The GPIUS2 subscales were then correlated with age and gender. Results show that age correlated weakly and negatively with CIU ($\tau = -.15, r = -.23, p < .05$), while gender correlated moderately and positively with MR ($\tau = .26, r = .40, p < .001$): women ($M = 15.59, SD = 4.99$) scored higher on MR than men ($M = 12.24, SD = 5.26$).

Table 4.3: Converted Pearson correlations (from rank order correlations) between PN, PF and GPIUS2CS subscales.

	PN	**PF**
POSI ($M = 9.04, SD = 4.80$)	.35***	.17
MR ($M = 13.81, SD = 5.39$)	.40***	.29**
CP ($M = 9.32, SD = 5.03$)	.14	.22*
CIU ($M = 10.20, SD = 5.27$)	.20*	.14
NO ($M = 6.03, SD = 4.41$)	.25**	.09

Note: $n = 149$, * $p < .05$. ** $p < .01$. *** $p < .001$.

Linear Regressions

The zero-order correlations in Table 4.3 indicate that PN and PF are correlated with GPIUS2 subscales. We further analyzed these relationships with linear regressions (Table 4). Moreover, we carried out multiple regressions (Table 4.4) when significant age and gender correlations were found with GPIUS2 subscales. Prior to each analysis, data were subjected to the Kolmogorov-Smirnov test which revealed a deviation from normality for all GPIUS2 subscales. Data were then transformed with a square root transformation to make the distributions normal or near-normal and to satisfy the linear or multiple regression assumptions. The linear regression involving PN and NO is not reported because two underlying assumptions were violated. Results indicated that PN is a predictor of POSI ($F(1, 147) = 15.87, p < .001$); PN and gender are predictors of MR ($F(2,146) = 18.34, p < .001$); PF and gender are predictors of MR ($F(2,146) = 13.27, p < .001$); PF is a predictor of CP ($F(1,147) = 6.09, p < .05$); PN and age are predictors of CIU ($F(2,146) = 6.21, p < .01$).

Table 4.4: Regression parameters for GPIUS2 subscales.

Dependent variable	**Adjusted** R^2	**Predictor**	β	**Stand.** β	**t**	**p**
POSI	.09	PN	.14	.31	3.98	$< .001$
MR	.19	PN	.87	.32	4.32	$< .001$
		Gender	1.08	.29	3.86	$< .001$
MR	.14	PF	.80	.24	3.08	$< .01$
		Gender	1.09	.29	3.78	$< .001$
CP*	.03	PF	.01	.20	2.47	$< .05$
CIU	.07	PN	.04	.16	2.00	$< .05$
		Age	-.01	-.21	-2.59	$< .05$

* The Kolmogorov-Smirnov test revealed a slight deviation from normality for residuals distribution ($p = .02$).

4.3.4 Discussion

The data collected by Psicoscrigno shows that a high number of Facebook users answered the questionnaires that we included into it. This result highlights the feasibility of using research in the large for distributing a health app that offers people reliable assessments of mental health conditions. Moreover, research in the large allowed us to carry out a correlational study, whose purpose was to examine the possible relationship between time perspective and problematic Internet use. The obtained results met our hypothesis: the PN and PF time frames were found to be correlated with PIU. Moreover, two regression analysis pointed out that PN and PF are predictors of PIU.

The study showed that TP is an individual difference that plays a significant role in PIU, in particular for people who have a PN or PF orientation. Individuals with such orientations might be more willing to use the Internet to alleviate their negative moods than individuals with a different time orientation: in particular, when correlating TP with the GPIUS2 subscales, the larger correlation value was obtained between PN and MR. Further analysis showed also that women had higher values on the MR subscale than men. As suggested by Larsen (2000), women use strategies that rely on social interactions for mood regulation more frequently and successfully than men. The fact that these strategies can rely on the use of the Internet and social networks might contribute to explain the higher MR values for women in the GPIUS2.

The other moderate correlation concerned PN and POSI. Individuals with a high PN value present a potentially disturbing portrait with minimal and unsatisfactory interpersonal relationships (Zimbardo and Boyd, 1999). This might contribute to explain the correlation with POSI: individuals with a past negative orientation could find the Internet to be a safer and better place for social interactions.

Given the negative outcomes that PIU could lead individuals to, in particular adolescents (Hawi, 2012; Moreno et al., 2011), TP should be included in efforts devoted to better understand PIU. Moreover, TP should be considered when studying people use of social network sites, such as Facebook, considering that they could contribute to PIU (Kittinger et al., 2012). More generally, the findings of this study provide additional support to the usefulness of TP in predicting problematic behaviors.

Conclusions

In this thesis, we have explored how some of the HCI methods, i.e. quantitative as well qualitative, in the lab as well as in situ, and a recently proposed approach, i.e. research in the large, can be applied for proposing effective health apps that target the general population. In particular, we presented three case studies in different domains, i.e. (i) mindfulness, (ii) neurological assessment, and (iii) psychological assessment, and proposed an app for each domain.

In the following, we summarize the results of our research and discuss future work directions for each one of the proposed health apps. Finally, we discuss the strengths and weaknesses of using research in the large that emerged from the studies with our proposed apps.

C.1 Design and Evaluation of Mobile Mindfulness Apps

In Chapter 2, this thesis developed a mobile app (AEON) to support people practice a mindfulness technique, i.e. distancing from thoughts, that requires individuals to be aware of their thoughts and to observe them while they pass by, without acting or grasping on them, and without trying to suppress them. We have illustrated the design process that led to the development of AEON, by analyzing previously proposed approaches aimed at helping people practice mindfulness as well as understanding the difficulties that individuals can encounter when approaching it. Then, we carried out a quantitative lab study to compare our proposed app with two traditional distancing from thoughts techniques not based on technology. Such study methodology allowed us to assess whether AEON could actually be beneficial to people, comparing its effect with two well-known techniques whose efficacy is known.

Results of the study pointed out that AEON was able to obtain better results in terms of achieved mindfulness, level of difficulty and degree of pleasantness than the two traditional techniques. It was also the preferred approach for users. To the best of our knowledge, this study is the first to provide evidence that a mobile app could be beneficial in helping users practice mindfulness. Moreover, it allowed us to answer the research question *RQ1*. We then broadened the investigation of the effectiveness of AEON. In particular, we aimed at deeply understanding its long-term effects when used by people in their everyday contexts. According to (Klasnja et al., 2011), this understanding, together with the design knowledge that results from it, is arguably the biggest contribution that HCI can make to the development of effective health-related systems.

We thus conducted a qualitative in situ long-term study. In particular, since mindfulness could be helpful in reducing some negative emotional states, such as worry and ruminative thinking (Querstret and Cropley, 2013), we investigated whether the long-term use of our app could help users ameliorate their worries. This study approach allowed us to thoroughly examine users' experience with AEON over a prolonged period of time, as well as gain deep insights on their perception in using the app for ameliorating worry. Moreover, as the app was used in participants' everyday contexts, the study allowed us to discover possible different patterns of use and design opportunities. Interestingly, results reveal that several participants experienced a mindfulness state and a reduction of their worries when using the app, suggesting that the prolonged usage of AEON might be beneficial for people. Such findings allowed us to partially answer the research question *RQ2* and motivated us to pursue the investigation of the long-term effects of AEON when used by people in their everyday contexts. To this purpose, we employed a quantitative approach and the research in the large paradigm to carry out an in situ 4-week study. Such study methodology allowed us to assess the effectiveness of the app with a wider sample of people than the other two user studies of AEON that we carried out. Moreover, it contributed to increase external and ecological validity (Miller, 2012). Overall, this study allowed us to address the research questions *RQ4* and *RQ6*. To the best of our knowledge, no study so far employed research in the large to evaluate a mobile mindfulness app. Interestingly, results pointed out that AEON helped participants increase their level of mindfulness during the period of the study. Thus, together with the qualitative in situ evaluation, this study allowed us to fully answer the research question *RQ2*.

Given the positive outcomes of our evaluations of AEON, in future studies we will extend the investigation of the possible effects of the app to the broader context of supporting physical and psychological health, in line with traditional mindfulness interventions in clinical settings (Keng et al., 2011). As a specific example, we have started collaborating with sexologists to employ AEON as an adjunct in the treatment of sexual dysfunctions, such as men's premature ejaculation and women's vaginal anorgasmia. A positive role that AEON might play in such therapeutic interventions is that of helping patients distance themselves from the distressful recurring thoughts that are typically associated with their particular sexual dysfunction.

