The iterative development and evaluation of the gamified stress management app “Stress-Mentor”

Vom Fachbereich Sozialwissenschaften

der Technischen Universität Kaiserslautern

zur Verleihung des akademischen Grades

Doktor der Naturwissenschaften (Dr. rer. nat)

genehmigte

D i s s e r t a t i o n

vorgelegt von

Alexandra Hoffmann

Tag der Disputation:	Kaiserslautern, 12.02.2020
Dekanin:	Prof. Dr. Shanley E. M. Allen
Vorsitzende/r:	Prof. Dr. Georg Wenzelburger
Gutachter/in:	1. Prof. Dr. Thomas Lachmann
	2. Prof. Dr. Didier Stricker

D386

March 2020

Table of Contents

List of Figures V

List of Tables VIII

Abbreviations X

Chapter 1: General Introduction 1

1.1 mHealth 1

1.1.1 Potential of mHealth 1

1.1.2 Gamification and Interactive Design 2

1.1.3 Ethical, Legal and Social Implications 3

1.1.4 Assessing the Quality of Health Apps and User Compliance 4

1.2 Stress and Stress Management 6

1.2.1 mHealth for Stress Management 6

1.2.2 mHealth for Pain Management 8

1.2.3 Overview and Gap in Research 9

Chapter 2: The Iterative Development of "Stress-Mentor" 12

2.1. Development of the MVP and Resulting Implications 12

2.2 Integration of a Diary Overview Chart and Resulting Implications 19

2.3 Integration of Stress Management Techniques 21

2.3.1 Background 21

2.3.2 Testing the Applicability of the Wearable EEG System MUSE for a Neurofeedback-Based Relaxation Exercise 22

2.3.2.1 Introduction 22

2.3.2.2 Study 1: Evaluation of MUSE's Reliability and Validity 27

2.3.2.2.1 Methods 27

2.3.2.2.2 Results 33

2.3.2.2.3 Discussion 37

2.3.2.3 Study 2: Evaluation of the Usability, User Experience and General Quality of the MUSE Wearable and Its Neurofeedback-Based Companion App 40

2.3.2.3.1 Methods 40

2.3.2.3.2 Results 42

2.3.2.3.3 Discussion 45

2.3.2.4 Conclusion 48

2.3.3 Testing the Applicability of Audio- and Text-Based Relaxation Exercises in a User Study 49

2.3.3.1 Introduction 49

2.3.3.2 Methods 50

2.3.3.3 Results 51

2.3.3.4 Discussion 51

2.3.4 Description of Included Stress Management Methods 53

2.4 Linking Gamification and Behavior Change Techniques in “Stress-Mentor” 54

2.4.1 Background 54

2.4.2 “Stress-Mentor’s” Concept 56

2.4.2.1 Diary 56

2.4.2.2 Avatar 58

2.4.2.3 Agent 59

2.4.2.4 Task of Day/Week 60

2.4.2.5 Visualization of Progress 62

2.4.2.6 Additional Features 62

2.4.3 Summary 64

Chapter 3: A Systematic Comparison of the MVP with the Final Version of “Stress-Mentor” Regarding Perceived App Quality 65

3.1 Introduction 65

3.1.1 Background 65

3.1.2 Gamification in Health Apps 66

3.1.3 “Stress-Mentor’s” Final Version vs. the MVP 66

3.2 Methods 69

3.2.1 Participants 69

3.2.2 Study Procedure 69

3.2.3 App Quality 70

3.2.4 Analysis 71

3.3 Results 71
3.4 Discussion 74
3.5 Conclusion 76
Chapter 4: Evaluation of "Stress-Mentor's" Gamified Concept 77
4.1 Introduction 77
4.1.1 Background 77
4.1.2 Short Summary of "Stress-Mentor's" Concept 80
4.1.3 Study Aim 82
4.2 Methods 82
4.2.1 Participants 82
4.2.2 Study Procedure 83
4.2.3 App Usage 84
4.2.4 Usability and User Experience 84
4.2.5 App Quality 84
4.2.6 Analysis 86
4.3 Results 86
4.3.1 Usage Behavior 86
4.3.2 Usability and User Experience 90
4.3.3 App Quality 90
4.4 Discussion 91
4.4.1 Usage Behavior 91
4.4.2 Usability, User Experience and Quality Ratings 93
4.4.3 Participants' Feedback to "Stress-Mentor" 94
4.4.4 Limitations 96
4.5 Conclusions 96
Chapter 5: Clinical Outlook: "Pain-Mentor" – Applying the Concept of "Stress-Mentor" in the Context of Chronic Pain Management 97
5.1 Introduction 97
5.1.1 Background 97

5.1.2 "Pain-Mentor" 100
5.1.3 Motivation 103
5.2 Methods 104
5.2.1 Participants 104
5.2.2 Study Procedure 104
5.2.3 App Quality 107
5.2.4 Additional App Specific Questions 108
5.3 Results 108
5.3.1 Quality of "Pain-Mentor" 108
5.3.2 Additional Questions Outcomes 109
5.4 Discussion 109
5.4.1 Principal Results 109
5.4.2 Limitations 114
5.5 Conclusions 114
Chapter 6: General Discussion 116
6.1 Principal Results 116
6.2 Ethical, Legal and Social Implications 119
6.3 Recommendations for the Development of Gamified Health and Medical Apps 121
6.4 Recommendations for the Use of Wearables in the Stress Management Context 123
6.5 Limitations and Outlook 125
6.6 Conclusion 125
References 127
Appendix 143
Acknowledgements 165
Notes 166
Declaration 167
Curriculum Vitae 168

List of Figures

Figure 1: Development stages of Stress-Mentor. 13

Figure 2: Screenshots of the MVP. Left: main screen with avatar and diary summary. Right: diary scales with color coded entries to provide feedback on the conformity of the user's behavior with the health recommendations. 15

Figure 3: Change of the avatar's appearance in accordance with the diary entries. 15

Figure 4: Means and standard errors of the MVP's (diary and avatar) uMARS ratings. 17

Figure 5: Displayed are the two chart types that were compared in a user study; a) depicts the bar chart and b) the point chart. 20

Figure 6: Diagram of the study procedure. The order the conditions were presented in was randomized across all participants. 29

Figure 7: Electrode placement according to the international 10-20 system. The locations used in this study are highlighted in grey. 30

Figure 8: Means and standard deviations for the alpha power measured by each of the two devices (wearable EEG MUSE and the gel-based EEG) in each electrode (A1, A2, Fp1, and Fp2) during each of the four experimental conditions (eyes open, eyes closed, breathing, and brainstorming). 36

Figure 9: Means and standard deviations for the beta power measured by each of the two devices (wearable EEG MUSE and the gel-based EEG) in each electrode (A1, A2, Fp1, and Fp2) during each of the four experimental conditions (eyes open, eyes closed, breathing, and brainstorming). 37

Figure 10: Study procedure for testing the wearable EEG MUSE and its companion app. 41

Figure 11: Self-Assessment Manekin ratings at the beginning of the study (T1), after connecting the wearable EEG MUSE to the MUSE companion app (T2), and after finishing the second meditation (T3). Depicted are the medians, maximum and minimum values, as well as first and third quartiles. Significant differences are marked with an asterisk (*). 43

Figure 12: Means and standard deviations of the User Experience Questionnaire (UEQ) ratings for the MUSE system (wearable and app). 43

Figure 13: Means and standard deviations of the uMARS ratings for the MUSE companion app. 44

Figure 14: Screenshot of an audio exercise. 52

Figure 15: App features timeline. 56

Figure 16: Left: screenshot of the diary. Middle: the app's main screen with a summary of the color-coded diary entries. Right: screenshot of the diary overview diagram. 58

Figure 17: a): Changes in the avatar's looks dependent on the diary entries (left: red entries, middle: yellow entries, and right: green entries). b): Avatar's development over time, dependent on the user's progress (left to right: egg, infant, child, teenager, and adult). 59

Figure 18: Screenshot of the photobook to track the avatar's development. 60

Figure 19: Left: the "wise owl" provides instructions and information about the app's usage. Right: the "wise owl" provides general information about stress management. 61

Figure 20: a) Screenshot of Stress-Mentor that shows the avatar and the suitcase which is filled with virtual items. The badges are placed on the tree trunk. The experience points are displayed in the upper right corner. b) Screenshot of the shop. c) The "wise owl" hands a new badge to the user, who has completed the tasks of the day on a regular basis. d) Progress bars depict the user's progress with regard to the next level and the badges. 63

Figure 21: Screenshots and content of Stress-Mentor's first test version (minimal viable product, MVP, left screenshot) and its full version (right screenshot). The app combines evidence-based stress management techniques with gamification. The content already included in the MVP is highlighted in bold print. 68

Figure 22: uMARS (user Mobile Application Rating Scale) ratings of the MVP (minimal viable product) and the full version of Stress-Mentor are displayed. Depicted are the medians, maximum and minimum values, as well as first and third quartiles. Significant differences between the groups are marked with an asterisk (*). 72

Figure 23: Screenshot of Stress-Mentor. 81

Figure 24: Differences and similarities between the gamified and the not gamified version of Stress-Mentor. 83

Figure 25: Diagram of the study procedure. 85

Figure 26: Means and standard errors of the number of performed tasks for the experimental group (EG) who received the gamified version of Stress-Mentor, and the control group (CG) who received a non-gamified version of the app, for young (18-35 years), middle aged (36-50 years), and old users (51-65 years). 90

Figure 27: Means and standard errors of the amount of time users spent with the app for the experimental group (EG) who received the gamified version of Stress-Mentor, and the control group (CG) who received a non-gamified version of the app, for young (18-35 years), middle aged (36-50 years), and old users (51-65 years). 90

Figure 28: User Mobile Application Rating Scale ratings for the experimental group (EG) who received the gamified version of Stress-Mentor, and the control group (CG) who received a non-gamified version of the app. Depicted are the medians, maximum and minimum values, as well as first and third quartiles. Significant differences between the ratings of both groups are marked with an asterisk (*). 91

Figure 29: Screenshot of Pain-Mentor. 102

Figure 30: Depiction of the participant acquisition and study procedure. 105

Figure 31: Sequential order of the app testing process that all experts followed. 106

Figure 32: Experts mean ratings with standard deviations for the MARS (Mobile Application Rating Scale) regarding Pain-Mentor. 109

List of Tables

Table 1: Diary categories of the MVP with entry evaluation and descriptions of the avatar's changing appearance depending on the diary entries. The values for sleep are displayed in a simplified manner, because the evaluation depends on a person's age. 16

Table 2: Means and standard deviations of Pearson correlation coefficients (r) between the power spectral densities of 0.1-110 Hz of session 1 and 2, regarding each experimental condition, device, and channel. All p-values were ≤ .001. 33

Table 3: Means and standard deviations of the Pearson correlation coefficients (r) between the power spectral densities of 0.1-110 Hz of the two devices regarding each experimental condition, session, and channel. All p-values were ≤ .001. 33

Table 4: Stress management exercises that are included in Stress-Mentor. 54

Table 5: Detailed description of the gamification techniques that are included in Stress-Mentor. 55

Table 6: Diary categories of Stress-Mentor's full version with entry evaluation. Sleep duration and alcohol are displayed in a simplified manner as the health recommendations depend on the user's age and gender respectively. 57

Table 7: Means (M), standard deviations (SD), and statistical values of Mann-Whitney-U tests for uMARS ratings of the MVP and the full version of Stress-Mentor by item. The new alpha level is set at .025 according to the Bonferroni correction method. 72

Table 8: MANOVA results for main effects and interactions. 86

Table 9: ANOVA results for main effects and interactions for number of performed tasks and amount of time spent with the app. 87

Table 10: Post-hoc t-test results for differences between young, middle aged and old users regarding the usage behaviors. 87

Table 11: T-test results for within-groups comparisons regarding the number of performed tasks per week. 88

Table 12: Post-hoc t-test results for differences between the experimental and control group with regard to the number of performed tasks per week. 88

Table 13: T-test results regarding differences between the weeks regarding the amount of time users spent with the app. 88

Table 14: T-test results for within-groups comparisons for young, middle aged, and old users regarding the number of performed tasks per week. 89

Table 15: Post-hoc t-test results for differences between young, middle aged, and old users regarding the number of performed tasks per week. 89

Table 16: Means (M), standard deviations (SD), and statistical values of Mann-Whitney-U tests for ratings of the User Experience Questionnaire of the MVP and the full version of Stress-Mentor. ____ 91

Table 17: Experts' mean ratings and standard deviations for the additional questions regarding Pain-Mentor. ____ 110

Abbreviations

app	application
CG	control group
CRS	Comfort Rating Scales
EEG	electroencephalography
EG	experimental group
ERP	event related potential
H	hypothesis
M	mean
MARS	Mobile Application Rating Scale
Mdn	median
mHealth	mobile health
MVP	minimal viable product
PMR	progressive muscle relaxation
PSD	power signal density
SAM	Self-Assessment Manikin
SD	standard deviation
SE	standard error
SUS	System Usability Scale
UEQ	User Experience Questionnaire
uMARS	user Mobile Application Rating Scale

Chapter 1

General Introduction

1.1 mHealth

Mobile technologies, such as smartphones and tablets, possess unique qualities that make them powerful tools for promoting health and healthy lifestyles (Klasnja and Pratt, 2012). They are widely adopted (e.g., approximately 70% of the German and 80% of the American population used smartphones in 2018, Statista, 2019a), always on, and people tend to carry them with them everywhere (Ventä et al., 2008). Furthermore, their functionalities are constantly increasing. This makes them the perfect medium to deliver public health information and interventions. In line with this thinking, the recent mobile health (mHealth) trend aims to help people to improve their personal health through mobile technologies (Harrison et al., 2011). So far, over 300,000 smartphone applications (apps) are available (with more being added each day) that aim to assist users to achieve diverse health objectives such as quitting smoking, taking more steps, and disease management (IQVIA, 2017).

1.1.1 Potential of mHealth

Health apps are a promising approach for the treatment and care of chronic diseases (Lee et al., 2018), as well as for improving general health and well-being in today's society (World Health Organization, 2011). In contrast to conventional therapies and other self-help approaches (e.g., self-help books), health apps allow the users to perform the training at a time and place that is convenient to them (Atienza and Patrick, 2011). This promises a more efficient use of available resources, and allows access to knowledge and critical medical information to both patients and physicians (Istepanian and Lacal, 2003). Previous research indicates that health apps can support the change of health relevant behaviors, such as medication adherence, physical activity or making healthy lifestyle choices (O'Reilly and Spruijt-Metz, 2013; Anglada-Martinez et al., 2015). They can also increase the user's education, tracking, motivation, and adherence (Handel, 2011; Ahtinen et al., 2013; Gibbons et al., 2018).

In addition, such apps support detailed tailoring and can therefore be easily adjusted to the user's personal demands (Atienza and Patrick, 2011). These features are said to improve both the adoption of self-management strategies and subsequently, disease prevention. This, in turn, can have a positive influence on the consumer's health (Gibbons et al., 2018). Consequently, mHealth can lead to substantial financial savings for both the user and the health care system (Luxton et al., 2011). That this approach is well received by users (Proudfoot et al., 2010) is reflected by the increasing number of individuals that use smartphone applications to manage their personal health (1.7 billion health apps were downloaded in 2013, 3.7 billion health apps were downloaded in 2017, Statista, 2019c).

However, in order to be effective, health apps need to be based on evidence-based content. This includes behavior change techniques which enable the user to achieve long-term behavior change (Abraham and Michie, 2008), as well as the combination of those techniques with gamification in order to ensure user compliance (Cafazzo et al., 2012).

1.1.2 Gamification and Interactive Design

Despite the potential usefulness of mHealth interventions, previous research suggests that the interest in health apps can be fleeting and that their use is, therefore, often perceived as a temporary commitment (Dennison et al., 2013). The exploration of strategies that motivate user compliance has been identified as an important research direction in order to ensure the effectiveness of mHealth (Klasnja and Pratt, 2012). Gamification, i.e. the use of game elements in nongame contexts, is aimed at increasing the motivation, engagement, and enjoyment of interventions (Deterding et al., 2011). This goes a long way to ensure the user's interest (Oinas-Kukkonen and Harjumaa, 2009) and thus, to increase his or her exposure to the intervention's content (Davies et al., 2012). A taxonomy by Hoffmann et al. (2017) identified 17 gamification techniques (e.g., levels, leaderboards, avatars, agents, time pressure, and digital rewards such as points and badges) that can be applied in the mHealth context. Empiric evidence supports that the use of such techniques can, indeed, increase the use of a service (Hamari, 2013). Taylor et al. (2019) for example reported an increase in compliance for keeping a gamified health diary, compared to a non-gamified version of the app. The technology acceptance model supports this finding. It states that a systems usage intention relies on the user's perceived joy of use (Venkatesh and Bala, 2008). This is an important consideration, because increasing the usage of health apps through gamification can reinforce desired behaviors (McKeown et al., 2016). It is therefore already applied in many different kinds of health apps (e.g., physical activity, diet and weight loss, and hygiene (Pereira et al., 2014)).

Indeed, previous studies provide evidence that the use of gamification can be an effective tool to support mHealth interventions. For example, the application of points in a diabetes intervention app (Cafazzo et al., 2012) and the combination of different gamification techniques (e.g., points, badges, leaderboards, time pressure) in a weight management app for children have had positive impact on health relevant behaviors (González et al., 2016). This is further supported by another study that focused on the combination of game elements (e.g., challenges, social pressure, social support, and digital rewards) with behavior change techniques (e.g., providing information about health consequences, goal setting, and self-monitoring through a digital diary) in an app for smoking cessation. The results of this study showed that this combination of gamification with aspects from behavior change theory was well received by users (Edwards et al., 2018). It is therefore hardly surprising that gamification has also been reported to increase the user experience and usability of health apps (Zagel and Bodendorf, 2014; Johnson et al., 2016).