C.2 Design and Evaluation of Mobile Health Apps for Neurological Assessment

While Chapter 2 developed a mindfulness app, in Chapter 3 we employed qualitative and quantitative studies to design a mobile gamified app (MotorBrain) for neurological assessment. MotorBrain offers computerized versions of six motor tests traditionally used by neurologists for the assessment of users' motor skills or impairments. In particular, we applied gamification to turn such tests into simple games, with the aim of making them attractive and pleasant to execute for users. In addition to providing a preliminary assessment of their motor performance to users, MotorBrain sends the acquired data to a remote server, allowing us to build a database of individual motor skills data, which can be used for normative neurological studies. For example, the collected data can be used for studying the aging of human motor performance and also in the diagnosis of movement disorders or in the rehabilitation process of a pathological motor function. To the best of our knowledge, MotorBrain will be the first publicly available health app to achieve the above described goals.

The thesis has illustrated the design process that led to the first prototype of MotorBrain. However, in this case, we could not ground the design process by analyzing previously proposed approaches, as we did with AEON. Thus, given the novelty of applying gamification in the neurological context, we first conducted a qualitative lab study to evaluate users' experience with the app, instead of a quantitative one. The study methodology we employed allowed us to assess the feasibility of our approach, i.e. applying gamification for turning motor tests into games, as well as identifying the factors that could interfere with the main purpose of the application, i.e. the collection of meaningful motor skills data. We thus modified the app by taking into account such problems, designing a second prototype of MotorBrain. Then, we conducted a quantitative pilot study to evaluate whether the data collected by MotorBrain can be effectively used for describing the aging of human motor performance. Such study approach allowed us to assess the usefulness of the collected data, since the obtained results show that it is suitable to make meaningful distinctions among different kinds of performance, and were able to highlight performance losses due to aging. Thus, they allowed us to answer the research question *RQ3*. Moreover, they are consistent with similar studies in the literature that employed digitizing tablets, desktop computers, or ad-hoc surfaces, e.g. (Raw et al., 2012; Welsh et al., 2007). Therefore, the results we obtained are promising with respect to the next goal of the research, i.e. to collect and analyze large datasets of human motor performance by making the MotorBrain app freely available to the general public through on-line app stores.

It is interesting to note that an instrument that is able to detect motor performance losses such as MotorBrain might be useful to detect the presence of movement disorders when they are still in a subclinical stage and cannot be detected by physician's clinical observation. This will be explored in our future work by conducting detailed clinical assessments of subjects who turn to be outliers in their age group. We will also study the possibility to use the app as an instrument in rehabilitation: patients could perform the tasks repeatedly over a period of time and the physician could measure possible motor learning obtained through such training.

C.3 Using Research in the Large with an On-Line Health App for Psychological Assessment

Unlike the previous two case studies, Chapter 4 developed a desktop on-line app (Psicoscrigno) that focuses on the domain of psychological assessment, and in particular, on problematic Internet use (PIU).

To develop the app, we first analyzed the literature to find a valid and reliable questionnaire for PIU assessment and included it into Psicoscrigno. Given the simplicity of the app, we did not carry out user centered evaluations of it as we did with AEON and MotorBrain. Instead, we exploited the possibility offered by research in the large to made it directly available on Facebook and conduct a correlational study to better describing PIU in terms of personality traits. To this purpose, we included into the app an assessment of time perspective (TP). Such research paradigm allowed us to offer valid and reliable assessments of both PIU and TP to a possibly large number of users, i.e. all Italian Facebook users, as well as studying the effects of TP in PIU. Indeed, results pointed out that individuals with particular time perspectives might be more inclined to experience a problematic Internet use than individuals with other time perspectives. Overall, this case study allowed us to address the research questions *RQ5* and *RQ6*.

We are now extending Psicoscrigno with other reliable and valid questionnaires. In particular, considering our research interests on mindfulness and worry (see Chapter 2), we are including into Psicoscrigno two validated questionnaire to assess them, i.e. the Mindfulness Attention Awareness Scale (Brown and Ryan, 2003) and the Penn State Worry Questionnaire (Meyer et al., 1990). This will allow us to exploit research in the large for carrying out correlational studies to describe PIU in terms of mindfulness as well as examine if worry might play a role in the relationships between PIU and TP. Indeed, studies in the literature have found that both constructs are related with anxiety, e.g. (Caplan, 2007; Zimbardo and Boyd, 1999), and a high level of worry is central in such disorder (American Psychiatric Association, 2013).

C.4 Strengths and Weaknesses of Using Research in the Large

In this thesis, we employed research in the large for carrying out scientific studies with two of our proposed health apps, i.e. AEON and Psicoscrigno. Such studies differed both in terms of: (i) goal and length, i.e. studying the long-term effect of the usage of the app on people (AEON) and carrying out a correlational study (Psicoscrigno), and (ii) the mechanisms we exploited for distributing such apps, i.e. on-line app stores for AEON and Facebook for Psicoscrigno. Overall, the results we obtained from both studies confirmed some strengths and weaknesses of research in the large (see Section 1.2) that we discuss in the following.

Focusing our attention to the goal and length of the two studies, Psicoscrigno required a short use to people to answer each questionnaire and collect data for the correlational study. Results confirmed the feasibility of using research in the large for studying characteristics of people with higher levels of external and ecological validity compared to common research settings (Wilson et al., 2012). The study of AEON in the large required instead a long-term continuative use. Although we could evaluate the effectiveness of the app when used by people in their everyday contexts and extend the sample of our previous studies of the app, on the other hand results revealed that only a small percentage of the initial participants reached the end of the study. In general, this limitation is typical of long-term evaluations carried out with research in the large and suggests that new mechanisms for keeping participants engaged in the long-term must be explored. In our study, we used the unlocking of new app features at the end of the study period as an incentive. For example, future studies should consider distributing rewards along the entire study period.

Focusing our attention on the distribution mechanisms that we employed, the integration of Psicoscrigno with Facebook allowed us to deal with a well-known limitation of research in the large, i.e. the collection of unverified demographic data (see Section 1.2.2.4). Indeed, Psicoscrigno asked people permission to access the age and gender they specified in their profiles, which tend to be accurate (Wilson et al., 2012). In the case of AEON, since users could enter their personal thoughts in the app, we did not exploit such integration as an additional measure to guarantee their privacy. In general, the integration of a research app with Facebook should be considered when demographic data is needed. The integration with Facebook could also allow a research app to exploit the principle of social facilitation, i.e. the fact that people are more likely to perform a target behavior if they can discern via technology that others are performing the behavior along with them (Fogg, 2002), and thus persuade people to use more it. For example, in the case of AEON, Facebook could offer users the possibility to share their progress in the practice of mindfulness. However, we cannot exclude that such feature could lower people's perception of privacy, since they have to share an internal state. To explore this consideration, future studies could compare two versions of AEON, one with the integration with Facebook and one without it, and analyze which version of the app has been used more by people.

A

Appendix 1

A.1 Interview Protocol

Thank you very much for coming back. Now, I am going to ask you some questions concerning the period during which you used AEON. Please, feel free to tell me anything that came into your mind when you were using the app...

1. What did you think when you were using the app?
2. How did you feel when you were using the app?
3. Did you notice anything new or unusual in your days or in yourself during the period of the study?
4. Did you relate to your worries or think about them differently during this period?
5. Now that you are familiar with the app, what do you think about it?
6. Is there anything that you would change or improve in the app?

Bibliography

21awake (2014). buddhify 2 - mindfulness & meditation for modern life (Version 2.0) [Mobile application software]. Available from: https://itunes.apple.com/us/app/buddhify-2/id687421118.

Abroms, L. C., Padmanabhan, N., Thaweethai, L., and Phillips, T. (2011). iPhone apps for smoking cessation: A content analysis. *American Journal of Preventive Medicine*, 40(3):279–285.

Adams, J. and Nettle, D. (2009). Time perspective, personality and smoking, body mass, and physical activity: An empirical study. *British Journal of Health Psychology*, 14(1):83–105.

Agostino, R., Berardelli, A., Formica, A., Accornero, N., and Manfredi, M. (1992). Sequential arm movements in patients with parkinson's disease, huntington's disease and dystonia. *Brain*, 115(5):1481–1495.

Ahtinen, A., Mattila, E., Välkkynen, P., Kaipainen, K., Vanhala, T., Ermes, M., Sairanen, E., Myllymäki, T., and Lappalainen, R. (2013). Mobile mental wellness training for stress management: Feasibility and design implications based on a one-month field study. *JMIR Mhealth and Uhealth*, 1(2):e11.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2):179–211.

AllTheTests.com (2014). AllTheTests.com [Web application]. Available from: http://www.allthetests.com/quiz23/quiz/1198496255/Are-you-addicted-to-the-Internet.

American Psychiatric Association (2013). *The Diagnostic and statistical manual of mental disorders: DSM 5*. American Psychiatric Publishing.

Andersson, G., Ljótsson, B., and Weise, C. (2011). Internet-delivered treatment to promote health. *Current Opinion in Psychiatry*, 24(2):168–172.

Araaya (2012). Let Go (Version 1.8) [Mobile application software]. Available from: https://play.google.com/store/apps/details?id=com.araaya.letgo&hl=en.

Azar, K. M. J., Lesser, L. I., Laing, B. Y., Stephens, J., Aurora, M. S., Burke, L. E., and Palaniappan, L. P. (2013). Mobile applications for weight management: Theory-based content analysis. *American Journal of Preventive Medicine*, 45(5):583–589.

Baños, R., Etchemendy, E., Castilla, D., García-Palacios, A., Quero, S., and Botella, C. (2012). Positive mood induction procedures for virtual environments designed for elderly people. *Interacting with Computers*, 24(3):131–138.

Baer, R. A. (2003). Mindfulness training as a clinical intervention: A conceptual and empirical review. *Clinical Psychology: Science and Practice*, 10(2):125–143.

Bain, P. G. and Findley, L. J. (1993). *Assessing tremor severity: A clinical handbook*. London: Smith-Gordon.

Barak, A., Klein, B., and Proudfoot, J. G. (2009). Defining Internet-supported therapeutic interventions. *Annals of Behavioral Medicine*, 38(1):4–17.

Bates, S. C. and Marquit, J. (2011). Space psychology: Natural elements in habitation design. *Personal and Ubiquitous Computing*, 15(5):519–523.

Bautmans, I., Vantieghem, S., Gorus, E., Grazzini, Y.-R., Fierens, Y., Pool-Goudzwaard, A., and Mets, T. (2011). Age-related differences in pre-movement antagonist muscle co-activation and reaction-time performance. *Experimental Gerontology*, 46(8):637–642.

Bender, J. L., Yue, R. Y. K., To, M. J., Deacken, L., and Jadad, A. R. (2013). A lot of action, but not in the right direction: Systematic review and content analysis of smartphone applications for the prevention, detection, and management of cancer. *Journal of Medical Internet Research*, 15(12):e287.

Bentley, F., Tollmar, K., Stephenson, P., Levy, L., Jones, B., Robertson, S., Price, E., Catrambone, R., and Wilson, J. (2013). Health Mashups: Presenting statistical patterns between wellbeing data and context in natural language to promote behavior change. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 20(5):30.

Berman, M. G., Jonides, J., and Kaplan, S. (2008). The cognitive benefits of interacting with nature. *Psychological Science*, 19(12):1207–1212.

Bishop, S. R., Lau, M., Shapiro, S., Carlson, L., Anderson, N. D., Carmody, J., Segal, Z. V., Abbey, S., Speca, M., Velting, D., and Devins, G. (2004). Mindfulness: A proposed operational definition. *Clinical Psychology: Science and Practice*, 11(3):230–241.

Błachnio, A., Przepiórka, A., and Rudnicka, P. (2013). Psychological determinants of using Facebook: A research review. *International Journal of Human-Computer Interaction*, 29(11):775–787.

Bless, J. J., Westerhausen, R., Arciuli, J., Kompus, K., Gudmundsen, M., and Hugdahl, K. (2013). "Right on all occasions?" - On the feasibility of laterality research using a smartphone dichotic listening application. *Frontiers in Psychology*, 4.

Blunck, H., Bouvin, N. O., Franke, T., Grø nbæk, K., Kjaergaard, M. B., Lukowicz, P., and Wüstenberg, M. (2013). On heterogeneity in mobile sensing applications aiming at representative data collection. In *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication (UbiComp '13)*, pages 1087–1098. ACM Press.

Boggs, J. M., Beck, A., Felder, J. N., Dimidjian, S., Metcalf, C. a., and Segal, Z. V. (2014). Web-based intervention in mindfulness meditation for reducing residual depressive symptoms and relapse prophylaxis: A qualitative study. *Journal of Medical Internet Research*, 16(3):e87.

Böhmer, M. and Krüger, A. (2014). A case study of research through the app store: Leveraging the system UI as a playing field for improving the design of smartphone launchers. *International Journal of Mobile Human Computer Interaction (IJMHCI)*, 6(2):32–45.

Boniwell, I. and Zimbardo, P. G. (2004). Balancing time perspective in pursuit of optimal functioning. In *Positive Psychology in Practice*, pages 165–178. John Wiley & Sons, Hoboken, NJ.

Boudreaux, E. D., Waring, M. E., Hayes, R. B., Sadasivam, R. S., Mullen, S., and Pagoto, S. (2014). Evaluating and selecting mobile health apps: Strategies for healthcare providers and healthcare organizations. *Translational Behavioral Medicine*, pages 1–9.

Bowers, M. (2011). The Shredder (Version 1.0) [Mobile application software]. Available from: http://itunes.apple.com/us/app/the-shredder/id473225371?mt=8.

Bradley, M. M. and Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1):49–59.

Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–110.

BreathTheWaves (2013). BreathTheWaves Full (Version 1.2) [Mobile application software]. Available from: https://play.google.com/store/apps/details?id=nl.breaththewaves.

Breton, E. R., Fuemmeler, B. F., and Abroms, L. C. (2011). Weight loss-there is an app for that! But does it adhere to evidence-informed practices? *Translational Behavioral Medicine*, 1(4):523–529.

Brown, K. W. and Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology*, 84(4):822–848.

Brown, K. W., Ryan, R. M., and Creswell, J. D. (2007). Mindfulness: Theoretical foundations and evidence for its salutary effects. *Psychological Inquiry*, 18(4):211–237.

Caplan, S. E. (2002). Problematic Internet use and psychosocial well-being: Development of a theory-based cognitive-behavioral measurement instrument. *Computers in Human Behavior*, 18(5):553–575.

Caplan, S. E. (2005). A social skill account of problematic Internet use. *Journal of Communication*, 55(4):721–736.

Caplan, S. E. (2007). Relations among loneliness, social anxiety, and problematic Internet use. *CyberPsychology & Behavior*, 10(2):234–242.

Caplan, S. E. (2010). Theory and measurement of generalized problematic Internet use: A two-step approach. *Computers in Human Behavior*, 26(5):1089–1097.

Carmody, J. and Baer, R. A. (2008). Relationships between mindfulness practice and levels of mindfulness, medical and psychological symptoms and well-being in a mindfulness-based stress reduction program. *Journal of Behavioral Medicine*, 31(1):23–33.

Chiesa, A. and Malinowski, P. (2011). Mindfulness-based approaches: Are they all the same? *Journal of Clinical Psychology*, 67(4):404–424.

Chiesa, A. and Serretti, A. (2011). Mindfulness based cognitive therapy for psychiatric disorders: A systematic review and meta-analysis. *Psychiatry Research*, 187(3):441–453.

Chittaro, L. and Vianello, A. (2013). Time perspective as a predictor of problematic Internet use: A study of Facebook users. *Personality and Individual Differences*, 55(8):989–993.

Chittaro, L. and Vianello, A. (2014). Computer-supported mindfulness: Evaluation of a mobile thought distancing application on naive meditators. *International Journal of Human-Computer Studies*, 72(3):337–348.

Chomutare, T., Fernandez-Luque, L., Å rsand, E., and Hartvigsen, G. (2011). Features of mobile diabetes applications: Review of the literature and analysis of current applications compared against evidence-based guidelines. *Journal of Medical Internet Research*, 13(3):e65.

Cicerale, A., Ambron, E., Lingnau, A., and Rumiati, R. I. (2014). A kinematic analysis of age-related changes in grasping to use and grasping to move common objects. *Acta Psychologica*, 151:134–142.