Notwithstanding these positive results, it has to be kept in mind that the effects of gamification are diverse and depend largely on the application context (Johnson et al., 2016). Though most studies report positive effects of gamification (Koivisto and Hamari, 2019), some also report neutral or even negative effects with regard to cognitive and behavioral outcomes (Johnson et al., 2016).

1.1.3 Ethical, Legal and Social Implications

In addition to content, developers need to consider certain ethical, legal, and social challenges when designing health apps. For one, the research codes that are followed when studying mHealth often depend on the researchers' profession. However, it has been called into question to what extent the applied codices can actually cover the field of mobile technologies. Nonetheless, ethical research guidelines also apply to mHealth studies. It has been suggested that the development of mobile health specific codices could solve this problem. For this purpose Weber (2015) introduced MEESTAR, a model that helps developers to critically reflect and evaluate their health technologies with respect to seven ethical dimensions (i.e., participation, security, care, autonomy, privacy, self-perception, and fairness). Notwithstanding the potential of such solutions, to this day no ethical codex has established itself (Albrecht and Fangerau, 2015). Still, if conducting a clinical trial, the research should meet the overall requirements for clinical research and, thus, be in line with the Declaration of Helsinki (World Medical Association, 2013).

Where appropriate, legal idiosyncrasies may have to be considered, such as medical product laws that, at least in Germany, apply to medical apps. This includes, for example, apps that support patients in dealing with their disease (e.g., a chronical illness) in everyday life or apps that are used to support medical professionals in diagnosing a disease (MPG, 2019).

Another major aspect that should be considered is the app's acceptability. In order to ensure that a health app is accepted it is important to involve the user in its development process. This ensures that user requirements, such as data protection and personalization, are taken into account (van Velsen et al., 2018). Additionally, app developers should make sure that the app is widely accessible and low cost, as these aspects can restrict its comprehensive use (Krebs and Duncan, 2015).

Also, a health app's acceptance is often decreased through a difficulty for both users and medical professionals to identify secure, high-quality health apps (Institute for Healthcare Informatics, 2013; Boudreaux et al., 2014). Though there are many apps on the market, their quality and effectiveness is often called into question (De La Vega and Miró, 2014). Yet, both aspects are important factors to ensure the acceptance and usage of health apps (Venkatesh and Bala, 2008; Akter et al., 2013).

1.1.4 Assessing the Quality of Health Apps and User Compliance

Because both perceived quality and effectivity are important factors to ensure a health app's acceptance (Venkatesh and Bala, 2008; Akter et al., 2013), they are often the focus of mHealth research. Different approaches are used to shed light on these aspects and can help potential users and medical professionals to identify suitable health apps (Jake-Schoffman et al., 2017).

One approach that is often found in the literature is content analysis, which aims at reviewing the evidence-based content of currently available health apps (Jake-Schoffman et al., 2017). Such reviews provide a good insight into which apps make use of proven concepts (e.g., behavior change techniques) and which are lacking in this regard (Breton et al., 2011; Christmann et al., 2017a; Hoffmann et al., 2017).

Another method that aims at determining the quality of health apps is the application of the Mobile Application Rating Scale (MARS, Stoyanov et al., 2015). The MARS was developed specifically to assess the quality of health apps in the eyes of health and IT experts. Later, a user centered version of the MARS was published (user Mobile Application Rating Scale (uMARS, Stoyanov et al., 2016), which focuses on assessing a health app's quality from the user's perspective. Both the MARS and uMARS questionnaires provide easily interpretable ratings regarding the general quality, the subjective quality, and the perceived impact of the contemplated health app.

Besides quality, how well an app functions has major impact on its intended use (Venkatesh and Bala, 2008). Usability testing (e.g., with the System Usability Scale (SUS, Brooke, 1996) can ascertain a health app's functionality (Jake-Schoffman et al., 2017). Though important, usability criteria do not cover all aspects that make a program appealing (Hassenzahl et al., 2000). To achieve a broader assessment of how an app is perceived by users, app developers should also consider the feelings, impressions, and attitudes that the user experiences with the product. Evaluating the user experience (e.g., with the User Experience Questionnaire (UEQ)) allows the comprehensive assessment of subjective reactions, specifically emotional aspects of an app's usage (Laugwitz et al., 2008).

Even though emotional or behavioral consequences are judged to be equally relevant, their outcomes might differ from one another (Hassenzahl, 2001). Thus, in addition to content, quality, usability, and user experience, the effectiveness evaluation through randomized controlled trials plays an important role in the assessment of health apps. This includes the evaluation of the user's compliance. One approach that is used to determine this aspect is the measurement of usage duration and quantity (e.g., Christmann et al., 2018b). The reason for this is that the regular use of an intervention ensures that the user comes into contact with the app's behavior change content (Vandelanotte et al., 2007; Webber et al., 2010). This can increase the effectiveness of an intervention (McKeown et al., 2016). In turn, a next step should include longitudinal user studies that evaluate both the app's effectiveness and its effects on long-term behavior change (Institute for Healthcare Informatics, 2013).

1.2 Stress and Stress Management

The previously addressed aspects also apply in the context of stress management. Stress, and with it stress management, have gained importance with regard to personal health in recent years. The prevalence of chronic stress in today's society is high. More than half of the working age society report stress and stress related symptoms (Wiegner et al., 2015). This poses a growing health risk to modern society, since chronic stress is directly linked to a person's physical, as well as mental, well-being (Malarkey and Mills, 2007). Furthermore, chronic stress and health are also indirectly connected through stress-related behaviors (e.g., smoking, alcohol and drug abuse, poor eating habits, sedentary lifestyle, and insufficient therapy adherence) (Baum and Posluszny, 1999; Cohen et al., 2007). As a result, preventing chronic stress through the learning of stress management strategies in healthy individuals, is an important way to improve society's health and well-being (American Psychological Association, 2012, 2015).

However, stress management interventions are not only important for disease prevention in healthy individuals. Poor stress management has also been linked to an increased health and mortality risk in patients with chronic conditions (Russ et al., 2012). It is therefore, not surprising that stress management training is a major part of many multimodal disease treatments, such as in the treatment of chronic pain (Flor et al., 1992; Elgar and McGrath, 2003; Barlow and Ellard, 2006).

Stress management methods for prevention or in disease treatment are usually taught in single or group therapy sessions. However, in recent years stress management apps have been becoming an increasingly popular medium to promote stress self-management (e.g., Morris et al., 2010; Ahtinen et al., 2013; Chittaro and Sioni, 2014).

1.2.1 mHealth for Stress Management

In line with this trend of using mHealth for stress management purposes, a rising number of stress management apps are now available (e.g., “Ovia”, “Mevi”, “DeStressify”, “StressEraser”, “AEON”, “Headspace”, “Healthy Mind” and “myCompass”). First evidence shows that the use of such apps can have a positive impact on a person's stress level

(Ahtinen et al., 2013; Chittaro and Sioni, 2014; Economides et al., 2018; Lee and Jung, 2018). Moreover, stress management apps can successfully reduce symptoms of stress and improve the user's psychological (Harrison et al., 2011) and physical well-being (Lee and Jung, 2018).

Like other mHealth programs, stress management apps are more and more often combined with wearable devices in order to monitor or manipulate the user's biological responses (e.g., heart rate, respiratory rate, peripheral skin temperature, and neuronal activity) through biofeedback. Examples of how wearables are utilized in stress management apps include the monitoring and controlling of relaxation levels, breathing patterns, and concentration (Kaushik et al., 2006; Lee, 2009; Maddox et al., 2015).

However, no matter whether mHealth interventions for stress management are based on wearables or not, in order to ensure their effectiveness (Morris et al., 2010; Harrison et al., 2011; Chittaro and Sioni, 2014) they have to combine predisposing factors (e.g., provide information and knowledge about stress and stress management, and motivate the user), enabling factors (e.g., teach stress self-management skills and allow the tracking of stress relevant behaviors), and reinforcing factors (e.g., rewards and feedback) (Payne et al., 2016). This shows that combining evidence-based content from behavior change theory, stress management theory, and gamification theory is a promising approach to ensure the effectiveness of stress management apps.

Notwithstanding the importance of integrating and combining evidence-based content from these backgrounds, app reviews show that few apps on the market make use of these concepts. The available apps vary largely with respect to the amount of included behavior change techniques and stress management methods (Coulon et al., 2016; Christmann et al., 2017a). While some apps do not include any evidence-based behavior change techniques or stress management methods, those that do mostly focus on one specific coping strategy (Coulon et al., 2016). This demonstrates that current stress management apps show controversial theoretical underpinnings (Lee et al., 2014). However, there are also exceptions that combine several behavior change and stress management techniques within one app (e.g., "Mevi") (Christmann et al., 2017a).

The implementation of gamification in stress management apps on the other hand, remains an exception. Moreover, most app designers do not combine gamification with evidence-based content from behavior change theory or stress management theory in order to improve user compliance (Hoffmann et al., 2017).

One reason why gamification might not be met with enthusiasm in this context, could be the ongoing discussion about whether gamification should be used in stress management applications. For one, little is known about how gamification affects psychological or behavioral outcomes in the context of stress management, or about how users perceive this aspect (Johnson et al., 2016; Sardi et al., 2017; Koivisto and Hamari, 2019). Furthermore, users of the stress management app "Ovia" did not wish for an integration of game elements in this context (Ahtinen et al., 2013). This suggests that users might neither want, nor accept game elements in stress management apps. A possible cause for this could be that the advantages of linking mHealth with gamification is still unknown to them (Spil et al., 2017).

Though gamification is rare in mHealth for stress management, there have been a few exceptions. For example, MUSE app measures the user's brain activity though a wearable electroencephalography (EEG) device and then feeds the data back to the user in real time (neurofeedback) in order to support meditation (InteraXon Inc.). Afterwards, the user receives digital rewards (e.g., points and badges) in accordance with his or her performance (Przegalinska et al., 2018). Another example that combines biofeedback with gamification is StressEraser. Here, the user receives points for controlling his or her heart rate variability through breathing exercises (Ebben et al., 2009). Though these programs pose a first step towards integrating gamification in mHealth stress management interventions, so far it remains unknown if these approaches are effective, as well as wanted and accepted by users.

1.2.2 mHealth for Pain Management

Besides preventing chronic stress in healthy individuals, stress and stress management play an important role for a large number of medical conditions (Russ et al., 2012). One area where stress management plays an integral part of multimodal treatment approaches is chronic pain therapy (Jensen et al., 2003; Macea et al., 2010; Irvine et al., 2015). Here it is applied in combination with medical treatments (e.g., medication, surgery rehabilitation, and physical therapy) (American Psychological Association, 2013) in order to achieve a better treatment outcome. Such multimodal treatments have shown to improve overall quality of life in chronic pain patients (Flor et al., 1992; Elgar and McGrath, 2003; Barlow and Ellard, 2006).

However, the integration of multimodal approaches into therapy has been slow (Breivik et al 2006). Indeed, the majority of patients never receive the required training to acquire the necessary self-management skills that allow them to better deal with their chronic pain in everyday life (Peng et al., 2007a; Peng et al., 2007b; Lynch et al., 2008).

Due to the rising number of available health apps, appropriate chronic pain treatment approaches have the potential to reach a larger number of affected patients (Demiris et al., 2008). First evidence suggests that such apps have great potential to support chronic pain treatment and that they are well received by patients (Irvine et al., 2015; Jamison et al., 2018).

Even though the inclusion of evidence-based content is crucial to ensure the effectiveness of pain management apps (Morris et al., 2010; Harrison et al., 2011; Chittaro and Sioni, 2014), reviews show that the content of apps is limited in this context. With the exception of the pain diary app "PainTracker" (Jamison et al., 2018), most pain apps only supply information, and lack a combination of evidence-based functionalities (Rosser and Eccleston, 2011; Wallace and Dhingra, 2014; Lalloo et al., 2015). They miss strategies for social support, evidence-based self-care skills, and the tracking of the multidimensional experience of pain. In addition, the educational content that is included is often of poor quality (Jamison et al., 2018). Moreover, so far no German smartphone app that combines these aspects with gamification is available for adults. This strongly indicates that comprehensive, clinically informed smartphone apps with evidence-based content that combine these concepts are highly needed in this context (Lalloo et al., 2015).

1.2.3 Overview and Gap in Research

As outlined above, the effective combination of insights from behavior change theory and gamification in mobile health systems is a promising approach to ensure the user's adherence and long-term behavior change (e.g., Davis et al., 2008; González et al., 2016). The combination of these concepts has been suggested as a prerequisite to ensure the effectiveness of stress management apps (Payne et al., 2016). Nevertheless, a recent app review showed that stress management and behavior change content is almost never combined with gamification in order to improve user compliance and intervention adherence. Moreover, no app that links behavior change theory and stress management theory with gamification is available in German (Hoffmann et al., 2017).

To close this gap, a novel stress management application, named Stress-Mentor, was designed and implemented in an iterative development approach. Including the user into the development process in this manner aims to ensure that areas for improvements are uncovered, that the user's requirements are met, that the app is accepted by users, and that the development results in a high quality product (Stinson et al., 2010; van Velsen et al., 2018). This includes the user's need for data protection which can impact the app's usage intention (Faust-Christmann et al., 2018), as well as personalization (Gay and Leijdekkers, 2012) and the support of the user's autonomy (Sax et al., 2018). Also, in order to ensure comprehensive usage, the app should be widely accessible and low cost (Krebs and Duncan, 2015). Moreover, to meet the three prerequisites (predisposition, engagement, and reinforcement) for effective stress management apps and to ensure the user's compliance with the intervention, Stress-Mentor's concept should combine behavior change techniques and a multitude of stress management techniques with gamification features (Payne et al., 2016).

In line with the iterative development approach, a first prototype (minimal viable product, MVP, (Lenarduzzi and Taibi, 2016)) of Stress-Mentor was developed that linked the self-monitoring of stress relevant behaviors through a diary with the appearance of an avatar (Christmann et al., 2018b). This MVP was then improved based on user feedback. This included the extension of the gamification structure, as well as the testing and integration of various stress management strategies. The development process led to an extended app framework, which linked a large variety of behavior change techniques and stress management methods with gamification (see Chapter 2 for details on the development process and resulting app concept).

The resulting full version of Stress-Mentor was then evaluated in a four-week user study, to reveal improvements compared to the MVP. The results of this analysis are presented in Chapter 3. As previously recommended (Olff, 2015), this study also assessed the impact of gamification on the app's usage behavior, usability, user experience, and perceived quality (see Chapter 4). As the last step, the applicability of the app's concept for other health contexts was determined in an expert study, using the example of chronic pain management (stress management plays an important role in the multimodal treatment of chronic pain (e.g., Irvine et al., 2015). The results of this evaluation are presented in Chapter 5.

Chapter 6 concludes this thesis with a general discussion of the presented research. It includes a summary of the studies' main results and their impact, and discusses broached ethical, legal, and social implications. Moreover, recommendations for the development of gamified health apps and the application of wearables in stress management apps will be deduced from the presented research. At the end, an outlook on the requirements of future research will be given, in addition to an overview of the studies' limitations.

The chapters of this thesis build on each other and can be seen as cumulative research. Nonetheless, all chapters provide the necessary theoretical background and can also be read independently.

Chapter 2

The Iterative Development of "Stress-Mentor"

This chapter focuses on the iterative development of Stress-Mentor, which progressed over four stages (see Figure 1 for an overview). In the first stage, a first prototype was developed. This MVP consisted of a health diary that was linked to the appearance of an avatar in the form of a bird-like cartoon animal (Christmann et al., 2018b). In the next step this MVP was extended through the development of a diary overview diagram (Christmann et al., 2017b). Afterwards, the applicability of the wearable EEG MUSE for a neurofeedback relaxation exercise for Stress-Mentor was tested. For this purpose, two studies were conducted. The first study tested the reliability and validity of the data obtained with the wearable, the second study focused on the usability, comfort and quality of the device and the MUSE neurofeedback app. Afterwards, additional audio- and text-based stress management exercises were realized and tested for their applicability. The user studies that were conducted at each development stage provided feedback on areas of improvement. This feedback was then used to develop an extended app concept that combines aspects from behavior change theory and stress management theory within an extended gamification framework (Christmann et al., 2018a).

2.1. Development of the MVP and Resulting Implications

Self-regulatory skills are a prerequisite in order to change maladaptive health behaviors (Kanfer and Gaelick-Buys, 1991). This self-regulation process can be divided into three distinct stages, namely self-monitoring, self-evaluation, and self-reinforcement (Kanfer, 1970). The first stage, self-monitoring, is aimed at deliberately paying attention to one's own behavior (Kanfer and Gaelick-Buys, 1991). Three requirements must be met to ensure the effectiveness of self-monitoring through diaries. First, the user must be truthful when making an entry; second, the entries must be made within temporal proximity to the tracked behavior; and third, the user must be consistent in keeping his or her diary (Bandura, 1998).