Cohn, A. M., Hunter-Reel, D., Hagman, B. T., and Mitchell, J. (2011). Promoting behavior change from alcohol use through mobile technology: The future of ecological momentary assessment. *Alcoholism: Clinical and Experimental Research*, 35(12):2209–2215.

Collyer, C. E., Broadbent, H. A., and Church, R. M. (1994). Preferred rates of repetitive tapping and categorical time production. *Perception & Psychophysics*, 55(4):443–453.

Consumer Health Information Corporation (2011). Motivating patients to use smartphone health apps. Retrieved from: http://www.prweb. com/releases/2011/04/prweb5268884.htm (last accessed on 15 December 2014).

Contreras-Vidal, J. L., Teulings, H. L., and Stelmach, G. E. (1998). Elderly subjects are impaired in spatial coordination in fine motor control. *Acta Psychologica*, 100(1):25–35.

Coolican, H. (2009). *Research methods and statistics in psychology*. Routledge.

Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., Leaver-Fay, A., Baker, D., Popović, Z., and Players, F. (2010). Predicting protein structures with a multiplayer online game. *Nature*, 466(7307):756–760.

Costa, P. T. and McCrae, R. R. (1992). Four ways five factors are basic. *Personality and Individual Differences*, 13(6):653–665.

Coulton, P. and Bamford, W. (2011). Experimenting through mobile "apps" and "app stores". *International Journal of Mobile Human Computer Interaction (IJMHCI)*, 3(4):55–70.

CounsellingResource.com (2014). Counselling Resource [Web application]. Available from: http://counsellingresource.com/lib/quizzes/net-addiction-testing/internet-addiction/.

Cowan, L. T., Van Wagenen, S. A., Brown, B. A., Hedin, R. J., Seino-Stephan, Y., Hall, P. C., and West, J. H. (2012). Apps of steel: Are exercise apps providing consumers with realistic expectations? A content analysis of exercise apps for presence of behavior change theory. *Health Education & Behavior*, 40(2):133–139.

Cramer, H., Rost, M., Belloni, N., Bentley, F., and Chincholle, D. (2010). Research in the large. using app stores, markets, and other wide distribution channels in Ubicomp research. In *Proceedings of the 12th ACM International Conference Adjunct Papers on Ubiquitous Computing-Adjunct*, pages 511–514, New York, New York, USA. ACM Press.

de la Vega, R. and Miró, J. (2014). mHealth: A strategic field without a solid scientific soul. A systematic review of pain-related apps. *PloS One*, 9(7):e101312.

DeCuir-Gunby, J. T., Marshall, P. L., and McCulloch, A. W. (2011). Developing and Using a Codebook for the Analysis of Interview Data: An Example from a Professional Development Research Project. *Field Methods*, 23(2):136–155.

Deterding, S., Dixon, D., Khaled, R., and Nacke, L. (2011). From game design elements to gamefulness: Defining gamification. In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, pages 9–15. ACM Press.

Doan, A., Ramakrishnan, R., and Halevy, A. Y. (2011). Crowdsourcing systems on the world-wide web. *Communications of the ACM*, 54(4):86–96.

Donker, T., Petrie, K., Proudfoot, J., Clarke, J., Birch, M.-R., and Christensen, H. (2013). Smartphones for smarter delivery of mental health programs: a systematic review. *Journal of Medical Internet Research*, 15(11):e247.

Dontworrycompany (2011). Worrydoll lite (Version 1.01) [Mobile application software]. Available from: http://itunes.apple.com/us/app/worrydoll-lite/id482387311?mt=8.

Drake, L., Duncan, E., Sutherland, F., Abernethy, C., and Henry, C. (2008). Time perspective and correlates of wellbeing. *Time & Society*, 17(1):47–61.

Dufau, S., Duñabeitia, J. A., Moret-Tatay, C., McGonigal, A., Peeters, D., Alario, F.-X., Balota, D. a., Brysbaert, M., Carreiras, M., Ferrand, L., Ktori, M., Perea, M., Rastle, K., Sasburg, O., Yap, M. J., Ziegler, J. C., and Grainger, J. (2011). Smart phone, smart science: How the use of smartphones can revolutionize research in cognitive science. *PloS One*, 6(9):e24974.

Dunkel, C. S. and Weber, J. L. (2010). Using three levels of personality to predict time perspective. *Current Psychology*, 29(2):95–103.

El-Gayar, O., Timsina, P., Nawar, N., and Eid, W. (2013). Mobile applications for diabetes self-management: Status and potential. *Journal of Diabetes Science and Technology*, 7(1):247–262.

EMarketer (2014). Worldwide Internet, social network and mobile users. Technical report, EMarketer.

Embodied (2014). Worry Bubble (Version 1.01) [Mobile application software]. Available from: https://itunes.apple.com/us/app/worry-bubble/id889560057?mt=8.

Epel, E. S., Bandura, A., and Zimbardo, P. G. (1999). Escaping homelessness: The influences of self-efficacy and time perspective on coping with homelessness. *Journal of Applied Social Psychology*, 29(3):575–596.

Erisman, S. M. and Roemer, L. (2010). Induced mindfulness on emotional responding to film clips. *Emotion*, 10(1):72–82.

Eysenbach, G. (2001). What is e-health? *Journal of Medical Internet Research*, 3(2):e20.

Feldman, G., Greeson, J., and Senville, J. (2010). Differential effects of mindful breathing, progressive muscle relaxation, and loving-kindness meditation on decentering and negative reactions to repetitive thoughts. *Behaviour Research and Therapy*, 48(10):1002–1011.

Fernandez-Luque, L., Karlsen, R., Krogstad, T., Burkow, T. M., and Vognild, L. K. (2010). Personalized health applications in the Web 2.0: The emergence of a new approach. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 1053–1056. IEEE.

Ferrari, J. and Diazmorales, J. (2007). Procrastination: Different time orientations reflect different motives. *Journal of Research in Personality*, 41(3):707–714.

Ferreira, D., Kostakos, V., and Dey, A. K. (2012). Lessons learned from large-scale user studies: Using Android market as a source of data. *International Journal of Mobile Human Computer Interaction (IJMHCI)*, 4(3):28–43.

Fiordelli, M., Diviani, N., and Schulz, P. J. (2013). Mapping mHealth research: A decade of evolution. *Journal of Medical Internet Research*, 15(5):e95.

Fleiss, J. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378–382.

Fogg, B. J. (2002). *Persuasive technology: Using computers to change what we think and do*. Morgan Kaufmann.

Fogg, B. J. (2009). A behavior model for persuasive design. In *Proceedings of the 4th International Conference on Persuasive Technology (Persuasive '09)*, page 40, New York, New York, USA. ACM Press.

Free, C., Phillips, G., Felix, L., Galli, L., Patel, V., and Edwards, P. (2010). The effectiveness of M-health technologies for improving health and health services: A systematic review protocol. *BMC Research Notes*, 3(1):250.

Free, C., Phillips, G., Watson, L., Galli, L., Felix, L., Edwards, P., Patel, V., and Haines, A. (2013). The effectiveness of mobile-health technologies to improve health care service delivery processes: A systematic review and meta-analysis. *PLoS Medicine*, 10(1):e1001363.

Fresco, D. M., Moore, M. T., van Dulmen, M. H. M., Segal, Z. V., Ma, S. H., Teasdale, J. D., and Williams, J. M. G. (2007). Initial psychometric properties of the experiences questionnaire: Validation of a self-report measure of decentering. *Behavior Therapy*, 38(3):234–246.

Glück, T. M. and Maercker, A. (2011). A randomized controlled pilot study of a brief web-based mindfulness training. *BMC Psychiatry*, 11(1):175.

Goblue International, L. (2014). Craving to Quit (Version 3.0.3) [Mobile application software]. Available from: https://itunes.apple.com/us/app/craving-to-quit/id585558426?mt=8.

Good, B. M. and Su, A. I. (2011). Games with a scientific purpose. *Genome Biology*, 12(12):135.

Good, B. M. and Su, A. I. (2013). Crowdsourcing for bioinformatics. *Bioinformatics*, 29(16):1925–1933.

GoToQuiz.com (2014). GoToQuiz.com [Web application]. Available from: http://www.gotoquiz.com/do_you_have_internet_addiction.

Grajales III, F. J., Sheps, S., Ho, K., Novak-Lauscher, H., and Eysenbach, G. (2014). Social media: A review and tutorial of applications in medicine and health care. *Journal of Medical Internet Research*, 16(2):e13.

Green, L. W. and Kreuter, M. W. (1991). Health promotion planning: An educational and environmental approach. In *Health promotion planning: An educational and environmental approach*. Mayfield, Mountain View, CA.

Griffiths, F., Lindenmeyer, A., Powell, J., Lowe, P., and Thorogood, M. (2006). Why are health care interventions delivered over the internet? A systematic review of the published literature. *Journal of Medical Internet Research*, 8(2):e10.

Gunaratana, H. (2002). *Mindfulness in plain english*. Wisdom Publications, Boston.

Hartnett, G. (2013). ThoughtBox (Version 1.0) [Mobile application software]. Available from: https://itunes.apple.com/us/app/thoughtbox/id688306970?mt=8.

Hawi, N. S. (2012). Internet addiction among adolescents in Lebanon. *Computers in Human Behavior*, 28(3):1044–1053.

Hayes, S. (2005). *Get out of you life and into your life: The new acceptance and commitment therapy*. New Harbinger Publications, Oakland, CA.

Hayes, S., Jacobson, N. S., Follette, V., and Dougher, M. J. (1994). *Acceptance and change. Content and context in psychotherapy*. Context Press, Reno, NV.

Hayes, S. C., Luoma, J. B., Bond, F. W., Masuda, A., and Lillis, J. (2006). Acceptance and commitment therapy: Model, processes and outcomes. *Behaviour Research and Therapy*, 44(1):1–25.

Hayes, S. C. and Shenk, C. (2004). Operationalizing mindfulness without unnecessary attachments. *Clinical Psychology: Science and Practice*, 11(3):249–254.

Hayes, S. C., Strosahl, K. D., and Wilson, K. G. (1999). *Acceptance and commitment therapy*. Guilford Press, New York.

HeadSpace (2014). Headspace.com - meditation & mindfulness (Version 2.0.1) [Mobile application software]. Available from: https://itunes.apple.com/us/app/headspace.com-meditation-mindfulness/id493145008?mt=8.

HealthyPlace.com (2014). HealthyPlace [Web application]. Available from: http://www.healthyplace.com/psychological-tests/internet-addiction-test/.

Henze, N. and Pielot, M. (2013). App stores: External validity for mobile HCI. *interactions*, 20(2):33–38.

Henze, N., Pielot, M., Poppinga, B., Schinke, T., and Boll, S. (2011a). My app is an experiment: Experience from user studies in mobile app stores. *International Journal of Mobile Human Computer Interaction (IJMHCI)*, 3(4):71–91.

Henze, N., Rukzio, E., and Boll, S. (2011b). 100,000,000 taps: Analysis and improvement of touch performance in the large. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '11)*, pages 133–142. ACM Press.

Henze, N., Sahami Shirazi, A., Schmidt, A., Pielot, M., and Michahelles, F. (2013). Empirical research through ubiquitous data collection. *Computer*, 46(6):74–76.

Hoehn, M. M. and Yahr, M. D. (1967). Parkinsonism: Onset, progression, and mortality. *Neurology*, 17(5):427–442.

Hubel, K. A., Reed, B., Yund, E. W., Herron, T. J., and Woods, D. L. (2013). Computerized measures of finger tapping: Effects of hand dominance, age, and sex. *Perceptual & Motor Skills*, 116(3):929–52.

Hund, P. M., Dowell, J., and Müller, K. (2014). Representation of time in digital calendars: An argument for a unified, continuous & multi-granular calendar view. *International Journal of Human-Computer Studies*, 72(1):1–11.

Hundert, A. S., Huguet, A., McGrath, P. J., Stinson, J. N., and Wheaton, M. (2014). Commercially available mobile phone headache diary apps: A systematic review. *JMIR Mhealth and Uhealth*, 2(3):e36.

Inside Outlier (2014). Let It Go – stress relief (Version 2.1.5) [Mobile application software]. Available from: https://play.google.com/store/apps/details?id=com.insideoutlier.letitgo&hl=en.

InteraXon (2014a). Muse Calm (Version 1.6.1) [Mobile application software]. Available from: https://itunes.apple.com/us/app/muse-calm/id849841170?mt=8.

InteraXon (2014b). Muse: The brain sensing headband Retrieved from: http://store.choosemuse.com/products/muse (last accessed on 15 December 2014).

Kabat-Zinn, J. (1990). *Full catastrophe living: Using the wisdom of your body and mind to face stress, pain, and illness*. Delta, New York.

Kabat-Zinn, J. (2003). Mindfulness-based interventions in context: Past, present, and future. *Clinical Psychology: Science and Practice*, 10(2):144–156.

Kabat-Zinn, J. (2005). *Coming to our senses: Healing ourselves and the world through mindfulness*. Hyperion, New York.

Kawrykow, A., Roumanis, G., Kam, A., Kwak, D., Leung, C., Wu, C., Zarour, E., Sarmenta, L., Blanchette, M., and Waldispühl, J. (2012). Phylo: A citizen science approach for improving multiple sequence alignment. *PloS One*, 7(3):e31362.

Keng, S. L., Smoski, M. J., and Robins, C. J. (2011). Effects of mindfulness on psychological health: A review of empirical studies. *Clinical Psychology Review*, 31(6):1041–1056.

Keough, K. A., Zimbardo, P. G., and Boyd, J. N. (1999). Who's smoking, drinking, and using drugs? Time perspective as a predictor of substance use. *Basic and Applied Social Psychology*, 21(2):149–164.

Keru (2011). Throw your worry away! (Version 1.2) [Mobile application software]. Available from: http://itunes.apple.com/us/app/throw-your-worry-away!/id437253188?mt=8.

Killingsworth, M. A. and Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science*, 330(6006):932–932.

Kittinger, R., Correia, C. J., and Irons, J. G. (2012). Relationship between Facebook use and problematic Internet use among college students. *Cyberpsychology, Behavior and Social Networking*, 15(6):324–327.

Klasnja, P., Consolvo, S., and Pratt, W. (2011). How to evaluate technologies for health behavior change in HCI research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*, pages 3063–3072. ACM Press.

Klasnja, P. and Pratt, W. (2012). Healthcare in the pocket: Mapping the space of mobile-phone health interventions. *Journal of Biomedical Informatics*, 45(1):184–198.

Klingeman, H. (2001). The time game: Temporal activity on emotional well-being among older australian women: Cross-sectional and longitudinal analysis. *Time & Society*, 10(2-3):303–328.

Kranz, M., Murmann, L., and Michahelles, F. (2013). Research in the large: Challenges for large-scale mobile application research - A Case study about NFC adoption using gamification via an app store. *International Journal of Mobile Human-Computer Interaction (IJMHCI)*, 5(1):45–61.

Kraut, R., Patterson, M., Lundmark, V., Kiesler, S., Mukopadhyay, T., and Scherlis, W. (1998). Internet paradox. A social technology that reduces social involvement and psychological well-being? *American Psychologist*, 53(9):1017–1031.

Kristjánsdóttir, Ó. B., Fors, E. A., Eide, E., Finset, A., van Dulmen, S., Wigers, S. H. r., and Eide, H. (2011). Written online situational feedback via mobile phone to support self-management of chronic widespread pain: A usability study of a Web-based intervention. *BMC Musculoskeletal Disorders*, 12(1):51.

Krusche, A., Cyhlarova, E., King, S., and Williams, J. M. G. (2012). Mindfulness online: A preliminary evaluation of the feasibility of a web-based mindfulness course and the impact on stress. *BMJ Open*, 2(3):e000803.

Krusche, A., Cyhlarova, E., and Williams, J. M. G. (2013). Mindfulness online: An evaluation of the feasibility of a web-based mindfulness course for stress, anxiety and depression. *BMJ Open*, 3(11):e003498.

Laconi, S., Rodgers, R. F., and Chabrol, H. (2014). The measurement of Internet addiction: A critical review of existing scales and their psychometric properties. *Computers in Human Behavior*, 41:190–202.

Laghi, F., Baiocco, R., D'Alessio, M., and Gurrieri, G. (2009). Suicidal ideation and time perspective in high school students. *European Psychiatry*, 24(1):41–46.

Lampe, C., Ellison, N., and Steinfield, C. (2007). A familiar face (book): Profile elements as signals in an online social network. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*, pages 435–444. ACM Press.

Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.