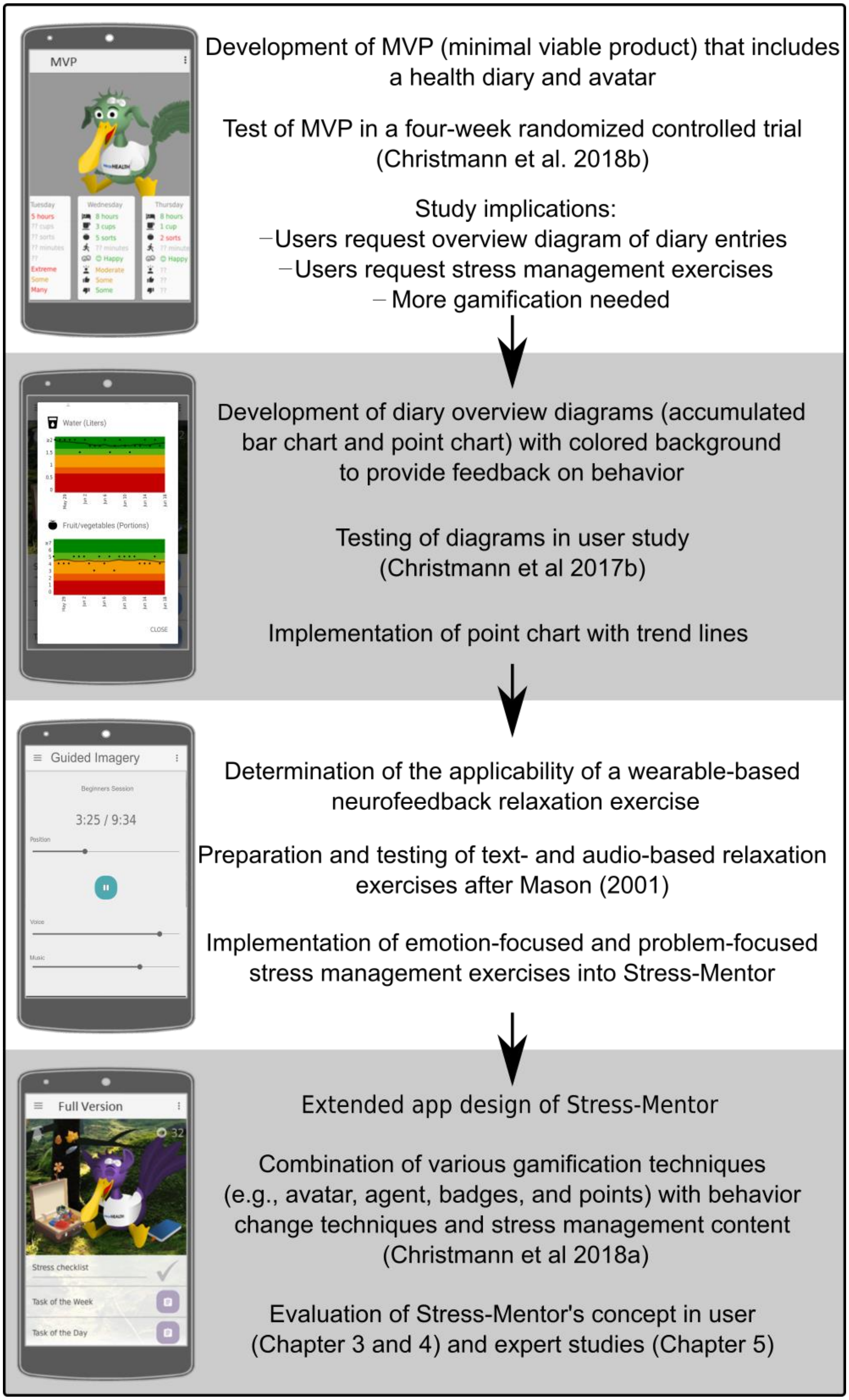

Figure 1: Development stages of Stress-Mentor.

In agreement with the recent mHealth trend, health apps that provide digital diaries can help to achieve these aspects (Stone et al., 2002; Wharton et al., 2014; Taylor et al., 2019). As a result, they can help users increase their awareness of the tracked behaviors, along with promoting behavior change (Acharya et al., 2011). Though the first two requirements for effective self-monitoring can be achieved with a diary app, digital diaries themselves might have problems to sustain the consistency of the user's entries (Solbrig et al., 2017). Vicarious reinforcement through avatars has been suggested as a solution to this problem (e.g., Taylor et al., 2019).

In line with this approach, Stress-Mentor's MVP combined self-monitoring through a diary with vicarious reinforcement through an avatar, i.e. a bird-like cartoon animal (the Rhineland-Palatinate Elwetritsch, displayed in Figure 2). Though vicarious reinforcement through avatars has been successfully applied in other health genres (e.g., in the context of physical exercise (Fox et al., 2009), or diet (Byrne et al., 2012)), Stress-Mentor's MVP was the first app that made use of this concept in the context of stress management (Hoffmann et al., 2017).

The diary included eight stress relevant constructs that could be tracked by the user on a daily basis (see Table 1 for a list of the included categories). For each entry, the color of the answer changed according to a traffic light coloring system, depending on whether the behavior was in line with official health recommendations or not. Entries that met the health recommendations were displayed in green, those that somewhat met the recommendations were displayed in yellow, and entries that did not meet the health recommendations were colored in red (Table 1 and Figure 2 for details).

Besides this direct feedback, the user also received indirect feedback through changes in the looks of his or her avatar. Each diary category was linked to one aspect of the avatar's appearance (see Table 1 for details). Overall, the avatar looked healthier if the user's entries were in accordance with the health recommendations (see Figure 3).

The MVP was tested in a four-week randomized controlled user study (Christmann et al., 2018b). With regard to usage consistency, this study revealed that the experimental group (EG), which received vicarious reinforcement through the avatar, did not show changes in the number of missing diary entries throughout the usage period. In contrast, the control group (CG) without vicarious reinforcement, showed a significant increase in missing diary entries over time.

Figure 2: Screenshots of the MVP. Left: main screen with avatar and diary summary. Right: diary scales with color coded entries to provide feedback on the conformity of the user's behavior with the health recommendations.

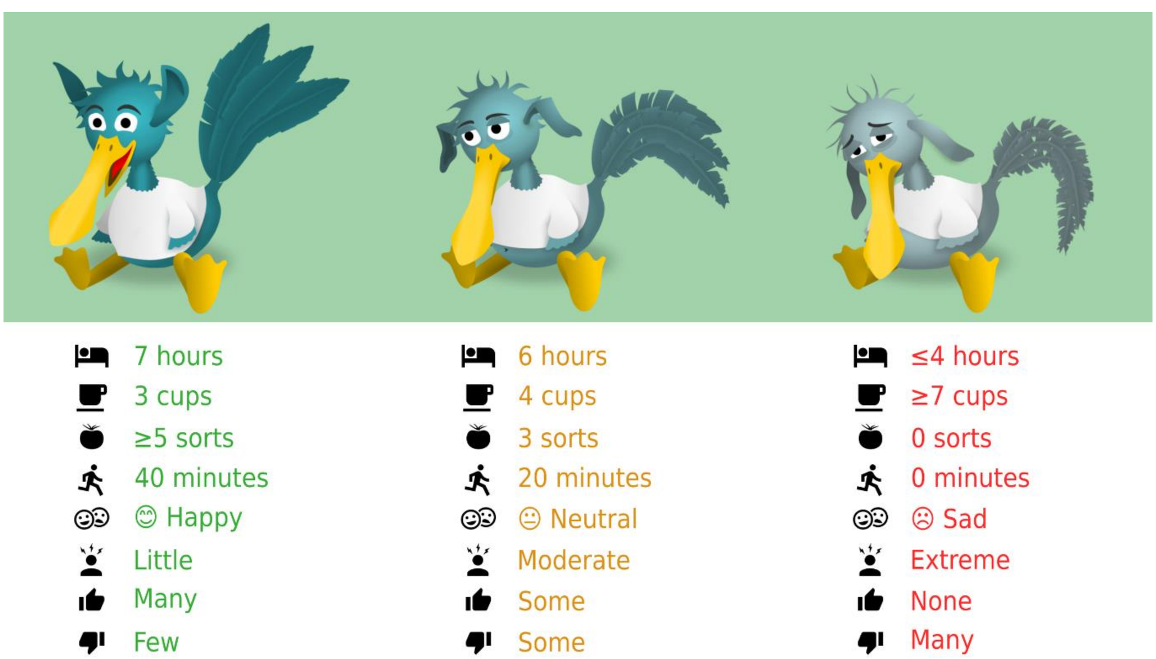

Figure 3: Change of the avatar's appearance in accordance with the diary entries.

Table 1: Diary categories of the MVP with entry evaluation and descriptions of the avatar's changing appearance depending on the diary entries. The values for sleep are displayed in a simplified manner, because the evaluation depends on a person's age.

Diary Categories	Green Evaluation	Yellow Evaluation	Red Evaluation	Change of Avatar's Appearance
Sleep in hours	7-9	6, 10	≤ 5, ≥ 11	Position of head, dark circles under the eyes
Caffeine intake in 100mg units	0-3	4-5	≥ 6	Trembling of the whole avatar
Portions of fruits and vegetables	≥ 5	3-4	0-2	Color saturation of plumage
Sport in minutes	≥ 30	10-20	0	Girth of abdomen
Emotional state	happy	neutral	sad, angry	Emotional expression
Stress level	0-1	2-3	4-5	Tidiness of plumage
Positive events	many	some	none, few	Position of the ears
Negative events	none, few	some	many	Position of the ears

This coincides with previous results that suggested that vicarious reinforcement can be an effective way to influence user behaviors (Fox et al., 2009; Byrne et al., 2012; Parks et al., 2014). However, the results with regard to the avatar's effect on the MVP's usage were somewhat ambiguous. The revealed effect of vicarious reinforcement in this study was not strong enough to result in significant group differences.

Still, the avatar was generally perceived in a positive manner by users of the EG. For example, they stated that watching the avatar change over time was fun. This attitude coincides with previous results on representing the user through an avatar's appearance (Lyles et al., 2017). This indicates that there appears to be no general dislike of gamification in the context of stress management apps.

In addition, the convergent validity of the diary categories demonstrates that the first property for effective self-monitoring, namely truthfulness (Bandura, 1998), can be achieved with the MVP, provided that the users are honest when making their entries and do not deceive themselves. The second property, i.e. temporal proximity, is also supported as the diary entries can be made anywhere and anytime. The last property for effective self-monitoring, namely consistency (Bandura et al., 1963), was supposed to be promoted through vicarious reinforcement. This could not be confirmed, as the results for the MVP's usage consistency were ambiguous.

Clearly, the avatar alone was not sufficient to keep the user entertained over a long usage period. This agrees with previous results, showing that user's wish for extensive motivational support (Thiele et al., 2002). Further support for this was found in the app's quality ratings. Though the uMARS ratings did not differ between the EG and CG (as revealed by Mann-Whitney U tests; Field, 2009), the MVP's ratings revealed room for improvement, especially regarding the subscales engagement and perceived impact (Figure 4).

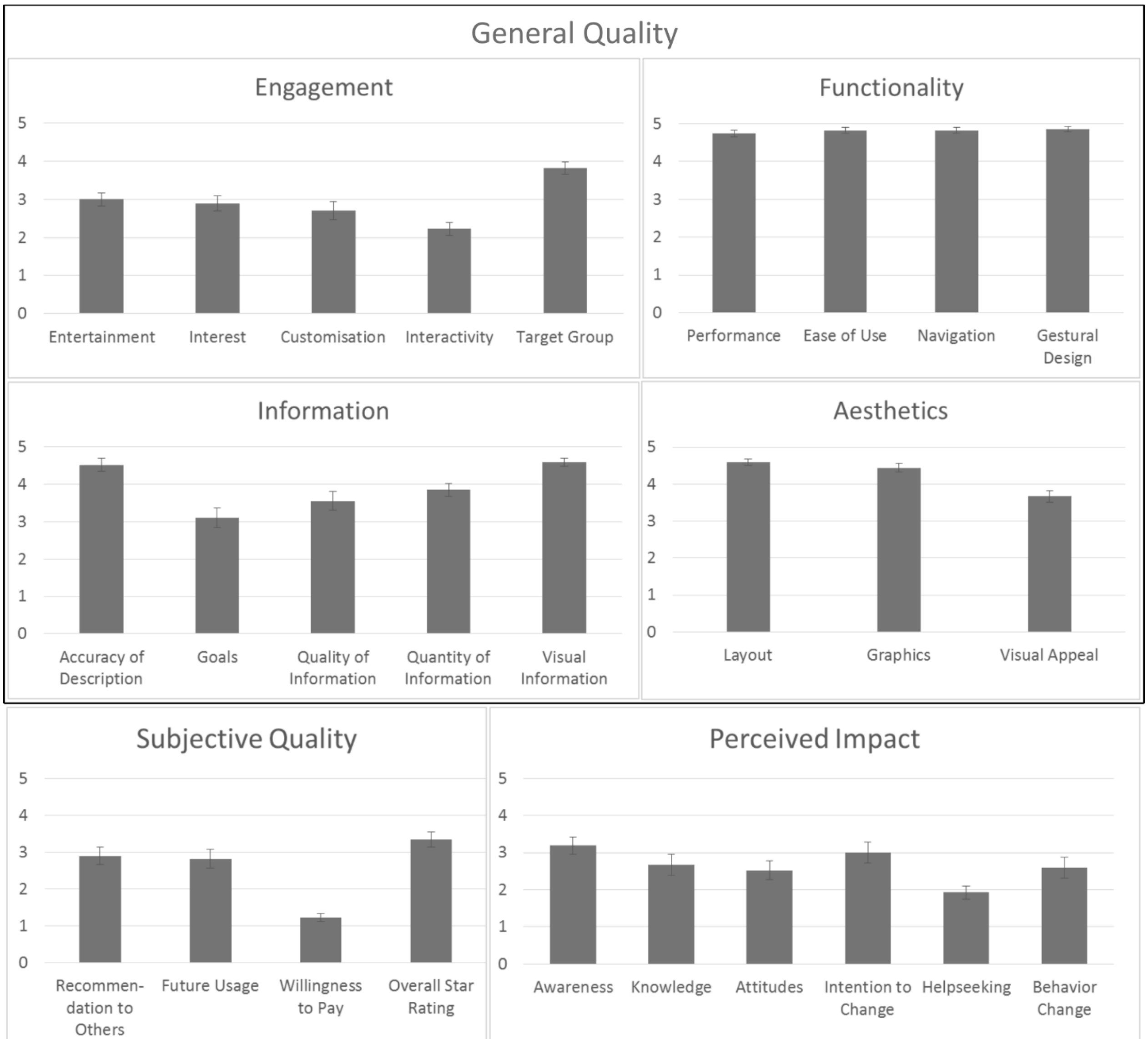

Figure 4: Means and standard errors of the MVP's (diary and avatar) uMARS ratings.

This finding was sustained by the users' comments. For one, the users requested an overview diagram to help them track their behaviors over time. Such charts are important features of health apps and support the user's self-evaluation, which is an important step for self-regulation. This, in turn, can promote behavior change (Kanfer, 1970).

Besides the overview diagram, the users wanted additional categories to be included in the diary, namely for water and alcohol consumption. Since stress can both directly impact a person's health through its physiological effects and also indirectly impact a person's health through maladaptive health behaviors such as poor eating habits or alcohol abuse (Baum and Posluszny, 1999; Cohen et al., 2007; Glanz et al., 2015), the app's diary was extended by two categories (namely, water and alcohol consumption). Furthermore, participants commented that the four provided smiley faces in the diary category mood were not always sufficient to represent their emotional state. This resulted in the addition of two more moods: worried and content. This matches with the users' general wish for a more detailed classification of the diary entries' color coding. Participants perceived the use of only three color assessments as too harsh, leading to the implementation of a more detailed subdivision of the diary's color coding system where it was appropriate.

Moreover, participants from both groups asked for a broader set of stress management methods besides self-monitoring (see Christmann et al., 2017a for taxonomy). An extension of the diary categories, as well as the implementation of stress management methods, could result in increased ratings of the app's perceived impact. It also goes a long way to ensure the app's effect on subjective stress level and coping, as users have difficulty to come up with practical coping strategies on their own (Kocielnik and Sidorova, 2015).

The results clearly showed that although the MVP was an adequate first approach, self-monitoring though a diary alone is not enough to satisfy the user long term. Moreover, the integration of an avatar for gamification might not be sufficient to improve the user's compliance. However, the ratings for engagement and, as a result, Stress-Mentor's usage consistency could be further improved through the integration of additional gamification techniques. Such approaches have shown great potential to ensure user compliance (Taylor et al., 2019) and as a result, the effectivity of health apps (e.g., Cafazzo et al., 2012; González et al., 2016).

2.2 Integration of a Diary Overview Chart and Resulting Implications

During the MVP study, the users requested a diary overview diagram in order to help them track their behavior change over time. This coincides with the fact that, besides self-monitoring through a diary, self-evaluation is an important step of the self-regulation process (Kanfer, 1970). Self-monitoring applications alone do not help people to reach their personal, actual goals (e.g., reducing subjective stress level). Instead, there is a risk that the collection of data itself becomes the goal (Marcengo and Rapp, 2014). Therefore, the reflective stage (Fleck and Fitzpatrick, 2010; Li et al., 2011) is crucial before the intended behavior change can occur (Fogg, 2002). Providing suitable visualizations (e.g., through charts; Li et al., 2011; Li et al., 2012; Cuttone et al., 2013) can facilitate self-evaluation (Cuttone et al., 2014) and thus, support the self-regulation of behaviors (Kanfer, 1970).

The line plot (Cuttone et al., 2014) and the bar chart (e.g., Sony LifeLog, Fitbit) are common long-term visualizations of health data. At first glance, these visualizations seem to be perfect for exploring progress over time as well as correlations between health behaviors (Li et al., 2011). However, these chart types do not consider how the respective health behavior should be appraised. Therefore, they do not allow the user to make decisions for or against behavior change (DiClemente et al., 2001). Exceptions, such as Fitbit, bridge this gap by providing additional goal fulfillment charts which depict to what degree the user's personal goals are achieved. However, visualizations that combine time series data and the appraisal of this data within one chart are rare.

To help Stress-Mentor's users to make decisions regarding behavior change, two alternate long-term visualizations (an accumulated bar chart and a point chart, see Figure 5) were developed and compared in a user study (Christmann et al., 2017b).