Lappalainen, R., Sairanen, E., Järvelä, E., Rantala, S., Korpela, R., Puttonen, S., Kujala, U. M., Myllymäki, T., Peuhkuri, K., Mattila, E., Kaipainen, K., Ahtinen, A., Karhunen, L., Pihlajamäki, J., Järnefelt, H., Laitinen, J., Kutinlahti, E., Saarelma, O., Ermes, M., and Kolehmainen, M. (2014). The effectiveness and applicability of different lifestyle interventions for enhancing wellbeing: The study design for a randomized controlled trial for persons with metabolic syndrome risk factors and psychological distress. *BMC Public Health*, 14:310.

Larsen, R. (2000). Toward a science of mood regulation. *Psychological Inquiry*, 11(3):129–141.

Lau, M., Bishop, S., Segal, Z., Buis, T., Anderson, N., Carlson, L., Shapiro, S., Carmody, J., Abbey, S., and Devins, G. (2006). The Toronto mindfulness scale: Development and validation. *Journal of Clinical Psychology*, 62(12):1445–1467.

Leahy, R. (2006). *The worry cure: Seven steps to stop worry from stopping you.* Three Rivers Press, New York.

Leversen, J. S. R., Haga, M., and Sigmundsson, H. (2012). From children to adults: Motor performance across the life-span. *PloS One*, 7(6):e38830.

Li, I., Dey, A. K., and Forlizzi, J. (2010). A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*, pages 557–566. ACM Press.

Li, I., Dey, A. K., and Forlizzi, J. (2011). Understanding my data, myself: Supporting self-reflection with ubicomp technologies. In *Proceedings of the 13th International Conference on Ubiquitous Computing (UbiComp '11)*, pages 405–414. ACM Press.

Lindley, C. A., Nacke, L., and Sennersten, C. C. (2008). Dissecting play - Investigating the cognitive and emotional motivations. In *Proceedings of 13th International Conference on Computer Games (CGames 2008)*, pages 9–16. University of Wolverhampton.

Linehan, M. (1993a). *Cognitive-behavioral treatment of borderline personality disorder*. Guilford Press, New York.

Linehan, M. (1993b). *Skills training manual for treating borderline personality disorder*. Guilford Press., New York.

Luengo-Oroz, M. A., Arranz, A., and Frean, J. (2012). Crowdsourcing malaria parasite quantification: An online game for analyzing images of infected thick blood smears. *Journal of Medical Internet Research*, 14(6):e167.

Lupton, D. (2014). Quantified sex: A critical analysis of sexual and reproductive self-tracking apps. *Culture, Health & Sexuality*, (ahead-of-print).

Ly, K. H., Trüschel, A., Jarl, L., Magnusson, S., Windahl, T., Johansson, R., Carlbring, P., and Andersson, G. (2014). Behavioural activation versus mindfulness-based guided self-help treatment administered through a smartphone application: A randomised controlled trial. *BMJ Open*, 4(1):e003440.

Lykins, E. L. B. and Baer, R. A. (2009). Psychological functioning in a sample of long-term practitioners of mindfulness meditation. *Journal of Cognitive Psychotherapy*, 23(3):226–241.

Malhotra, N. (2008). Completion time and response order effects in web surveys. *Public Opinion Quarterly*, 72(5):914–934.

Martínez-Pérez, B., de la Torre-Díez, I., and López-Coronado, M. (2013). Mobile health applications for the most prevalent conditions by the World Health Organization: Review and analysis. *Journal of Medical Internet Research*, 15(6):e120.

McMillan, D., Morrison, A., Brown, O., Hall, M., and Chalmers, M. (2010). Further into the wild: Running worldwide trials of mobile systems. In *Proceedings of the 8th International Conference on Pervasive Computing (Pervasive '10)*, pages 210–227. Springer Berlin Heidelberg.

McMillan, D., Morrison, A., and Chalmers, M. (2013). Categorised ethical guidelines for large scale mobile HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*, pages 1853–1862, New York, New York, USA. ACM Press.

Mental Health America (2014). Mental Health Screening Tools [Web application]. Available from: http://www.mentalhealthamerica.net/mental-health-screening-tools.

Meyer, T. J., Miller, M. L., Metzger, R. L., and Borkovec, T. D. (1990). Development and validation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy*, 28(6):487–495.

Middelweerd, A., Mollee, J. S., van der Wal, C., Brug, J., and Te Velde, S. J. (2014). Apps to promote physical activity among adults: A review and content analysis. *International Journal of Behavioral Nutrition and Physical Activity*, 11(1):97.

Miller, G. (2012). The smartphone psychology manifesto. *Perspectives on Psychological Science*, 7(3):221–237.

Miluzzo, E., Lane, N. D., Lu, H., and Campbell, A. T. (2010). Research in the app store era: Experiences from the cenceme app deployment on the iPhone. In *First Workshop on Research in the Large at UbiComp*. ACM Press.

MindApps (2014). The Mindfulness App (Version 1.7.3) [Mobile application software]. Available from: http://itunes.apple.com/it/app/the-mindfulness-app/id417071430?mt=8.

Mindfulapps (2012). Mindfulness TS (Version 2.1) [Mobile application software]. Available from: http://itunes.apple.com/en/app/mindfulness-ts/id451690670?mt=8.

Miralles, F., Tarongí, S., and Espino, A. (2006). Quantification of the drawing of an Archimedes spiral through the analysis of its digitized picture. *Journal of Neuroscience Methods*, 152(1-2):18–31.

Moorhead, S. A., Hazlett, D. E., Harrison, L., Carroll, J. K., Irwin, A., and Hoving, C. (2013). A new dimension of health care: Systematic review of the uses, benefits, and limitations of social media for health communication. *Journal of Medical Internet Research*, 15(4):e85.

Morahan-Martin, J. (1999). The relationship between loneliness and Internet use and abuse. *CyberPsychology & Behavior*, 2(5):431–439.

Morahan-Martin, J. (2008). Internet abuse: Emerging trends and lingering questions. In *Psychological Aspects of Cyberspace Theory Research Applications*, pages 32–69. Cambridge University Press.

Moreno, M. A., Jelenchick, L., Cox, E., Young, H., and Christakis, D. A. (2011). Problematic internet use among US youth: a systematic review. *Archives of Pediatrics & Adolescent Medicine*, 165(9):797–805.

Morris, M. E., Kathawala, Q., Leen, T. K., Gorenstein, E. E., Guilak, F., Labhard, M., and Deleeuw, W. (2010). Mobile therapy: Case study evaluations of a cell phone application for emotional self-awareness. *Journal of Medical Internet Research*, 12(2):e10.

Mosa, A. S. M., Yoo, I., and Sheets, L. (2012). A systematic review of healthcare applications for smartphones. *BMC Medical Informatics and Decision Making*, 12(1):67.

Muessig, K. E., Pike, E. C., Legrand, S., and Hightow-Weidman, L. B. (2013). Mobile phone applications for the care and prevention of HIV and other sexually transmitted diseases: A review. *Journal of Medical Internet Research*, 15(1):e1.

Myhealthapps.net (2014). My Health Apps. Retrieved from: http://myhealthapps.net (last accessed on 15 December 2014).

Nature Applications (2014). Let It Go: Vent and Release (Version 2.1) [Mobile application software]. Available from: https://itunes.apple.com/us/app/let-it-go-vent-and-release/id740828283?mt=8.

Nes, A. A. G., van Dulmen, S., Eide, E., Finset, A., Kristjánsdóttir, O. B., Steen, I. S. v., and Eide, H. (2012). The development and feasibility of a web-based intervention with diaries and situational feedback via smartphone to support self-management in patients with diabetes type 2. *Diabetes Research and Clinical Practice*, 97(3):385–393.

Orkin, J. and Roy, D. (2007). The Restaurant Game: Learning social behavior and language from thousands of players online. *Journal of Game Development*, 3(1):39–60.

Pagoto, S., Schneider, K., Jojic, M., DeBiasse, M., and Mann, D. (2013). Evidence-based strategies in weight-loss mobile apps. *American Journal of Preventive Medicine*, 45(5):576–582.

Perlman, D. M., Salomons, T. V., Davidson, R. J., and Lutz, A. (2010). Differential effects on pain intensity and unpleasantness of two meditation practices. *Emotion*, 10(1):65–71.