Although both charts resulted in satisfying overall quality scores (participants rated the quality of both charts with the uMARS), the findings were in favor of the point chart. For one, fluctuations were more reliably detected when presented in the point chart, whereas the accumulated bar chart covered up some of the periodic variance in the data. The detection of such periodic patterns has been pointed out to be fundamental (Larsen et al., 2013; Cuttone et al., 2014). Additionally, participants were more often unable to rate questions regarding the correlation between different health behaviors when they were displayed as bar charts.

Figure 5: Displayed are the two chart types that were compared in a user study; a) depicts the bar chart and b) the point chart.

Another important finding of this study was that the participants were not overloaded when both the raw data and the corresponding appraisals of the data were provided. This was supported by the ratings of perceived difficulty to answer, which did not differ between the two types of visualization. However, the point chart resulted in fewer trials in which participants were unable to pick an answer option (Christmann et al., 2017b).

Also, the results indicated that the presentation of the full color spectrum of the appraisals appears to play a role for the effective presentation of health data. Participants more frequently rated the behavior as health conscious when it was presented in the bar chart. While the full color spectrum was always visible in the point chart, this was not the case for the bar chart. It is likely that participants were not aware of the full color spectrum when rating the health consciousness of the behaviors displayed in the bar chart. This bias might be stronger for inexperienced users of the app, since the traffic light feedback system should be adopted easily over time (Eikey et al., 2015). Still, these results highlight the importance of displaying the full color spectrum for appraisals, especially regarding new users.

Overall, this study showed that combining health data with traffic light feedback in a bar or point chart can help the user to appraise his or her respective behaviors. However, there was a preference towards the point chart. In addition, some participants suggested using trend lines instead of single points in order to facilitate the observation of correlations in the point charts, as this approach reduces unavoidable noise (Cuttone et al., 2014). As there was a slight preference towards the point chart in this study, this visualization type, including the suggested trend lines, was implemented into Stress-Mentor.

2.3 Integration of Stress Management Techniques

2.3.1 Background

Though users can learn about their stress patterns based on the information they obtain through diaries, only very few are able to come up with practical coping strategies for improving their stress level on their own (Kocielnik and Sidorova, 2015). Thus, besides facilitating self-regulation through self-monitoring, self-evaluation and self-reinforcement, the inclusion of evidence-based stress coping strategies plays an integral role in ensuring the effectiveness of stress management apps (Williams and Kemper, 2010).

Lazarus (1985) distinguished between two broad types of coping strategies: problem-focused and emotion-focused coping. Problem-focused coping is aimed at altering a person's interaction with his or her environment (e.g., time management). Emotion-focused coping, on the other hand, is aimed at reducing stress sensations (e.g., progressive muscle relaxation (PMR)). Notwithstanding their positive effects, it has been shown that not all coping techniques work equally well for every individual (Barabasz and Perez, 2007). Stress management group interventions and self-help books, therefore, usually teach a broad range of coping techniques (e.g., Mason, 2001; Davis et al., 2008). Such multi-technique approaches seem to be more effective than using a single technique (Murphy, 1996; Jones and Johnston, 2000). They are, however, rarely found in current stress management apps (Coulon et al., 2016; Christmann et al., 2017a).

To close this gap, the goal was to provide a large variety of stress management techniques in Stress-Mentor. This combines the advantages of mHealth (e.g., being able to perform the stress management exercises anywhere and anytime; Blackburne et al., 2016) with those of multi-technique approaches (e.g., finding out which stress management techniques are best suited for the individual; Murphy, 1996; Koldijk et al., 2016). To determine how stress management exercises should be implemented into Stress-Mentor, several emotion-focused coping techniques were tested for their applicability in user studies.

2.3.2 Testing the Applicability of the Wearable EEG System MUSE for a Neurofeedback-Based Relaxation Exercise

2.3.2.1 Introduction

One example for an emotion-focused coping technique that aims at increasing relaxation and can help to handle and decrease stress is neurofeedback (Hjelm and Browall, 2000). In neurofeedback, brain activity (EEG waves) is analyzed by a computer in real time, broken down according to its frequency components and transformed into sensory inputs (e.g., visual or auditory) that provide feedback to the user. The distribution of EEG frequencies depends on the state of attention or consciousness (e.g., awake, sleeping, attentive, relaxed, and stressed). Alpha activity (8-13 Hz), for instance, appears to play an important role when providing neurofeedback that is targeted at relaxation. Previous studies have shown that an increase in relaxation is mirrored by an increase of alpha activity in the brain (Shaw, 2003). Moreover, users of neurofeedback can learn to distinguish between different brain frequencies, such as high and low alpha frequencies (Frederick, 2012) and control their production of alpha activity (Beatty, 1972). Such exercises can lead to an increase of perceived relaxation (Hjelm and Browall, 2000; van Boxtel et al., 2012).

Even though providing neurofeedback has a huge potential to support the user in staying autonomous and healthy, such applications have mostly remained in the context of science and have not yet reached consumers (Blankertz et al., 2010; McDowell et al., 2013). This is hardly surprising, considering that the use of such systems remains unpractical. One major reason for this is that the use of conventional EEG systems for neurofeedback requires the appliance of electrode gel in order to lower impedances between the electrodes and the scalp. The use of such “wet” electrodes, therefore, requires long preparation times and can lead to skin irritations caused by the electrode gel (Sullivan et al., 2007; Liao et al., 2011). Another reason such systems have not reached the market so far, is their cost (Cutrell and Tan, 2008; Hairston and Lawhern, 2015).

While the advantages might prevail for people who are dependent on them, it remains a fact that most EEG systems are not tenable for truly mobile or user oriented applications, due to issues regarding power, size, weight and comfort (Müller et al., 2008; McDowell et al., 2013; Gramann et al., 2014; Hairston and Lawhern, 2015). This makes them unappealing for healthy users and, therefore, leads to a shift in demands and requirements for such systems (Zander and Kothe, 2011; Ekandem et al., 2012; Lee et al., 2013). In order to be accepted on the market, such devices need to be well designed and gel-free (Müller et al., 2008). They also need to be applicable in everyday life (De Vos et al., 2014). Consequently, such systems must be simple to set up and troubleshoot, and they likely will have to be relatively transparent or even invisible to the user in order to ensure user compliance (McDowell et al., 2013). Additionally, they should be user friendly, robust, easy to use and designed in a 'fool-proof' manner (Mihajlovic et al., 2015), as well as low cost, lightweight and comfortable to wear in order to ensure their usability across different contexts (Hairston and Lawhern, 2015).

One possible approach to meet these challenges poses the use of dry electrode wearable EEGs. These systems omit the use of skin preparation and electrode gel and can operate with minimal wiring. This leads to significantly lower preparation times and effort and allows the user freedom of motion (Matthews et al., 2008). Moreover, the use of dry electrode wearable EEG systems helps to make the devices smaller and lighter, as user's will have to carry less hardware around (Wilson and Russell, 2003; De Vos et al., 2014). If designed in a user-friendly way such systems can be easily applied in health care and, thus, also in the context of stress management (Lee et al., 2013). To determine whether this is the case, different studies have assessed the usability of consumer-grade wearable EEGs. The results suggest that, though the devices often receive good usability ratings, there is still room for improvement (Rebolledo-Mendez et al., 2009; Leape et al., 2016).

EEG systems that could potentially be used in everyday life include, e.g. Neurosky MindWave (http:/neurosky.com/), Emotiv EPOC (https://www.emotiv.com/), imec EEG headset (https://www.imec-int.com/en/home), Cognionics Agile-10 (https://www.cognionics.net), and InteraXon MUSE (http://www.choosemuse.com). The integration of such systems into mHealth interventions can pose an effective alternative to self-help literature with respect to reducing stress (Miller, 2012). For example, wearable EEGs have been applied to control the user's concentration and relaxation levels (Lee, 2009; Maddox et al., 2015). However, their integration into mobile stress management interventions and their combination with gamified health apps that try to target user compliance is slow. One exception is the wearable EEG system MUSE (InteraXon Inc.), which is linked with a meditation app that provides the user with gamified neurofeedback. On one side, the app provides online auditory feedback during the exercise. On the other side, the MUSE companion app collects and visualizes the EEG data of a meditation session in a simplified and gamified manner (e.g., it provides points and badges) and compares the results of the current session to previous ones (Przegalinska et al., 2018).

Such combinations of wearable EEG systems with gamified mobile applications have great potential in the context of personal health care (Mihajlovic et al., 2015). However, in order to provide effective neurofeedback to the user, it is essential that the used EEG devices are both reliable and accurate in their measurement of the used EEG characteristic (e.g., ERPs or frequency band powers). Both the acceptability of the data collection method and very high accuracies are obligatory to ensure the users' acceptance (Wilson and Russell, 2003). The benefits of dry electrode wearable devices are, however, often counterbalanced by low signal reliability and validity (McFarland and Wolpaw, 2011).

There are already studies comparing the performance of dry electrode (e.g., Estepp et al., 2009) or wearable (e.g., Matthews et al., 2008; Bleichner et al., 2016) EEG systems to that of research-grade ones that are gel-based. However, only few studies have evaluated the validity of commercial, dry electrode wearable systems. Badcock et al. (2013) tested the performance of the Emotiv EPOC and compared it to that of a gel-based research EEG regarding auditory event related potentials (ERPs). For this they simultaneously recorded EEG signals with a research and the gaming EEG during 566 standard (1000 Hz) and 100 deviant (1200 Hz) tones under passive (non-attended) and active (attended) conditions for 21 adult participants. From these, they then obtained late auditory ERPs (P1, N1, P2, N2, and P3 peaks) and the mismatch negativity in active and passive listening conditions for each participant and each device. The study revealed that the two EEG systems were mostly comparable with respect to size and timing of the investigated ERPs. They only found small differences between the two devices for P1 in one of their experimental conditions, as well as for the peak and latency of P3. Also, mismatch negativity waveforms of the research and wearable EEG were only similar if they were not noisy. Despite these exceptions, the study concluded that the Emotiv EPOC might be a plausible alternative to research-grade EEG systems (Badcock et al., 2013). Their results are supported by a different study, which reported that the Emotiv EPOC's performance is as good as that of current research-grade systems in many paradigms (Melnik et al., 2017). Another study tested the validity of a three-channel dry electrode wearable system developed by Brain Products. In this study too, the EEG signal was recorded simultaneously with the wearable and a state-of-the-art research system. Two experimental paradigms were used to validate the device. On one hand, the researchers investigated ERPs with a variation of the oddball paradigm. On the other hand, they investigated features of the frequency domain with a paradigm that included occipital alpha. They found that the amplitude and temporal structure of the ERPs, as well as the features of the frequency domain did not differ between the two EEG systems. The results therefore supported the validity of the dry electrode wearable system with regard to both ERPs and features of the frequency domain, including occipital alpha (Zander and Kothe, 2011). Yet another study investigated the reliability and validity of Mindwave, a single electrode commercial EEG by NeuroSky. Again, the signal was simultaneously recorded with the wearable and a research-grade device during a resting state paradigm with an eyes open and an eyes closed condition and a driving simulator paradigm were participants performed an ecological and a dynamic driving task. The results indicated that the wearable was less reliable but still provided data that was comparable to that of the gel-based system regarding power spectra, temporal aspects and signal-to-noise ratio (Rieiro et al., 2019).

While lower resolution components have previously been suggested as a plausible way to improve design, power, and cost of dry electrode, wearable EEG systems (Hairston and Lawhern, 2015), these studies show that such trade-offs might no longer be necessary. Instead they indicate that such systems can be a useful tool in a variety of novel applications (Lopez-Gordo et al., 2014). On the other hand, a review of wearable EEGs for daily applications suggests that the devices require improvements with regard to their personalization, sensory input, brain signal acquisition and analysis, as well as their feedback generation (Mihajlovic et al., 2015).

Though the wearable EEG MUSE (InteraXon Inc.) was specifically designed to support relaxation through meditation exercises (InteraXon Inc.; Ijjada et al., 2015) and is used in a variety of other application scenarios (e.g., mental state recognition (Bashivan et al., 2016), evaluation of user enjoyment (Abujelala et al., 2016), real time drowsiness detection

(Rohit et al., 2017), measuring focus states (Przegalinska et al., 2018), as well as concentration and stress levels (Maddox et al., 2015)), so far only one study has focused on evaluating its validity (Krigolson et al., 2017). Here, the researchers investigated ERPs using a visual oddball paradigm and reward positivity in a reward learning task. Again, the wearable EEG's data was compared to that of a gel-based device. However, opposite to Badcock et al. (2013) and Zander and Kothe (2011), this study made use of a between-subjects design. One group performed the paradigm while their brain activity was recorded with the wearable device, while the EEG of the control group was recorded with a state-of-the-art research system. The results supported MUSE's validity with regard to both ERP observation and quantification regarding N200, P300, and reward positivity components (Krigolson et al., 2017).

Besides ERP oriented study designs, frequency bands are often used in the context of both research and applied neurofeedback methods for relaxation (Hjelm and Browall, 2000). As a result, frequency analysis is often used as a basis for testing the validity of new EEG systems (e.g., Harland et al., 2002; Estepp et al., 2009; Grozea et al., 2011). Eyes open and eyes closed are two regularly used conditions in this context, due to their distinctive pattern of alpha activity (Kirschstein, 2008). During the eyes open condition, less alpha activity is expected because of the alpha blockade (Berger, 1929; Shaw, 2003). In contrast, the eyes closed condition results in an increase of alpha activity. This difference in alpha band power is especially apparent in electrodes located in close proximity to the occipital lobe (Shaw, 2003; Kirschstein, 2008).

Besides eyes open and eyes closed conditions, belly breathing exercises could pose as an additional experimental task. Such exercises are well-known relaxation techniques. Because relaxation is related to higher alpha activity in the brain (Shaw, 2003), an increase in alpha band power is expected in this condition. Besides tasks that focus on alpha activity, other experimental conditions could be applied in order to validate EEG devices with respect to other frequency bands. For example, brainstorming should elicit higher beta activity (14-30 Hz) due to an increase in cognitive load (Ray and Cole, 1985; Kirschstein, 2008). Thus, this task could potentially be applied to validate new EEGs with regard to the beta frequency band. However, beta activity can be found in a variety of locations across the cortex (Kropotov, 2009). This makes it harder to detect beta activity with a restricted electrode setup, as is often the case with consumer-oriented EEG systems. Even if the activity is harder to detect though, the wearables measurements should still be comparable to those of a gel-based device.

Even though MUSE was created and is often used for frequency dependent neurofeedback (e.g., Rohit et al., 2017), to this point no evidence exists that supports the system's reliability and validity with respect to frequency analysis. The first study in this thesis, therefore, aimed at examining both the reliability and validity of the MUSE wearable with regard to alpha and beta frequency domains across different experimental conditions (i.e., eyes open, eyes closed, breathing and brainstorming).

However, as mentioned above, the device's validity and reliability are not the only aspects that determine how MUSE is received by users. Rather, its usability, user experience, and perceived comfort majorly define whether users adopt the system for neurofeedback training in the end (Müller et al., 2008; McDowell et al., 2013; De Vos et al., 2014; Hairston and Lawhern, 2015). This, in turn, determines the device's applicability for stress management applications. A previous study already tested MUSE's usability for applied healthcare research. Here, a researcher rated the device's ease of use and form factor to be medium (Leape et al., 2016). However, the interpretation of the results is limited due to the use of a single rater and narrow evaluation criteria. Thus, the second study presented here focused on the evaluation of these aspects with a larger sample, as well as on determining how users perceive the neurofeedback supported meditation exercises that are provided in the MUSE companion app.

2.3.2.2 Study 1: Evaluation of MUSE's Reliability and Validity

2.3.2.2.1 Methods

Paradigm

The EEG-signal was recorded simultaneously with the wearable EEG system MUSE by InteraXon and the research-grade gel-based EEG system actiCAP by Brain Products. The experiment consisted of two sessions, with a 30 minute break between sessions. Each session entailed four experimental conditions: eyes open, eyes closed, breathing, and brainstorming. Spontaneous EEG features were recorded for each of these experimental tasks, causing different band powers for the alpha (8–13 Hz) and beta band (14-30 Hz) (e.g., Birbaumer and Schmidt, 2006). The four conditions were aimed to compare MUSE's performance across different experimental tasks with that of the gel-based research EEG. By performing every condition once in each of the two sessions, the test-retest reliability of the wearable EEG was tested. The parallel recording of the EEG signal with both the wearable and gel-based system allowed to compare the systems' reliabilities and to estimate convergent validity, as specific patterns of alpha and beta activity were expected for each experimental task (eyes open, eyes closed, breathing and brainstorming).

In the eyes open condition, participants were asked to think of nothing in particular and focus on a fixation cross placed on the wall in front of them. Because of the alpha blockade (Shaw, 2003), it was expected to find less alpha activity in this condition as compared to the eyes closed condition (Berger, 1929).

In the eyes closed condition, participants were again asked to think of nothing in particular and keep their eyes closed. Due to the closed eyes, larger alpha band power was expected in this condition compared to the eyes open condition (Shaw, 2003; Kirschstein, 2008). Both eyes open and eyes closed are tasks commonly used to validate new EEG systems with regard to the alpha frequency band (e.g., Estepp et al., 2009, Harland et al., 2002; Grozea et al., 2011).

In the breathing condition participants were instructed to close their eyes, breathe deeply through their belly and count their breath from 1 to 10. Once they arrived at count 10 or if they lost track of their count they started to count from the beginning and so on. Belly breathing is a well-known relaxation technique and relaxation is related to higher alpha activity in the brain (Shaw, 2003). However, because the participants were not trained in belly breathing alpha power in this condition was expected to be similar to that of the other conditions with closed eyes (brainstorming and eyes closed).