Plaza, I., Demarzo, M. M. P., Herrera-Mercadal, P., and García-Campayo, J. (2013). Mindfulness-based mobile applications: Literature review and analysis of current features. *JMIR Mhealth and Uhealth*, 1(2):e24.

Poppinga, B., Cramer, H., Böhmer, M., Morrison, A., Bentley, F., Henze, N., Rost, M., and Michahelles, F. (2012). Research in the Large 3.0: App stores, wide distribution, and big data in MobileHCI research. In *Proceedings of the 14th international Conference on Human-Computer Interaction with Mobile Devices and Services Companion (MobileHCI '12)*, pages 241–244. ACM Press.

Project Implicit (2014). Project Implicit Mental Health [Web application]. Available from: https://implicit.harvard.edu/implicit/user/pimh/index.jsp.

Proudfoot, J., Klein, B., Barak, A., Carlbring, P., Cuijpers, P., Lange, A., Ritterband, L., and Andersson, G. (2011). Establishing guidelines for executing and reporting Internet intervention research. *Cognitive Behaviour Therapy*, 40(2):82–97.

Psych Central (2014). Psych Central [Web application]. Available from: http://psychcentral.com/quizzes/netaddiction.htm.

Psychology-Tools.com (2014). Psychology Tools [Web application]. Available from: http://psychology-tools.com/internet-addiction-test/.

Pullman, S. L. (1998). Spiral analysis: A new technique for measuring tremor with a digitizing tablet. *Movement Disorders*, 13(S3):85–89.

Querstret, D. and Cropley, M. (2013). Assessing treatments used to reduce rumination and/or worry: A systematic review. *Clinical Psychology Review*, 33(8):996–1009.

Rabuffetti, M., Ferrarin, M., Spadone, R., Pellegatta, D., Gentileschi, V., Vallar, G., and Pedotti, A. (2002). Touch-screen system for assessing visuo-motor exploratory skills in neuropsychological disorders of spatial cognition. *Medical and Biological Engineering and Computing*, 40(6):675–686.

Raw, R. K., Kountouriotis, G. K., Mon-Williams, M., and Wilkie, R. M. (2012). Movement control in older adults: does old age mean middle of the road? *Journal of Experimental Psychology: Human Perception and Performance*, 38(3):735–745.

Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual & Motor Skills*, 8(3):271–276.

research2guidance (2013). The market for mHealth app services will reach $26 billion by 2017. Retrieved from: http://research2guidance.com/the-market-for-mhealth-app-services-will-reach-26-billion-by-2017/ (last accesssed on 15 December 2014).

Ritterband, L. M. and Tate, D. F. (2009). The science of Internet interventions. *Annals of Behavioral Medicine*, 38(1):1–3.

Rodriguez Fornies, H. (2014). Lose Weight - Mindfulness Meditation for weight loss and mindful eating (Version 2.3) [Mobile application software]. Available from: https://itunes.apple.com/us/app/lose-weight-mindfulness-meditation/id624565470?mt=8.

Rogers, Y. (2004). New theoretical approaches for human-computer interaction. *Annual Review of Information Science and Technology*, 38(1):87–143.

Rosenbaum, D. A. (2009). *Human motor control*. Academic Press.

Rosser, B. A. and Eccleston, C. (2011). Smartphone applications for pain management. *Journal of Telemedicine and Telecare*, 17(6):308–312.

Rothspan, S. and Read, S. J. (1996). Present versus future time perspective and HIV risk among heterosexual college students. *Health Psychology*, 15(2):131–134.

Runyan, J. D., Steenbergh, T. A., Bainbridge, C., Daugherty, D. A., Oke, L., and Fry, B. N. (2013). A smartphone ecological momentary assessment/intervention "app" for collecting real-time data and promoting self-awareness. *PloS One*, 8(8):e71325.

Saagara, L. (2014). Universal Breathing: Pranayama (Version 2.9.0) [Mobile application software]. Available from: https://itunes.apple.com/us/app/universal-breathing-pranayama/id433571551?mt=8.

Safran, J. and Segal, Z. V. (1996). *Interpersonal process in cognitive therapy*. Basic Books, New York.

Sager, C. (2014). Lotus Bud Mindfulness Bell (Version 2.3) [Mobile application software]. Available from: http://itunes.apple.com/it/app/lotus-bud-mindfulness-bell/id502329366?mt=8.

Sahami Shirazi, A., Henze, N., Dingler, T., Pielot, M., Weber, D., and Schmidt, A. (2014). Large-scale assessment of mobile notifications. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*, pages 3055–3064. ACM Press.

Sahami Shirazi, A., Rohs, M., Schleicher, R., Kratz, S., Müller, A., and Schmidt, A. (2011). Real-time nonverbal opinion sharing through mobile phones during sports events. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*, pages 307–310. ACM Press.

Sama, P. R., Eapen, Z. J., Weinfurt, K. P., Shah, B. R., and Schulman, K. A. (2014). An evaluation of mobile health application tools. *JMIR Mhealth and Uhealth*, 2(2):e19.

Saunders-Pullman, R., Derby, C., Stanley, K., Floyd, A., Bressman, S., Lipton, R. B., Deligtisch, A., Severt, L., Yu, Q., Kurtis, M., and Pullman, S. L. (2008). Validity of spiral analysis in early Parkinson's disease. *Movement Disorders*, 23(4):531–537.

Scaife, M. and Rogers, Y. (1996). External cognition: How do graphical representations work? *International Journal of Human-Computer Studies*, 45(2):185–213.

Schleicher, R., Sahami Shirazi, A., Rohs, M., Kratz, S., and Albrecht, S. (2011). WorldCupinion experiences with an Android app for real-time opinion sharing during soccer world cup games. *International Journal of Mobile Human Computer Interaction (IJMHCI)*, 3(4):18–35.

Schrope, M. (2013). Solving tough problems with games. *Proceedings of the National Academy of Sciences of the United States of America*, 110(18):7104–7106.

Segal, Z. V., Williams, J. M. G., and Teasdale, J. D. (2002). *Mindfulness-based cognitive therapy for depression: A new approach to preventing relapse*. Guilford Press, New York.

Seidler, R. D., Bernard, J. A., Burutolu, T. B., Fling, B. W., Gordon, M. T., Gwin, J. T., Kwak, Y., and Lipps, D. B. (2010). Motor control and aging: Links to age-related brain structural, functional, and biochemical effects. *Neuroscience & Biobehavioral Reviews*, 34(5):721–733.

serve-you.net (2014). QuizMeme [Web application]. Available from: http://www.quizmeme.com/internetaddict/quiz.php.

Shapiro, S. L., Carlson, L. E., Astin, J. A., and Freedman, B. (2006). Mechanisms of mindfulness. *Journal of Clinical Psychology*, 62(3):373–386.

Shaw, C., Gromala, D., and Seay, A. F. (2007). The Meditation Chamber: Enacting autonomic senses. In *Proc. of ENACTIVE/07*, volume 7, pages 405–408. ACROE Press.

Singer, A. R. and Dobson, K. S. (2007). An experimental investigation of the cognitive vulnerability to depression. *Behaviour Research and Therapy*, 45(3):563–575.

Siorpaes, K. and Hepp, M. (2008). Games with a purpose for the semantic web. *IEEE Intelligent Systems*, 23(3):50–60.

Small-n-Tall (2014). ZenFriend - Meditation Timer, Tracker, Log & Community for your Mindfulness, Zen, Zazen and Yoga Practice (Version 1.3.1) [Mobile application software]. Available from: https://itunes.apple.com/us/app/zenfriend-meditation-timer/id880004532?mt=8.

Speer, R., Krishnamurthy, J., Havasi, C., Smith, D., Lieberman, H., and Arnold, K. (2009). An interface for targeted collection of common sense knowledge using a mixture model. In *Proceedings of the 14th International Conference on Intelligent User Interfaces (IUI '09)*, pages 137–146. ACM Press.

Stafford, T. and Dewar, M. (2014). Testing theories of skill learning using a very large sample of online game players. *Psychological Science*, 25(2):511–518.

Stone, A. A., Shiffman, S., Schwartz, J. E., Broderick, J. E., and Hufford, M. R. (2003). Patient compliance with paper and electronic diaries. *Controlled Clinical Trials*, 24(2):182–199.

Surangsrirat, D. and Thanawattano, C. (2012). Android application for spiral analysis in Parkinson's Disease. In *Southeastcon, 2012 Proceedings of IEEE*, pages 1–6. IEEE.