During the brainstorming condition, participants were asked to keep their eyes closed. They were then given a word category (e.g., countries), for which they had to find as many matching words as possible in their head, without speaking them out loud, within one minute. Then they were given the next word category. The condition covered five word categories and resulted in a total of five minutes of recording time. To eliminate repetition effects, different words were used in both sessions. As cognitive load causes an increase in beta activity, it is expected that this task results in a higher beta band power compared to the other conditions (Ray and Cole, 1985; Kirschstein, 2008).

Each experimental task lasted five minutes. After each condition participants were asked whether they had followed the instructions. All experimental tasks were performed in both experimental sessions. The order of the tasks was randomized across participants. Event markers were used to mark the beginning and end of an experimental condition in the data of the gel-based EEG's signal. The study procedure is illustrated in Figure 6.

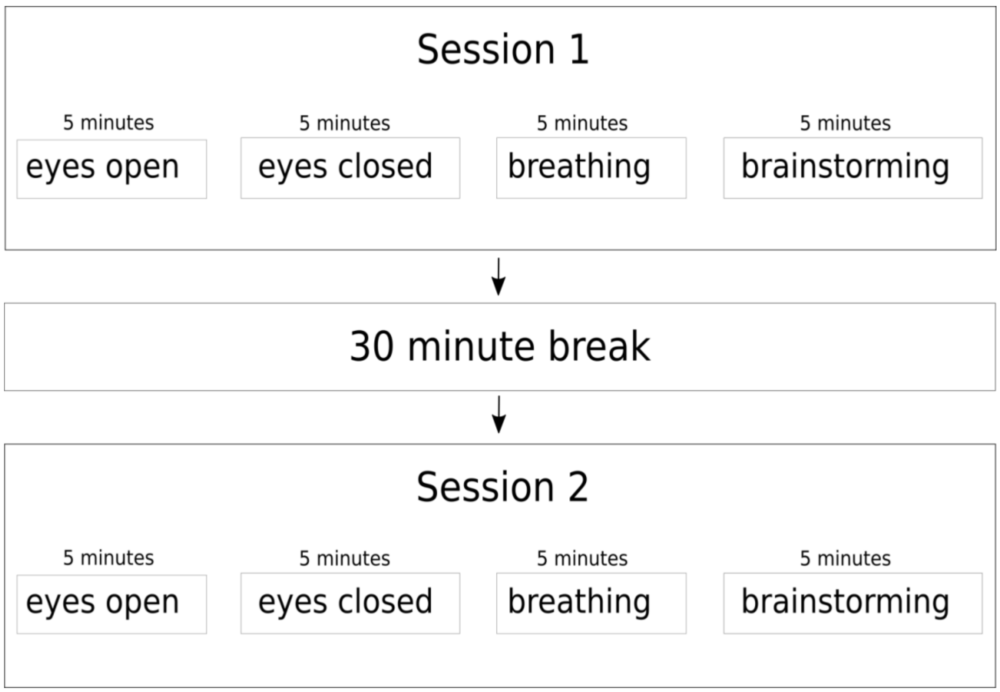

Figure 6: Diagram of the study procedure. The order the conditions were presented in was randomized across all participants.

Participants

The ethics committee of the Department of Social Sciences at the Technische Universität Kaiserslautern approved the methods used to test participants in this study. All subjects gave written informed consent in accordance with the Declaration of Helsinki and received 15 € for their participation. A total of 20 participants took part in this study. All participants were neurologically healthy. One participant had to be excluded due to repeated signal-loss in the electrodes of the gel-based EEG. The data of 19 participants (age range 19-29 years, mean (M) age = 24.1 years, standard deviation (SD) = 2.7 years, right-handed, 10 female) was pre-processed and used for statistical analysis.

Study Procedure

For this study, all participants were seated in an electronically shielded cubicle. Before attaching the electrodes, participants' skin was cleaned on the forehead and behind the ears, using first an abrasive gel and then alcohol. Afterwards, the electrodes of the two EEG systems (the wearable consumer EEG system MUSE by InteraXon and the gel-based research-grade EEG system actiCAP by Brain Products) were attached. First the wearable device was placed. The electrode connection for the wearable EEG was verified with the visual system provided by InteraXon. The EEG signal of the wearable EEG was recorded using MUSELab by InteraXon with a sampling rate of 220 Hz.

Next, the electrodes of the gel-based system were attached in a 1.5-2 cm distance to those of the wearable using adhesive rings. The electrodes were placed at the locations A1, A2, Fp1, and Fp2, grounding and reference electrodes were placed at Fpz according to the international 10-20 system (Figure 7), because this roughly matches the position of the four electrodes of MUSE. After attaching the electrodes to the scalp, conductive gel was applied to the gel-based system in order to improve connectivity. The electrode impedances were kept below 5 kΩ. The EEG signal of the gel-based system was amplified with actiCHamp and recorded using the Brain Vision Analyzer software by Brain Products with a sampling rate of 500 Hz. Participants were asked to sit still and to avoid moving as much as possible during recording.

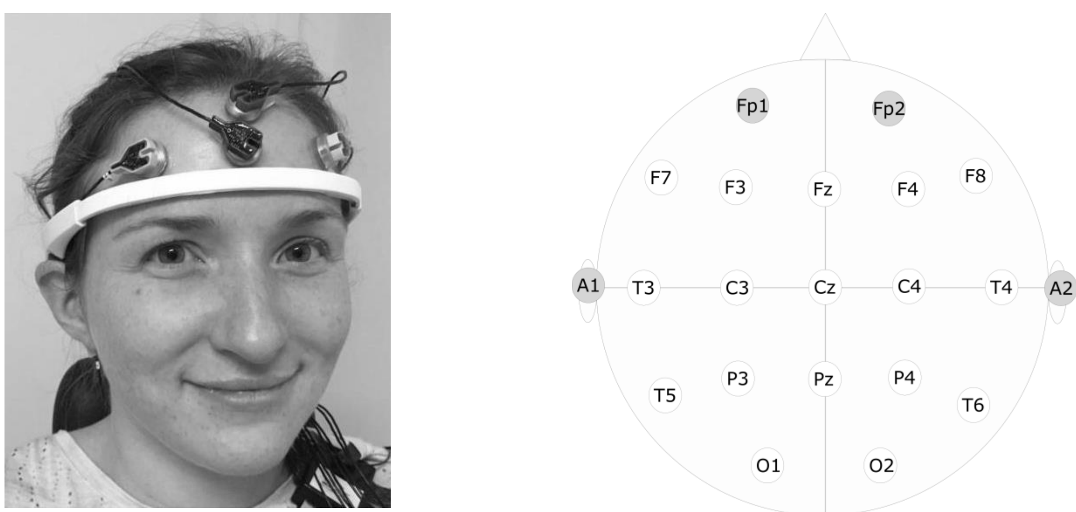

Figure 7: Electrode placement according to the international 10-20 system. The locations used in this study are highlighted in grey.

Pre-processing

MATLAB (version R2016a) and the toolboxes EEGlab (Delorme and Makeig, 2004) and ERPlab (version 13_6_5b) (Lopez-Calderon and Luck, 2014) were used to pre-process and analyze the data. The data from the time of the trigger to 300 seconds following the trigger for the eyes open, eyes closed, and breathing condition, were used for analysis. For the brainstorming condition, the data following 60 seconds after each trigger's occurrence (5 times one minute of brainstorming) was used for analysis. This, too, added up to a total of 300 seconds of recorded data.

Because the sampling rate of the wearable was set to 220 Hz, the sampling rate of the gel-based data was also reduced to 220 Hz during data pre-processing. A high pass filter at 1 Hz was used to eliminate baseline drifting and to reduce activity at very low frequencies, which is often an artifact. A powerline notch filter at 50 Hz was applied to remove artifacts stemming from electrical instruments in the vicinity (Libenson, 2010).

In the next step, a spectral analysis on the resulting data following Welch's method was conducted for frequencies between 0.1-110 Hz. This resulted in power signal density (PSD) values for all sessions, devices, conditions, and electrodes for each participant. In addition, the mean power of the alpha and beta frequency bands was derived.

Statistical Analysis

Pearson correlations between the two sessions for each device and the two devices in each session were calculated for the PSDs of each participant for each condition and each channel, as indicators for test-retest reliability. The resulting correlation coefficients were then averaged across all participants. High correlation coefficients (> .90) speak of high reliability, values between .80 - .90 are interpreted to be moderate and < .80 as insufficient reliability (Vincent, 2008).

Univariate analyses of variance (ANOVA) with repeated measures and subsequent t-tests were used to detect differences between the wearable device and gel-based system with regard to alpha and beta band powers (Bortz, 2016). A five-way ANOVA with the factors device (wearable and gel-based), channel (A1, A2, Fp1, Fp2), condition (eyes open, eyes closed, breathing, and brainstorming), frequency band (alpha and beta) and session (one and two), were conducted. The Greenhouse-Geisser adjustment was applied to correct for violations of sphericity. T-tests were calculated to further analyze the interaction effects. Bonferroni correction was used to adjust the t-test results for an accumulation of alpha errors (Field, 2009). The effect size (*d*) according to Cohen (1988) was calculated. Effects d = 0. 20 - 0.40 correspond to a small effect, d = 0.50 – 0.70 a moderate effect and d ≥ 0.80 a large effect.

Hypotheses

Reliability

H1: The data of session one and session two show high positive correlations (> .90) (Vincent, 2008) between the PSDs for both the wearable and the gel-based device, in all four channels and all four conditions.

H2: The ANOVA reveals no main effect of session and no interaction of session with any of the other factors.

H3: High positive correlations (> .90) (Vincent, 2008) are found between the PSDs of the wearable and gel-based device in both sessions, all four conditions, and all four channels.

Validity

H4: Both devices show reduced alpha activity in the eyes open condition compared to the eyes closed, breathing, and brainstorming conditions due to the alpha blockade (Shaw, 2003; Kirschstein, 2008).

H5: Both devices show higher alpha activity at the mastoids (A1 and A2) than at the frontal locations (Fp1 and Fp2) in the eyes closed, breathing, and brainstorming conditions compared to the eyes open condition, due to the alpha blockade (Shaw, 2003; Kirschstein, 2008).

H6: Both devices show comparable alpha activity in the breathing, eyes closed, and brainstorming condition, due to closed eyes (Shaw, 2003).

H7: Both devices show higher beta activity in the brainstorming condition compared to the eyes open, eyes closed, and breathing conditions (Ray and Cole, 1985; Kirschstein, 2008).

H8: There exist no differences between the two devices in the mean alpha and beta power measured in each condition and channel.

2.3.2.2.2 Results

Reliability

Both the gel-based and wearable device showed high positive correlations (all $r \geq .90$) between session one and two for the PSDs in all electrodes in all conditions (see Table 2). Correlation coefficients between the two devices were low (≤ .80) to high (≥ .90) for all electrodes and conditions in session one and two (see Table 3). The ANOVA showed no main effect of the factor session and no interactions of session with any of the other factors (see Appendix 1).

Table 2: Means (*M*) and standard deviations (*SD*) of Pearson correlation coefficients (*r*) between the power spectral densities of 0.1-110 Hz of session 1 and 2, regarding each experimental condition, device, and channel. All p-values were ≤ .001.

	Eyes open		Eyes closed		Breathing		Brainstorming	
	Wearable	**Gel-based**	**Wearable**	**Gel-based**	**Wearable**	**Gel-based**	**Wearable**	**Gel-based**
Fp1	*M* = .925; *SD* = .076	*M* = .936; *SD* = .069	*M* = .955; *SD* = .137	*M* = .983; *SD* = .024	*M* = .906; *SD* = .052	*M* = .983; *SD* = .046	*M* = .953; *SD* = .045	*M* = .993; *SD* = .007
Fp2	*M* = .934; *SD* = .103	*M* = .974; *SD* = .032	*M* = .956; *SD* = .068	*M* = .977; *SD* = .036	*M* = .943; *SD* = .070	*M* = .977; *SD* = .050	*M* = .937; *SD* = .127	*M* = .988; *SD* = .012
A1	*M* = .946; *SD* = .145	*M* = .974; *SD* = .032	*M* = .950; *SD* = .046	*M* = .964; *SD* = .047	*M* = .960; *SD* = .095	*M* = .964; *SD* = .089	*M* = .965; *SD* = .047	*M* = .989; *SD* = .013
A2	*M* = .956; *SD* = .058	*M* = .974; *SD* = .032	*M* = .974; *SD* = .064	*M* = .972; *SD* = .037	*M* = .945; *SD* = .024	*M* = .972; *SD* = .050	*M* = .957; *SD* = .088	*M* = .988; *SD* = .012

Table 3: Means (*M*) and standard deviations (*SD*) of the Pearson correlation coefficients (*r*) between the power spectral densities of 0.1-110 Hz of the two devices regarding each experimental condition, session, and channel. All p-values were ≤ .001.

	Eyes open		Eyes closed		Breathing		Brainstorming	
	Session 1	**Session 2**	**Session 1**	**Session 2**	**Session 1**	**Session 2**	**Session 1**	**Session 2**
Fp1	*M* = .800; *SD* = .207	*M* = .746; *SD* = .261	*M* = .897; *SD* = .136	*M* = .910; *SD* = .042	*M* = .878; *SD* = .082	*M* = .922; *SD* = .039	*M* = .873; *SD* = .062	*M* = .866; *SD* = .090
Fp2	*M* = .823; *SD* = .229	*M* = .748; *SD* = .247	*M* = .926; *SD* = .078	*M* = .901; *SD* = .049	*M* = .900; *SD* = .065	*M* = .921; *SD* = .076	*M* = .815; *SD* = .164	*M* = .864; *SD* = .088
A1	*M* = .938; *SD* = .031	*M* = .932; *SD* = .027	*M* = .912; *SD* = .041	*M* = .923; *SD* = .024	*M* = .910; *SD* = .059	*M* = .934; *SD* = .037	*M* = .908; *SD* = .042	*M* = .923; *SD* = .036
A2	*M* = .921; *SD* = .046	*M* = .922; *SD* = .032	*M* = .910; *SD* = .046	*M* = .925; *SD* = .043	*M* = .888; *SD* = .047	*M* = .919; *SD* = .046	*M* = .919; *SD* = .039	*M* = .932; *SD* = .026

Validity

The ANOVA revealed main effects of the factors device, channel, task, and frequency band. Twofold interactions were detected between channel*frequency band, channel*task, frequency band*task, channel*device, frequency band*device, and task*device. Triple interactions were found for channel*task*device, and frequency band*task*device. A fourfold interaction was detected for channel*frequency band*task*device (see Appendix 1).

Consistently, t-tests revealed that alpha activity was higher at the mastoids than at the frontal electrodes (all $p \leq .001$, see Appendix 2 for details). Higher beta band power was detected in Fp2 ($t(18)= 3.28$, $p = .004$, $d = 0.75$) and A1 ($t(18) = 3.85$, $p = .001$, $d = 0.86$) compared to A2.

In line with the interaction between frequency band*task, t-tests revealed reduced alpha activity in the eyes open condition compared to the eyes closed, breathing, and brainstorming tasks (all $p \leq .001$, see Appendix 4). However, this effect was only observed in the research-grade device (see Appendix 9 for details). Also, alpha activity was lower during brainstorming than during the eyes closed or breathing task (all $p \leq .001$). Beta band power was also lower during brainstorming than during eyes closed ($t(18) = 3.76$, $p = .001$, $d = 0.86$). See Appendix 4 for details.

Lower overall power was detected at the mastoids during eyes closed compared to the other three experimental conditions (all $p \leq .001$). Moreover, higher power was detected in A1 and A2 during eyes closed than during brainstorming (both $p \leq .001$). Also see Appendix 3 for details.

Overall, the power measured by the two devices did not differ at the frontal electrodes. The electrodes at the mastoids both showed significantly less power for the gel-based device than for the wearable (both $p \leq .001$, see Appendix 5 for details). However, the mean power was higher for MUSE during eyes closed, breathing, and brainstorming, while it was lower during the eyes open task (see Appendix 7). In line with this, MUSE also resulted in a higher mean alpha ($t = 3.55$, $p = .002$, $d = 0.81$) and mean beta band power ($t = 4.53$, $p \leq .001$, $d = 1.04$) than the research-grade EEG system (also see Appendix 6). Moreover, power differed between all electrodes of both devices in all four experimental conditions (see Appendix 8).

To analyze the fourfold channel*frequency band*task*device interaction, an additional four ANOVAs (one for each channel) were performed with the factors frequency band*task*device. All four electrodes showed an interaction of frequency band*task*device. Fp1, Fp2 and A2 showed a main effect of frequency band. Main effects of device and task, as well as an interaction between frequency band*device were revealed for A1 and A2. Fp1, Fp2, and A1 showed a task*device interaction (see Appendix 10).

Though the wearable showed lower alpha activity during the eyes open condition compared to the other three tasks in Fp1 and Fp2, these differences were not found for the gel-based device (see Appendices 11 and 12 for details). Both devices showed lower alpha power at A1 and A2 during the eyes open condition than during eyes closed, breathing and brainstorming (all $p \leq .001$). Lower alpha activity was also detected during brainstorming than eyes closed and breathing in A1 and A2 of both devices (see Appendices 13 and 14 for details).

T-tests did not uncover differences in the alpha power measured by the two devices in all four experimental conditions in Fp1 (Appendix 11). A similar pattern was observed for Fp2, with the exception of the eyes open task, where alpha power was higher in the gel-based device than in MUSE ($t = 2.77$, $p = .01$, $d = 0.64$, see Appendix 12 for details). The opposite was observed for A1 and A2. Though no differences were found between the two systems for these electrodes during the eyes closed and brainstorming tasks, alpha power was significantly higher in MUSE during the eyes open condition (see Figure 8 and Appendices 13 and 14). Moreover, alpha power was also higher in A1 of the wearable during breathing.

Though beta band power increased in Fp1 and Fp2 during the eyes open condition compared to the other three tasks for the gel-based device, this pattern did not exist for the wearable (see Appendices 11 and 12 for details). This is further supported by the differences that were uncovered between the beta power measured in each condition with the research-grade system and MUSE. For A1 and A2, the research-grade EEG showed higher beta activity in the eyes closed and breathing condition compared to brainstorming (all $p \leq .001$, see Appendices 13 and 14).

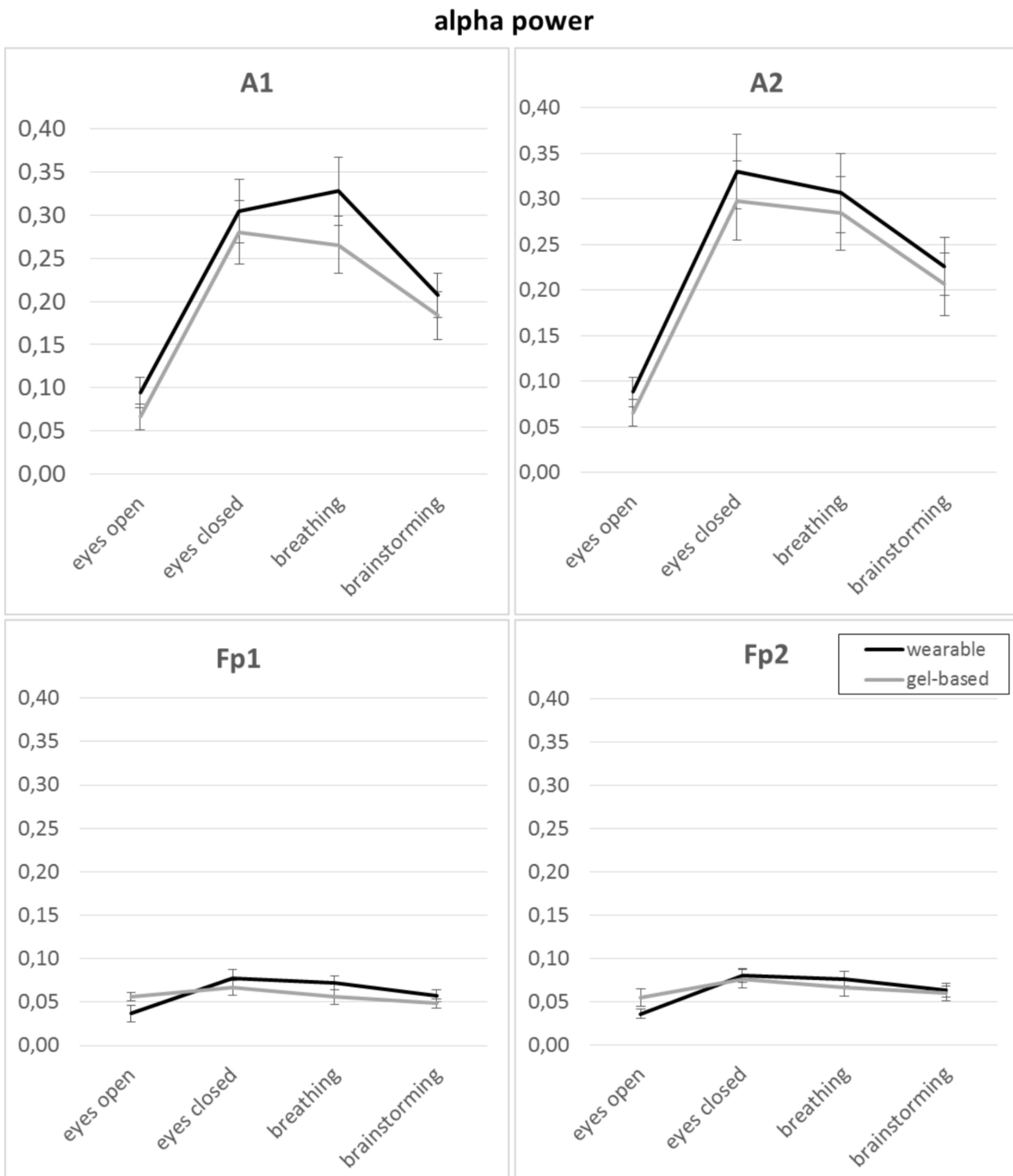

Figure 8: Means and standard deviations for the alpha power measured by each of the two devices (wearable EEG MUSE and the gel-based EEG) in each electrode (A1, A2, Fp1, and Fp2) during each of the four experimental conditions (eyes open, eyes closed, breathing, and brainstorming).

In contrast, MUSE only measured higher beta activity during breathing than brainstorming in A1 (t = 3.87, $p \leq .001$, d = 0.81) and in the eyes closed condition compared to brainstorming in A2 (t = 5.05, $p \leq .001$, d = 1.16). Additional differences in beta band power were detected for Muse between the eyes open task and breathing (A1 and A2), eyes closed (A2) and brainstorming (A2). This result is further supported by the fact that beta power was higher in MUSE during the eyes open condition compared to the gel-based device (see Figure 9 and Appendices 13 and 14).

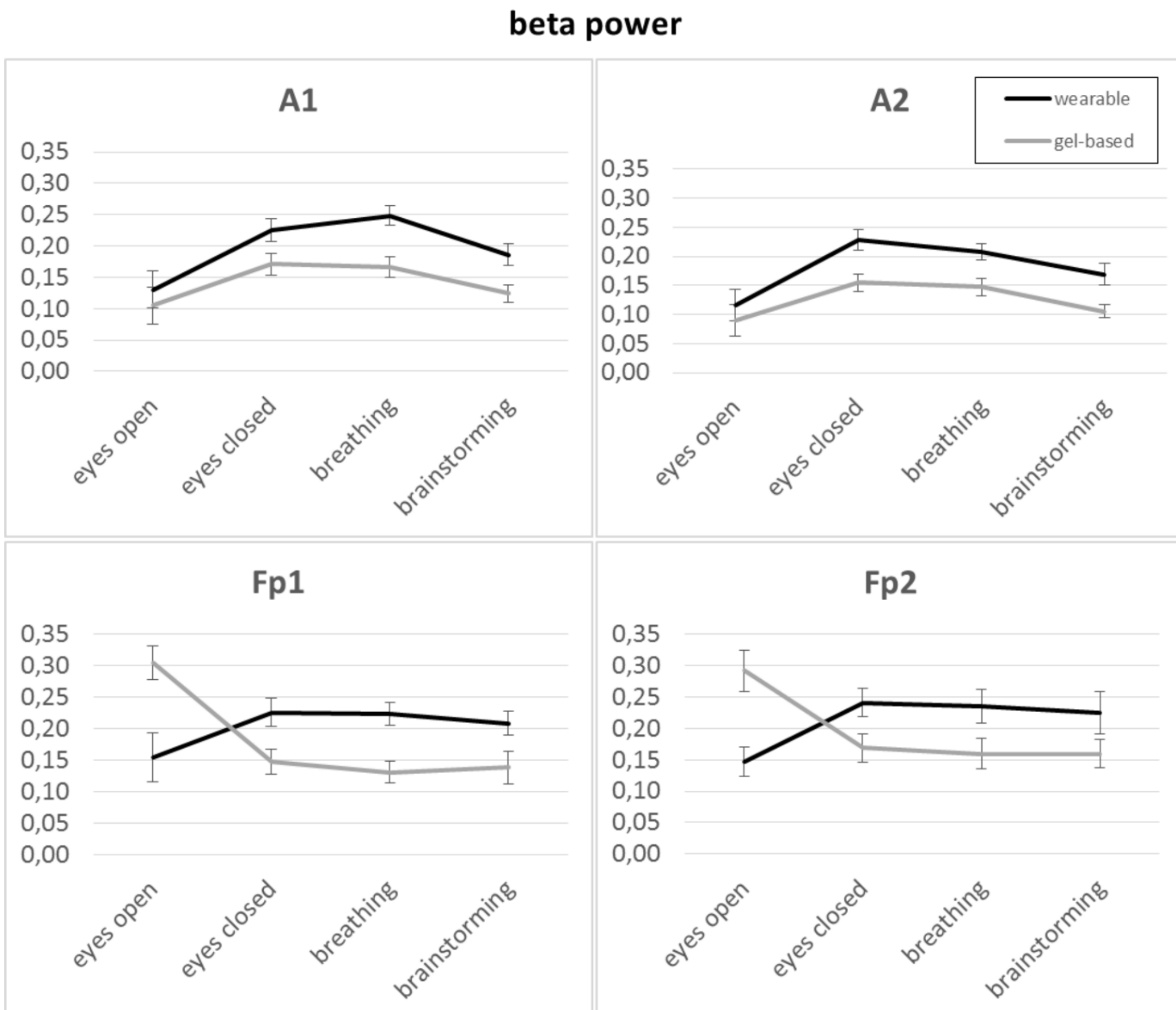

Figure 9: Means and standard deviations for the beta power measured by each of the two devices (wearable EEG MUSE and the gel-based EEG) in each electrode (A1, A2, Fp1, and Fp2) during each of the four experimental conditions (eyes open, eyes closed, breathing, and brainstorming).

2.3.2.2.3 Discussion

The results show that with regard to the systems' reliability, H1 and H2 could be confirmed. The high positive correlations (all $r \geq .90$) between the PSDs in session one and two that were uncovered for all four electrodes in all four experimental conditions speak for a high test-retest reliability of both devices (Vincent, 2008). This is further confirmed by the fact that the ANOVA revealed no main effect of session and no interactions of session with any of the other factors. Even though previous results show that high correlations between the wearable and research-grade EEGs are possible (Estepp et al., 2009; Zander and Kothe, 2011), the correlations between MUSE and the research-grade device differed strongly. Although most were high, medium and low correlations were also observed, especially in Fp1 and Fp2 during the eyes open and brainstorming condition, but also during the breathing task in Fp1 and A2. These variances indicate that there are severe differences in the measurements of the two devices. Thus, H3 could not be confirmed. This could be due to MUSE being more affected by signal loss, given the fact that wearable setups still face difficulties, such as motion artefacts (Chi et al., 2010).

In line with H4, alpha powers were significantly lower in the eyes open condition due to the alpha blockade, compared to the other three conditions in which participants kept their eyes closed (Shaw, 2003). Also, higher alpha power was detected at the mastoids compared to the frontal locations during eyes closed, breathing, and brainstorming than during the eyes open condition. This was expected (H5), as closed eyes lead to less visual input and subsequently to higher alpha activity in the brain's visual cortex (Berger, 1929; Shaw, 2003; Birbaumer and Schmidt, 2006). As expected, alpha activity was comparable during the eyes closed, breathing, and brainstorming conditions (Shaw, 2003). This confirms H6. Because brainstorming leads to a higher cognitive load, this task was expected to result in higher beta band powers (H7) in both devices. This was not confirmed by the results. A cause for this could be the fact that beta activity may be found in a variety of locations across the cortex, including the central areas (Kropotov, 2009). As the present electrode setup only covered the mastoids and frontal areas of the scalp, it is quite possible that the electrodes simply did not pick up on these signals.

Though there were differences with regard to the systems alpha activity during the eyes open task, the two devices were comparable in their performance during the other three experimental conditions. This goes in line with another study that also found variabilities between two investigated devices for ERPs in specific experimental conditions (Badcock et al., 2013). Still, akin to the results of Grozea et al. (2011), the two devices used in this study were very similar with respect to their measurements of alpha band power.

On the other hand, MUSE and the gel-based system differed more strongly with respect to beta activity. For example, the research-grade device showed an increase in beta activity in Fp1 and Fp2 during the eyes open condition. This was not the case for MUSE. This goes in line with previous results that showed that the investigated devices differed more strongly at electrodes located at the back of the head (Estepp et al., 2009) and over the left and right temporal lobe (Bleichner et al., 2016). Also, the beta band power in this study was higher in MUSE than in the gel-based EEG during the eyes open task. Thus, the main effect of device and the interactions of device with the other factors of the ANOVA indicates that H8 can only be confirmed in part. Still, previous studies demonstrate that such differences between the signals of two EEG systems are not uncommon (Krigolson et al., 2017; Melnik et al., 2017). Variability in amplifiers and contact materials, movement interference, mechanical pressure or electrode location are mentioned as possible reasons for variances across two EEG devices (Estepp et al., 2009; Badcock et al., 2013; Fiedler et al., 2014; Bleichner et al., 2016).

However, while Grozea et al. (2011) reported that the two devices used in their study were very similar for alpha and beta frequency band powers, this study found similarities between the two devices mostly with respect to alpha activity. This indicates that MUSE might be less suitable for applications involving the beta frequency band and demonstrates that MUSE’s validity strongly depends on the context of its application and targeted frequency domain aspects. Furthermore, this study only investigated the validity and reliability of the system in a stationary sitting position. Future studies should look into the device’s suitability for other tasks and for daily life applications that involve movement to determine its motion tolerance. Real world evidence of such recordings is still scarce (Wang et al., 2017), but one study indicates that recordings with MUSE under real-life conditions is noisy and unstable (Przegalinska et al., 2018). Though problems such as potential signal loss and movement artifacts might not be a problem for the analysis of frequencies over longer time periods, it remains a question for further research whether this poses an actual problem for other application scenarios (Gramann et al., 2011).

Besides the system’s reliability and validity, the study also revealed other areas of concern. For one, MUSE is still not completely unobtrusive and concealed. Hence, it remains to be seen whether the system is capable of meeting the users’ demands (McDowell et al., 2013). Furthermore, two study participants complained about the system being uncomfortable to wear for a longer period of time, showing a need to further examine the system's design and usability (Müller et al., 2008). In addition, though first evidence exists that neurofeedback exercises can result in the user's relaxation (Hjelm and Browall, 2000), it is still unknown whether this approach is wanted and accepted by users. For this reason, the next study investigated the usability, user experience, and general quality of MUSE and its companion app, which provides neurofeedback-based meditation exercises.

2.3.2.3 Study 2: Evaluation of the Usability, User Experience and General Quality of the MUSE Wearable and Its Neurofeedback-Based Companion App

2.3.2.3.1 Methods

Participants

A total of 19 participants (8 male/ 11 female, ages 22-29 years, M = 25.21, SD = 2.74 years) were recruited via an e-mail distribution list of the Technische Universität Kaiserslautern. Seven of the participants were currently employed, 11 were university students. All participants gave written consent to participate in accordance with the declaration of Helsinki. The study procedure was approved by the ethics committee of the Department of Social Sciences at the Technische Universität Kaiserslautern. In order to take part in the study, participants had to be between 18-30 years of age, be fluent in German (all app instructions were only available in German), not meditate regularly, and not have prior experience with the MUSE companion app.

Study Procedure

Each participant tested the MUSE app on a Lenovo TB-4706F tablet, according to a specific protocol (see Figure 10). Before testing the app, after connecting the wearable EEG to the app, and after finishing the second meditation, the participant rated his or her emotional state on three scales: valence, arousal and dominance, using the Self-Assessment Manikin (SAM, Bradley and Lang, 1994).

After testing, the well-established System Usability Scale (SUS, Brooke, 1996) was used to assess the usability of the MUSE system (combination of MUSE app and MUSE wearable EEG). The questionnaire provides reliable results even for small samples (Sauro, 2011). It consists of 10 questions that the participant evaluated on a five-point Likert scale from 1 (strongly disagree) to 5 (strongly agree).

In addition, the User Experience Questionnaire (UEQ, Laugwitz et al., 2008) was used to appraise the system's user experience. The UEQ includes 26 pairs of opposites that are rated on a seven-point scale and assess a system's attractiveness, perspicuity, dependability, efficiency, stimulation and novelty. These aspects can be further summarized in three major categories: attractiveness, pragmatic quality and hedonic quality.

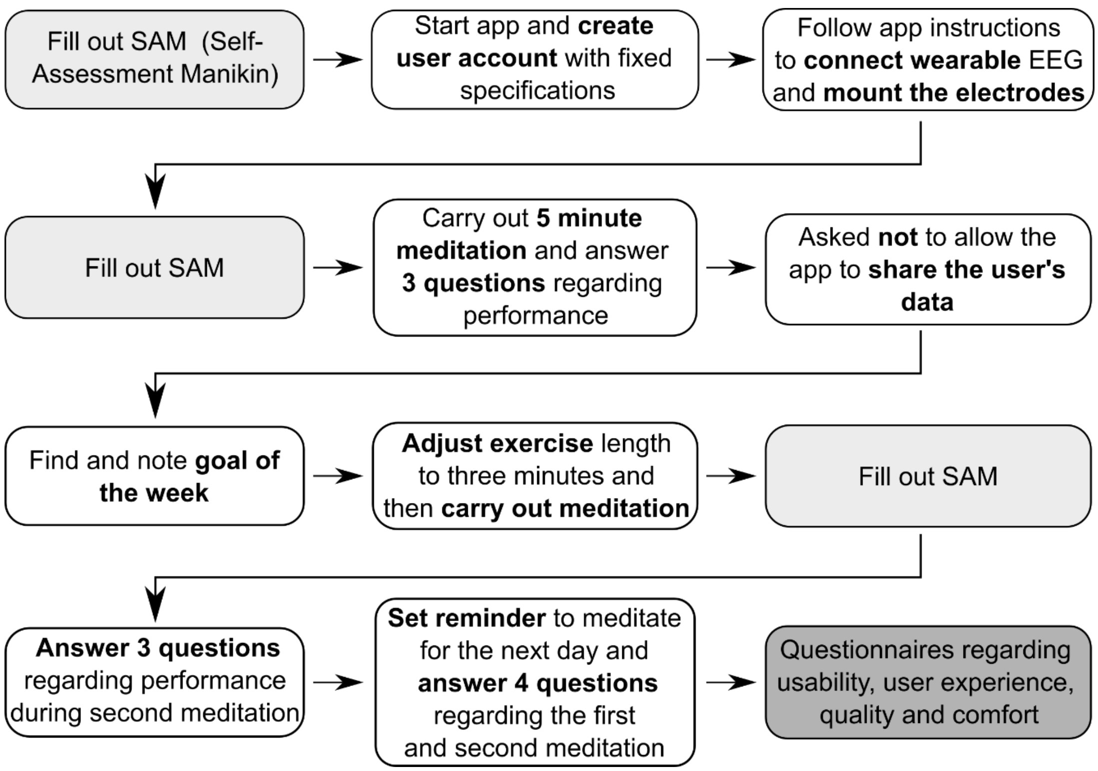

Figure 10: Study procedure for testing the wearable EEG MUSE and its companion app.