Teasdale, J., Segal, Z. V., Williams, J. G., Ridgeway, V. A., Soulsby, J., and Lau, M. (2000). Prevention of relapse/recurrence in major depression by mindfulness-based cognitive therapy. *Journal of Consulting and Clinical Psychology*, 68(4):615–623.

Teasdale, J. D., Moore, R. G., Hayhurst, H., Pope, M., Williams, S., and Segal, Z. V. (2002). Metacognitive awareness and prevention of relapse in depression: Empirical evidence. *Journal of Consulting and Clinical Psychology*, 70(2):275–287.

Texifter (2007). CAT (Version 1.6.1) [Web application]. Available from: http://cat.ucsur.pitt.edu/default.aspx.

The Center for Internet Addiction Recovery (2014). Netaddiction [Web application]. Available from: http://netaddiction.com/internet-addiction-test/.

The Guardian.com (2013). Are you an Internet addict? Retrieved from http://www.theguardian.com/technology/quiz/2013/oct/07/are-you-an-internet-addict-quiz (last accessed on 17 December 2014).

Thieme, A., Wallace, J., Johnson, P., McCarthy, J., Lindley, S., Wright, P., Olivier, P., and Meyer, T. D. (2013). Design to promote mindfulness practice and sense of self for vulnerable women in secure hospital services. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*, pages 2647–2656. ACM Press.

Thompson, B. L. and Waltz, J. (2007). Everyday mindfulness and mindfulness meditation: Overlapping constructs or not? *Personality and Individual Differences*, 43(7):1875–1885.

Thompson, N. J., Walker, E. R., Obolensky, N., Winning, A., Barmon, C., Diiorio, C., and Compton, M. T. (2010). Distance delivery of mindfulness-based cognitive therapy for depression: Project UPLIFT. *Epilepsy & Behavior*, 19(3):247–254.

Tippett, W. J. and Sergio, L. E. (2006). Visuomotor integration is impaired in early stage Alzheimer's disease. *Brain Research*, 1102(1):92–102.

Tools for Peace (2014). Stop, Breathe & Think (Version 2.0) [Mobile application software]. Available from: https://itunes.apple.com/us/app/stop-breathe-think/id778848692.

Unity Technologies (2013). Unity (Version 4.3) [Software]. Available from: http://www.unity3d.com.

US Department of Health and Human Services (2008). *Health Education Curriculum Analysis Tool*. Centers for Disease Control and Prevention, Atlanta, GA.

Valtchanov, D., Barton, K. R., and Ellard, C. (2010). Restorative effects of virtual nature settings. *Cyberpsychology, Behavior, and Social Networking*, 13(5):503–512.

van Mechelen, D. M., Van Mechelen, W., and Verhagen, E. A. (2014). Prevention apps assessed against the available scientific evidence: A review. *British Journal of Sports Medicine*, 48(11):878–882.

Vidyarthi, J., Riecke, B., and Gromala, D. (2012). Sonic Cradle: Designing for an immersive experience of meditation by connecting respiration to music. In *Proceedings of the Designing Interactive Systems Conference (DIS '12)*, pages 408–417. ACM Press.

Vidyarthi, J. and Riecke, B. E. (2014). Interactively mediating experiences of mindfulness meditation. *International Journal of Human-Computer Studies*, 72(8-9):674–688.

von Ahn, L. and Dabbish, L. (2004). Labeling images with a computer game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*, pages 319–326, New York, New York, USA. ACM Press.

von Ahn, L. and Dabbish, L. (2008). Designing games with a purpose. *Communications of the ACM*, 51(8):58–67.

von Ahn, L., Kedia, M., and Blum, M. (2006). Verbosity: A game for collecting common-sense facts. In *Proceedings of the SIGCHI conference on Human Factors in computing systems (CHI '06)*, pages 75–78. ACM Press.

Walker, D. A. (2003). JMASM9: Converting Kendall's tau for correlational or meta-analytic analyses. *Journal of Modern Applied Statistical Methods*, 2(2):525–530.

Weaver, E. R., Horyniak, D. R., Jenkinson, R., Dietze, P., and Lim, M. S. (2013). "Let's get wasted!" and other apps: Characteristics, acceptability, and use of alcohol-related smartphone applications. *JMIR Mhealth and Uhealth*, 1(1):e9.

Webb, T. L., Joseph, J., Yardley, L., and Michie, S. (2010). Using the Internet to promote health behavior change: A systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. *Journal of Medical Internet Research*, 12(1):e4.

Weinstein, D. and bambooyou.com (2014). Interactive Quiz [Web application]. Available from: http://courses.cs.vt.edu/professionalism/Workplace/Addiction.test.html.

Wells, A. (2006). Detached mindfulness in cognitive therapy: A metacognitive analysis and ten techniques. *Journal of Rational-Emotive & Cognitive-Behavior Therapy*, 23(4):337–355.

Welsh, T. N., Higgins, L., and Elliott, D. (2007). Are there age-related differences in learning to optimize speed, accuracy, and energy expenditure? *Human Movement Science*, 26(6):892–912.

Werbach, K. and Hunter, D. (2012). *For the win: How game thinking can revolutionize your business*. Wharton Digital Press.

West, J. H., Hall, P. C., Hanson, C. L., Barnes, M. D., Giraud-Carrier, C., and Barrett, J. (2012). There's an app for that: Content analysis of paid health and fitness apps. *Journal of Medical Internet Research*, 14(3):e72.

Westin, J., Dougherty, M., Nyholm, D., and Groth, T. (2010a). A home environment test battery for status assessment in patients with advanced Parkinson's disease. *Computer Methods and Programs in Biomedicine*, 98(1):27–35.

Westin, J., Ghiamati, S., Memedi, M., Nyholm, D., Johansson, A., Dougherty, M., and Groth, T. (2010b). A new computer method for assessing drawing impairment in Parkinson's disease. *Journal of Neuroscience Methods*, 190(1):143–148.

Wilson, R. E., Gosling, S. D., and Graham, L. T. (2012). A review of Facebook research in the social sciences. *Perspectives on Psychological Science*, 7(3):203–220.

Wise, S. P. and Shadmehr, R. (2002). Motor control. In *Encyclopedia of the Human Brain*, volume 3, pages 1–21. Elsevier Science.

Wolf, J. A., Moreau, J. F., Akilov, O., Patton, T., English, J. C., Ho, J., and Ferris, L. K. (2013). Diagnostic inaccuracy of smartphone applications for melanoma detection. *JAMA Dermatology*, 149(4):422–426.

Wolffram, J. and TrueSelfSoft (2011). Just Let Go (Version 1.0) [Mobile application software]. Available from: http://itunes.apple.com/us/app/just-let-go/id439683407?mt=8.

Yan, T. and Tourangeau, R. (2008). Fast times and easy questions: The effects of age, experience and question complexity on web survey response times. *Applied Cognitive Psychology*, 22(1):51–68.

Yang, P. C., Chang, C. C., Chen, Y. L., Chiang, J. H., and Hung, G. C. L. (2014). Portable assessment of emotional status and support system. *Lecture Notes in Computer Science*, 8423:175–183.

York, J. and Pendharkar, P. C. (2004). Human–computer interaction issues for mobile computing in a variable work context. *International Journal of Human-Computer Studies*, 60(5-6):771–797.

Young, K. S. (1998a). *Caught in the Net: How to recognize the signs of Internet addiction and a winning strategy for recovery*. Wiley, New York.

Young, K. S. (1998b). Internet addiction: The emergence of a new clinical disorder. *CyberPsychology & Behavior*, 1(3):237–244.

Yu, M. C., Wu, H., Lee, M. S., and Hung, Y. P. (2012). Multimedia-assisted breathwalk-aware system. *IEEE Transactions on Biomedical Engineering*, 59(12):3276–3282.

Zichermann, G. and Cunningham, C. (2011). *Gamification by design: Implementing game mechanics in web and mobile apps*. O'Reilly Media, Sebastopol, CA.

Zimbardo, P. G. and Boyd, J. N. (1999). Putting time in perspective: A valid, reliable individual-differences metric. *Journal of Personality and Social Psychology*, 77(6):1271–1288.

Zimbardo, P. G., Keough, K. A., and Boyd, J. N. (1997). Present time perspective as a predictor of risky driving. *Personality and Individual Differences*, 23(6):1007–1023.