The MUSE app's perceived quality was assessed with the user Mobile Application Rating Scale (uMARS, Stoyanov et al., 2016). The uMARS comprises 26 questions from three categories: subjective quality, perceived impact, and general quality. Participants rated each question on a five-point Likert scale from 1 (inadequate) to 5 (excellent). Questions from the category "perceived impact" were adapted, i.e. the term "health behavior" was replaced with "meditation".

The Comfort Rating Scales (CRS, Knight et al., 2002) were used to evaluate the comfort of the MUSE wearable. The CES appraises a wearable's comfort in six categories (emotion, attachment, harm, perceived change, movement, and anxiety) through the participant's agreement with 11 statements. The participant rated his or her agreement with the statements on a five-point scale from 1 (strongly agree) to 5 (strongly disagree).

Moreover, seven additional questions regarding the wearable's continued use, daily use, ease of use, visual appeal, use in private and public, as well as the willingness to pay for the device were assessed by the participant on a five-point scale from 1 (strongly agree) to 5 (strongly disagree). Also, three open ended questions that targeted the system's best and worst features, as well as additional comments, were included to allow for differentiated user feedback.

Statistical Analysis

Because the assumption of normal distribution was violated (all $p < .05$), differences in SAM ratings during the study were assessed with Friedman tests instead of repeated measures ANOVA. Wilcoxon signed-rank tests were used instead of t-tests for post-hoc analysis (Field, 2009). Bonferroni correction was used to adjust the statistical results for an accumulation of alpha errors (new alpha-level $p \leq .02$). The effect size (r) according to Cohen (1992) was calculated. Effects $r = 0.10 – 0.30$ correspond to a small effect, $r = 0.30 - 0.50$ a moderate effect, and $r \geq 0.50$ a large effect.

2.3.2.3.2 Results

The testing of the system took approximately 36 minutes. When answering the 11 questions with regard to the session summary and interpretation of the data, the participants made an average of $M = 1.2$ errors ($SD = 0.90$).

No differences were found between the participants' reported valence ($\chi^2(2) = 1.56$, $p = .47$) and dominance ($\chi^2(2) = 3.87$, $p = .16$) ratings. However, Friedman tests revealed differences with regard to arousal (χ^2 (2) = 15,49, $p \leq .001$). Arousal levels decreased after the second meditation, compared to the baseline rating ($Z = 2.45$, $p = .01$, $r = 0.56$) and the rating the participants gave after connecting the wearable EEG to the MUSE app ($Z = 2.47$, $p = .01$, $r = 0.57$). See Figure 11 for details.

The system's usability was rated as good, reaching 76% of the maximum possible SUS score (Bangor et al., 2009). The user experience showed mixed results, with the systems attractiveness, perspicuity, efficiency, and novelty being rated as good, stimulation as below average, and dependability as bad (see Figure 12). Notwithstanding the mixed results in the subscales, the means for the systems attractiveness ($M = 1.47$, $SD = 0.83$), pragmatic quality ($M = 1.05$, $SD = 1.00$) and hedonic quality ($M = 0.96$, $SD = 0.88$) were all above 0.8, thus speaking for a positive evaluation of the system's user experience (Schrepp, 2015).

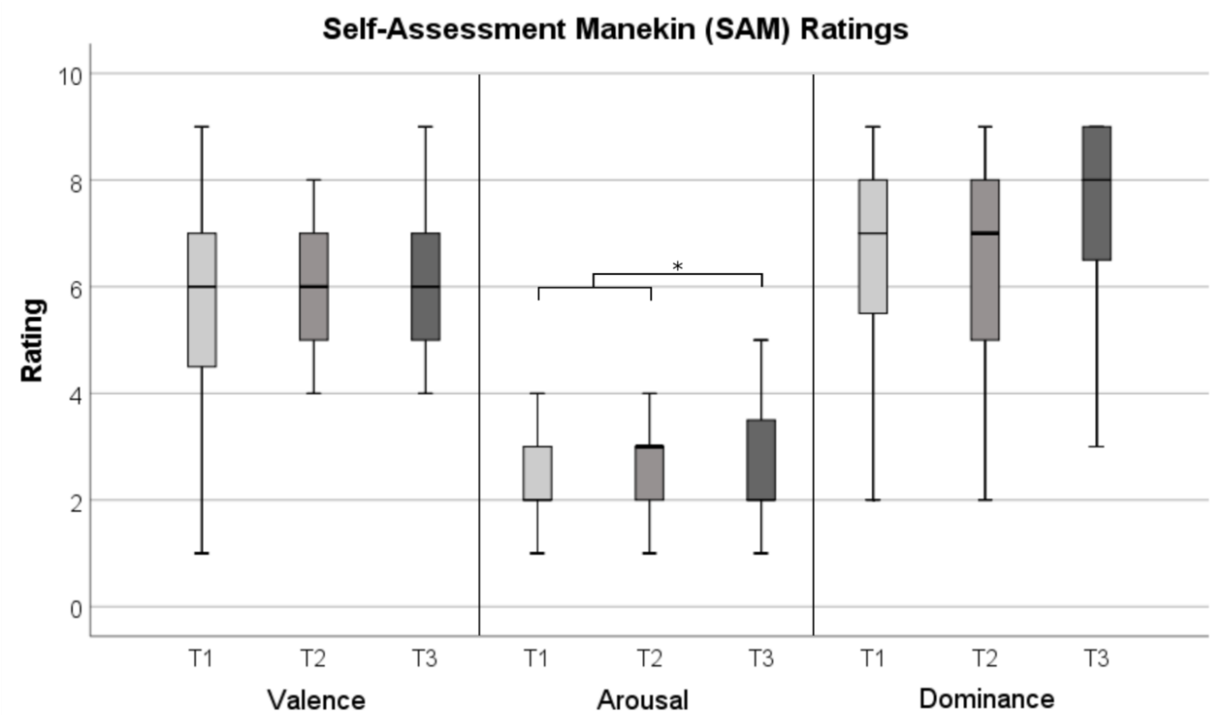

Figure 11: Self-Assessment Manekin ratings at the beginning of the study (T1), after connecting the wearable EEG MUSE to the MUSE companion app (T2), and after finishing the second meditation (T3). Depicted are the medians, maximum and minimum values, as well as first and third quartiles. Significant differences are marked with an asterisk (*).

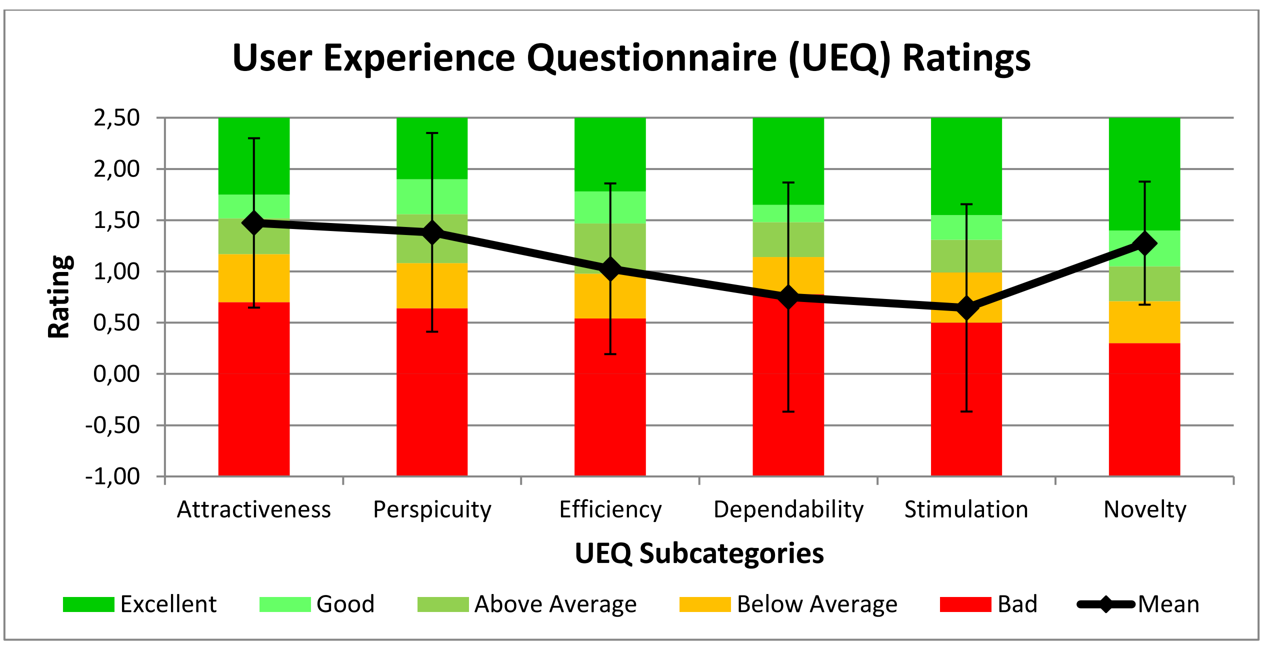

Figure 12: Means and standard deviations of the User Experience Questionnaire (UEQ) ratings for the MUSE system (wearable and app).

MUSE app's general quality was rated to be good (M = 3.76, SD = 0.41 out of 5). The subcategories entertainment and functionality mirrored this result. Even though the subcategory information was also perceived as good, its scale assessing the credibility of the app's source only received an average rating.

Additionally, the subscale aesthetics was only rated as average. Though its scales regarding the MUSE app's layout and visual appeal were rated to be good, the app's graphics were perceived as very poor. The app's subjective quality (M = 3.22, SD = 0.67 out of 5) and perceived impact (M = 2.84, SD = 0.87 out of 5) were assessed to be average. See Figure 13 for details.

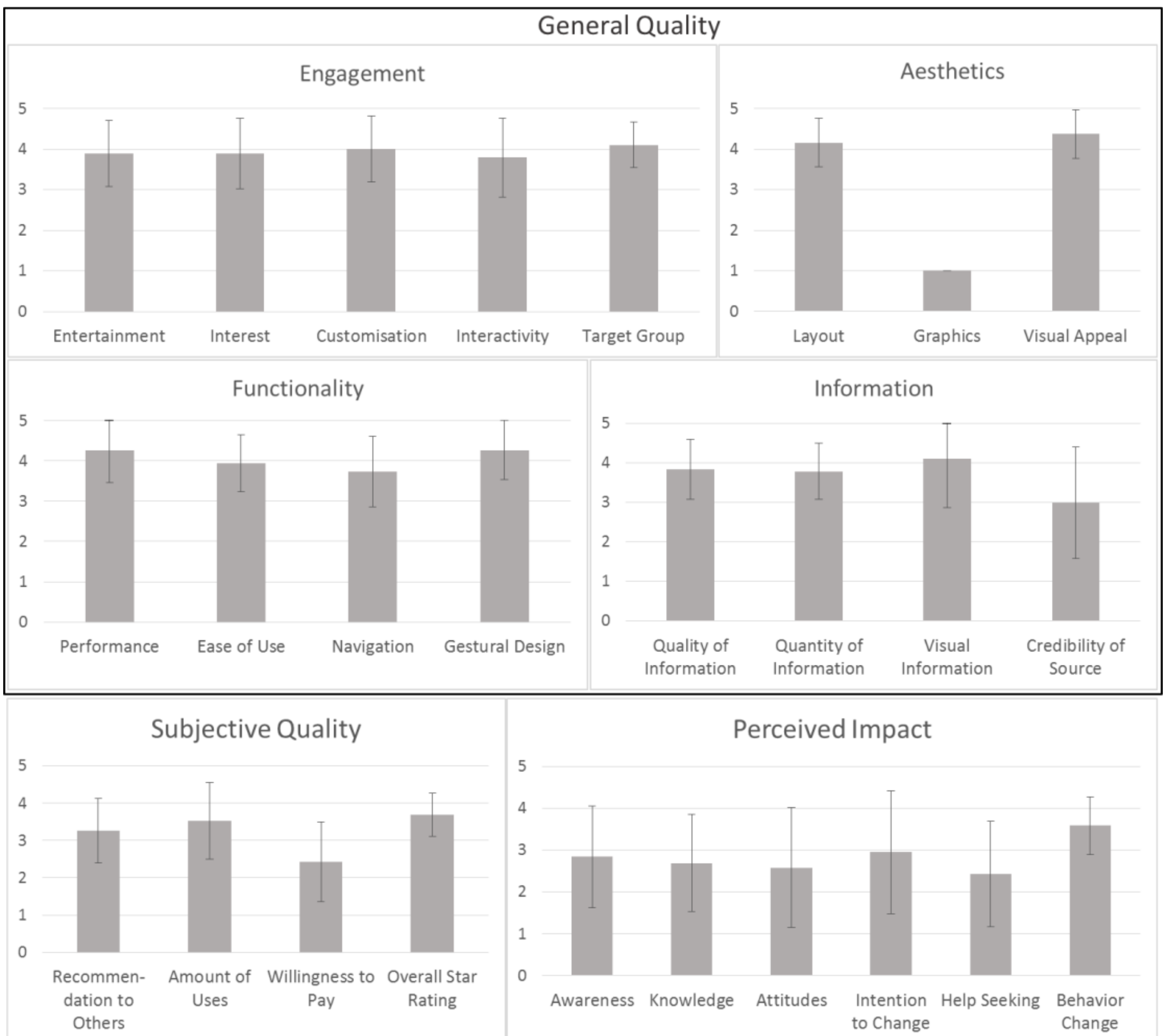

Figure 13: Means and standard deviations of the uMARS ratings for the MUSE companion app.

The wearable EEG's comfort was evaluated positively with regard to emotion (*M* = 1.68, *SD* = 0.93), harm (*M* = 1.45, *SD* = 0.78), movement (*M* = 2.23, *SD* = 1.61) and anxiety (*M* = 1.21, *SD* = 0.42). Perceived change (*M* = 3.05, *SD* = 1.38) received a neutral and attachment (*M* = 3.82, *SD* = 1.38) a negative rating.

The additional questions revealed that the participants found the wearable easy to use (*M* = 3.84, *SD* = 0.98) and visually appealing (*M* = 3.84, *SD* = 0.60). Though the rating with regard to using MUSE for meditation in the future was neutral (*M* = 3.00, *SD* = 1.11), they were unlikely to use it in daily life (*M* = 2.37, *SD* = 1.30). If they would use it, it would most likely be in private (*M* = 4.26, *SD* = 0.73), while they would avoid using the device in public (*M* = 1.63, *SD* = 0.83).

2.3.2.3.3 Discussion

The results showed a positive evaluation of the systems usability and user experience. This is slightly better than the rating of Leape et al. (2016), who evaluated the systems ease of use to be medium. However, their ratings were focused solely on the wearable, while the ratings of this study involved both MUSE and its companion app. In addition to its generally positive rating, the UEQ scores indicated room for improvement regarding MUSE's stimulation and dependability. The participants liked the motivational aspects, such as points and badges that were included in the app. As suggested by Christmann et al. (2018b), they even wanted more motivational aspects to be included to keep them interested long-term. In line with Rebolledo-Mendez et al. (2009), the results of this study show that assessing the usability and user experience of wearable EEGs is an effective approach to identify domains that require improvement.

Though the MUSE app's general quality was rated to be good, the app's graphics were assessed as poor. The participants thought that the graphs and graphics that were used to display the user's performance at the end of each exercise held little meaning and were hard to understand. App developers should visualize the data according to its application purpose and do it in a way that makes sense to the user, e.g. by educating and helping the user to interpret the data (Mihajlovic et al., 2015). A possible solution to this was implemented for Stress-Mentor's diary overview diagram, where in addition to displaying the user's data, the graph provides color coded feedback in order to simplify its interpretation (Christmann et al., 2017b).

In agreement with a previous assessment (Leape et al., 2016), the wearable EEG's comfort was also evaluated positively. Nonetheless, the participants had concerns regarding MUSE's comfort during longer sessions. This agrees with the comments from study 1, where the participants had to wear the device for a longer duration. Here, two participants complained about MUSE being uncomfortable. Furthermore, two participants from study 2 complained that the EEG was uncomfortable in combination with their glasses. These aspects are important considerations, because comfort is one factor that determines the system's acceptance (Hairston and Lawhern, 2015). Developers should make sure that wearable EEGs are comfortable to wear independent of the user's head shape and usage duration.

With regard to operating and navigating the system, the participants showed particular difficulty in connecting the wearable EEG with the companion app and establishing a good data signal. They felt that they could only learn how to operate the device correctly through trial and error. This coincides with a previous study that reported difficulties in the electrode setup of the MUSE EEG (Krigolson et al., 2017). This contradicts the fact that such systems have to be easy to set up and troubleshoot in order to ensure their usability and acceptance (McDowell et al., 2013). When integrating wearables into health apps, the developers should therefore make sure that connecting the device to the app is easily done, e.g. by providing a step by step description of the procedure (Mihajlovic et al., 2015). Establishing a good data signal could be improved through automatically adjusting the instructions to the user's current situation. Such intelligent EEG solutions could help the user to obtain a suitable data quality by guiding the user in readjusting the electrodes (Mihajlovic et al., 2015).

Though the wearable's visual appeal was rated as good, the participants were unwilling to wear the device in public, probably because it is very noticeable to observers (Leape et al., 2016). This coincides with the previous conclusion that wearable EEGs need to be transparent or even invisible (McDowell et al., 2013), as well as applicable in everyday life (De Vos et al., 2014) if developers want to ensure the user's compliance. It also contradicts previous predictions that the use of such sensors might become more acceptable (Leape et al., 2016).

In addition, data privacy is an important decision-making factor when it comes to downloading health apps (Meingast et al., 2006). The participants of this study noted that they had to disclose a lot of information that would keep them from downloading and using the app on their own smartphones and with their own data. It was also mentioned that the login process should not be dependent on the internet and that the data should only be saved locally instead of online. When utilizing commercially available wearables such as MUSE, it should be kept in mind that the users' data can potentially be intercepted or exposed to third parties (Alahäivälä and Oinas-Kukkonen, 2016).

Out of the 19 study participants, only two stated that they would be willing to buy MUSE for the retail price. One of them further restricted his answer by saying that he would only buy it to access the devices raw data and use this to build his own application, but not to use it with the MUSE app. This coincides with the assessment of the other 17 participants who said that the device was too expensive, especially considering its limited functionality and applicability. Seven participants would be willing to buy the MUSE wearable if it were cheaper. On average, participants were willing to pay approximately 80 € (range 30-200 €) for the device. This shows that the previous assumption that such devices must be low cost to ensure their usage held true (Hairston and Lawhern, 2015). It also demonstrates that, even though MUSE is one of the cheapest wearable EEG systems currently available, potential users are unlikely to acquire the device in order to perform neurofeedback-based relaxation exercises. Though the wearable EEG MindWave Mobile 2 (NeuroSky), which provides single electrode brain monitoring, is the cheapest one available at 99.99 $, its price is still above what the participants in this study were willing to pay. Moreover, such low-price solutions usually come at a trade-off for signal quality and robustness (Mihajlovic et al., 2015).

The results for the SAM show that the participants' arousal decreased after the second meditation, compared to the baseline measure and the rating after the connection of the EEG system with the app was completed. This suggests that neurofeedback-based meditations can have positive impact on perceived arousal (Hjelm and Browall, 2000; van Boxtel et al., 2012). Nonetheless, due to a missing control group, it remains questionable whether it is really the neurofeedback that causes a decrease in reported arousal. In fact, the participants said that they found the received neurofeedback to be disturbing. They stated that the audio-feedback was distracting them from the original task (meditation) and that it was hard for them to relax during the exercise. One even said she found the feedback to be stressful because she felt pressured into doing the exercise right. This coincides with a recent study on a biofeedback app to support deep and slow abdominal breathing. Here biofeedback instructions also did not result in immediate effects on the subjective relaxation level (Faust-Christmann et al., 2019).

In addition to this, the participants thought that MUSE's online-feedback was incorrect. Very high accuracies are required to ensure the users' acceptance of the device (Wilson and Russell, 2003). Besides, it was unclear how the app evaluated the user's data. If the content of a health app is perceived to be incorrect or irritating this likely causes distrust and discourages further usage (Dennison et al., 2013). This could have resulted in the relatively bad scores with respect to the credibility of the app's source, subjective quality, and perceived impact. The results therefore, indicate that apps that include techniques with a proven effect on well-being (e.g., the mindfulness meditation app "Headspace") might be preferable (Bostock et al., 2019).

Reinforcement for this came from the users' statements which indicated that the participants prefer guided relaxation exercises over ones with online feedback and that a choice of different exercises would be favored, so they could choose the ones they liked best. This makes sense because not all techniques work equally well for all users (Murphy, 1996; Koldijk et al., 2016). This challenge is often met by providing multiple techniques within one intervention (Murphy, 1996; Jones and Johnston, 2000).

2.3.2.4 Conclusion

While previous studies have investigated the validity of other wearable and consumer EEG systems (e.g., Matthews et al., 2008; Zander and Kothe, 2011) and one study focused specifically on MUSE's validity in the application of ERP measurements (Krigolson et al., 2017), this is the first study that looked into the suitability of MUSE with regard to frequency aspects. Study 1 therefore adds to the existing literature comparing wearable EEG systems and classical EEG setups.

The mixed results of this study indicate that MUSE's reliability and validity depends largely on electrode location, task, and the investigated frequency domains. This proves that there is still reason to question the accuracy of dry electrode and consumer EEG systems (Cutrell and Tan, 2008). Hence, the MUSE's applicability for neurofeedback-based relaxation exercises and other contexts is problematic.

This was further supported by the participant's perception that the feedback received with the MUSE app was incorrect. In addition, they thought the provided feedback was distracting them from meditating and felt that they could not relax during the exercise. In line with Mihajlovic et al. (2015), this indicates that like other wearable EEGs, MUSE requires improvement with regard to signal acquisition, feedback generation, and personalization.

Furthermore, potential users do not want to acquire the wearable device because it is limited to one application context. However, in order to reach as many users as possible, the technologies that are utilized for mHealth purposes should be accessible to people from a multitude of backgrounds (Alahäivälä and Oinas-Kukkonen, 2016). In view of these findings, the focus was put on integrating text- and audio-guided relaxation exercises into Stress-Mentor, instead of ones that are based on online feedback.

2.3.3 Testing the Applicability of Audio- and Text-Based Relaxation Exercises in a User Study

2.3.3.1 Introduction

In addition to the neurofeedback exercise provided by the MUSE companion app, four other relaxation exercises (guided imagery, progressive muscle relaxation, mindfulness meditation, and abdominal breathing) from Mason's (2001) "Guide to Stress Reduction" were translated into German and tested in a user study. The first exercise, guided imagery, describes a procedure during which an individual's personalized images are mentally produced (Weigensberg et al., 2009). Guided imagery often includes sensations such as smell, taste, sound, sight, and feel (Naparstek, 1995). Another proven stress management technique is PMR. PMR consists of the progressive tensing and relaxing of muscle groups. It is aimed at learning the sensations associated with tension, as well as recognizing the connection between tension and relaxation, while at the same time relaxing the complete skeletal musculature (Scheufele, 1999). Mindfulness meditation, on the other hand, trains a person to pay attention to internal and external experiences (e.g., sounds, heartbeat, and breathing) (Klatt et al., 2009), while breathing exercises are based on the manipulation of breath movements. For example, during abdominal breathing, a person's abdomen, rather than his or her chest, expands during inhalation (e.g., Varvogli and Darviri, 2011). In addition to these four techniques, listening to music or sounds has proven effective for reducing stress (Labbé et al., 2007). However, not all types of music promote relaxation. Classic and new age music for example seem to be better suited than e.g., heavy metal (Mornhinweg, 1992).

Each of the exercises was aimed at a beginner's level and included detailed instructions that the participants were asked to follow. To make following the provided instructions easier, the longer exercises (i.e., mindfulness meditation, guided imagery, and PMR) were recorded in an audio format. Because the instruction for the breathing exercise was short, it was presented as text. All four exercises were backed by new age ambient music. The same music was then used as an additional relaxation exercise, resulting in five experimental conditions: meditation, guided imagery, PMR, breathing, and music. In order to ensure that the resulting exercises were feasible to be integrated into Stress-Mentor, their applicability was tested in a user study.

2.3.3.2 Methods

The five exercises were tested in a user study to examine their practical suitability. For this purpose, N = 30 participants were recruited via the email distribution system of the Technische Universität Kaiserslautern. All participants gave written consent in accordance with the declaration of Helsinki.

The study was based on a within-subjects design. Each participant experienced all five stress management exercises, namely mindfulness meditation, guided imagery, PMR, breathing, and music over three sessions. Because mindfulness meditation, guided imagery and breathing were comparably short (each lasted approximately 20 minutes), they were presented in one session. The order of presentation for these three exercises was cross-balanced across participants. The PMR and music condition lasted 60 minutes each. This resulted in a total of three 60 minutes long sessions for every participant. The presentation order of the sessions was, again, cross-balanced across participants. After performing each exercise (mindfulness meditation, guided imagery, PMR, breathing, and music), the participant filled out a questionnaire regarding his or her perception of the task. In this questionnaire he or she was presented with different statements and was instructed to rate how strongly he or she agreed with each statement on a five-point Likert scale (from 1 “strongly disagree” to 5 “strongly agree”). See Appendix 15 for details regarding the presented statements. In addition, a final questionnaire was filled out at the end of the study (Appendix 16).

The Shapiro-Wilk test indicated that the data was not normally distributed (all $p > .05$). As a result, non-parametric tests (namely: Friedman test instead of ANOVA with repeated measures, and Wilcoxon signed-rank test for two related samples, Bortz, 2016) were used for statistical analysis. The Bonferroni correction method was used to control for effects of multiple testing. The effect size (r) according to Cohen (1992) was calculated. Effects r = 0. 10 – 0.30 correspond to a small effect, r = 0.30 - 0.50 a moderate effect and $r \geq 0.50$ a large effect.

2.3.3.3 Results

The Friedman tests showed no differences between the performed tasks with regard to the perceived ease of understanding the instructions, how much the participants enjoyed the exercise, the perceived quality of the audio recording, and as to how pleasant they perceived the background music. However, the tasks differed with regard to the perceived ease of executing the exercise ($\chi^2(4) = 15.71$, $p = .003$). Wilcoxon signed-rank tests confirmed that the instructions for the music condition were rated as easier to follow than those for guided imagery ($Z = -3.04$, $p = .002$, $r = 0.56$), mindfulness meditation ($Z = -2.98$, $p = .003$, $r = 0.54$), and PMR ($Z = -3.21$, $p = .001$, $r = 0.59$). Though no significant difference was found between the music exercise and the breathing exercise ($\alpha = .007$ due to Bonferroni correction), a general trend revealed that the execution of the music exercise was perceived as easier than that of the other conditions.

These results were further confirmed by the post-test questionnaire. The exercises did not differ in their final rating of pleasantness ($\chi^2(4) = 3.08$, $p = .54$). The Friedman test revealed differences between the five relaxation techniques regarding the final ratings of the perceived ease of performing the exercises ($\chi^2(4) = 11.05$, $p = .02$). However, Wilcoxon singed-rank tests did not confirm this result (all $p > .05$).

2.3.3.4 Discussion

Overall, the results revealed a positive evaluation of all presented stress management exercises. The instructions were perceived as easy to follow and the quality of the audio recordings was rated to be good. Moreover, participants generally enjoyed the exercises and rated them to be easy to execute. The speaking rate was perceived neither as too fast, nor as too slow. The speaker's voice was rated as pleasant and very comprehensible. However, two participants stated that they would have preferred a female instead of a male voice for the audio instructions. This indicates that this aspect might depend on individual preference. App designers should, therefore, consider providing the user the choice between different speakers (e.g., a male and a female), when implementing audio instructions, as is often done in navigation devices (e.g., TomTom).

Though the background music was perceived as pleasant, the study also revealed that 5/30 participants commented negatively on the used music. For example, one said that he perceived the music as boring, another described it as monotonous. In response to this feedback, a slider was implemented that allows the user to adjust the volume of the background music or turn it off completely, if he or she wishes (Figure 14).

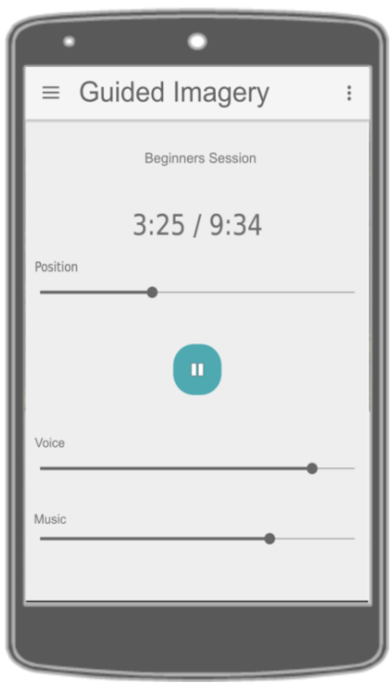

Figure 14: Screenshot of an audio exercise.

Moreover, instead of providing a preset music exercise, a task was added that allows the user to choose his or her own music. This exercise includes information on which music is generally better or worse suited for relaxation (Mornhinweg, 1992). Despite these adjustments, the results spoke for a positive evaluation of the presented stress management exercises and indicated that they could be implemented into the app.

2.3.4 Description of Included Stress Management Methods

Though biofeedback exercises have successfully been applied to reduce the stress levels of individuals (e.g., Hjelm and Browall, 2000), the presented studies with the wearable EEG MUSE showed that the application of such exercises might not be wanted by users of stress management apps. In fact, it was revealed that users preferred stress management exercises that were based on audio- or text instructions.

This coincides with the fact that the realized audio- and text-based relaxation exercises were well received by the users. As a result, the tested relaxation exercises for beginners, as well as the advanced exercises that are based on them, were included in Stress-Mentor. This agreed with the users' wish for graded tasks. As was the case in the study, longer and more complex exercises (e.g., PMR, meditation, and guided imagery) were presented in audio format. The instructions for shorter exercises with simpler instructions (e.g., breathing) were presented as text. In addition to PMR, meditation, guided imagery, and breathing, several other emotion-focused coping exercises were implemented. This included tasks for self-massage, mindfulness meditation, euthymic methods, and physical stress relief.

In addition to relaxation exercises, it has been recommended that stress management interventions should also include psycho-education, the analyses of individual reactions, assertiveness training, time management, and cognitive restructuring (Jones and Johnston, 2000). Thus, a variety of problem-focused coping strategies were also included in text format, namely refuting irrational ideas, time management, setting priorities, and assertiveness training. Because these exercises are aimed at altering an individual's interaction with his or her environment, they pose a good supplement to the emotion-focused techniques which are more strongly targeted at furthering relaxation (Lazarus, 1985). Since these exercises are often more complex, take longer to learn, and often involve planning in order to be executed, they were not tested before implementation.

Table 4 provides an overview of the included stress management exercise categories. All tasks were based on exercises from an established self-help book by Mason (2001). To support the user in finding additional help, information regarding important professional contacts and links to websites, as well as books with further information about stress and stress management, were included in several exercises. As a result, Stress-Mentor now includes 79 stress management tasks from 14 technique types. This multi-technique approach allows the user to find out for him- or herself which stress management techniques are best suited for him or her (Murphy, 1996; Koldijk et al., 2016).

Table 4: Stress management exercises that are included in Stress-Mentor.

Type of Stress Management Technique	Category	Classification
Avoiding perfectionism	Cognitive	Problem-focused
Refuting irrational ideas	Cognitive	Problem-focused
Assertiveness training	Cognitive	Problem-focused
Social support/change	Social support	Problem-focused
Set and identify priorities	Time management	Problem-focused
Observing and planning time division	Time management	Problem-focused
Euthymic methods	Relaxation	Emotion-focused
Self-massage	Relaxation	Emotion-focused
Breathing	Relaxation	Emotion-focused
Passive and active PMR	Relaxation	Emotion-focused
Guided imagery	Relaxation	Emotion-focused
Meditation	Relaxation	Emotion-focused
Mindfulness meditation	Relaxation	Emotion-focused
Physical stress relief	Relaxation	Emotion-focused

2.4 Linking Gamification and Behavior Change Techniques in "Stress-Mentor"

2.4.1 Background

Based on the feedback from the previous studies, an extended app concept for Stress-Mentor was developed. This concept links established stress management methods and behavior change techniques with an extensive gamification framework (Christmann et al., 2018a). Such approaches have already shown great potential in ensuring the effectiveness of health apps (e.g., Edwards et al., 2018).

All in all, Stress-Mentor includes 8 out of 17 gamification techniques that were presented in a taxonomy by Hoffmann et al. (2017). All techniques that could potentially lead to increased pressure on the user were not integrated because they were perceived as unsuitable for the context of stress management (Hoffmann et al., 2017). Moreover, real world prizes and 3-D environments were not feasible for the intended application of the app, though it is conceivable that these solutions could be implemented in the future (e.g., a virtual reality based stress management exercise or a combination of Stress-Mentor with a loyalty program of health insurance companies).

Appendix 17 provides an overview on which gamification techniques are included in Stress-Mentor and how they are realized. These gamification aspects are linked to a large variety of evidence-based behavior change techniques that are based on a taxonomy by Abraham and Michie (2008) (see Appendix 18 for details on which techniques were integrated and how they were realized). As recommended by Koldijk et al. (2016), the integrated behavior change techniques were chosen based on their suitability for their application in the context of stress management. In the following, a detailed description of Stress-Mentor and its features is provided, including how gamification was linked with the behavior change techniques (also see Table 5 for an overview). The behavior change techniques are italicized in the text. Stress-Mentor's most distinctive characteristics are limited usage duration by design to support the user's autonomy and a number of tiered gamification methods to keep the user interested (Figure 15).

Table 5: Detailed description of the gamification techniques that are included in Stress-Mentor.

Gamification Technique	Behavior Change Technique
Avatar	Provide feedback on performance, model or demonstrate the behavior
Agent	Agree on behavioral contract, prompt practice, provide instructions on app usage, provide information about the behavior health link and behavioral consequences, provide general encouragement, use follow up prompts
Health diary & diary overview diagram	Prompt self-monitoring of behavior, provide feedback on performance, provide information about behavior health link
Photo book	Provide opportunities for social comparison, provide feedback on performance
Shop	Provide feedback on performance
Experience points	Provide contingent rewards
Progress bars	Provide feedback on performance
Badges	Provide contingent rewards, provide feedback on performance
Tasks	Stress management (e.g., breathing exercises, meditation, PMR, refuting irrational ideas), plan social support and social change, time management, barrier identification, model or demonstrate the behavior, prompt practice, prompt specific goal setting, prompt review of behavioral goals, provide information about health behavior and behavioral consequences, teach to use prompts or cues, prompt self-talk, provide general encouragement, prompt intention formation, provide instructions, set graded tasks

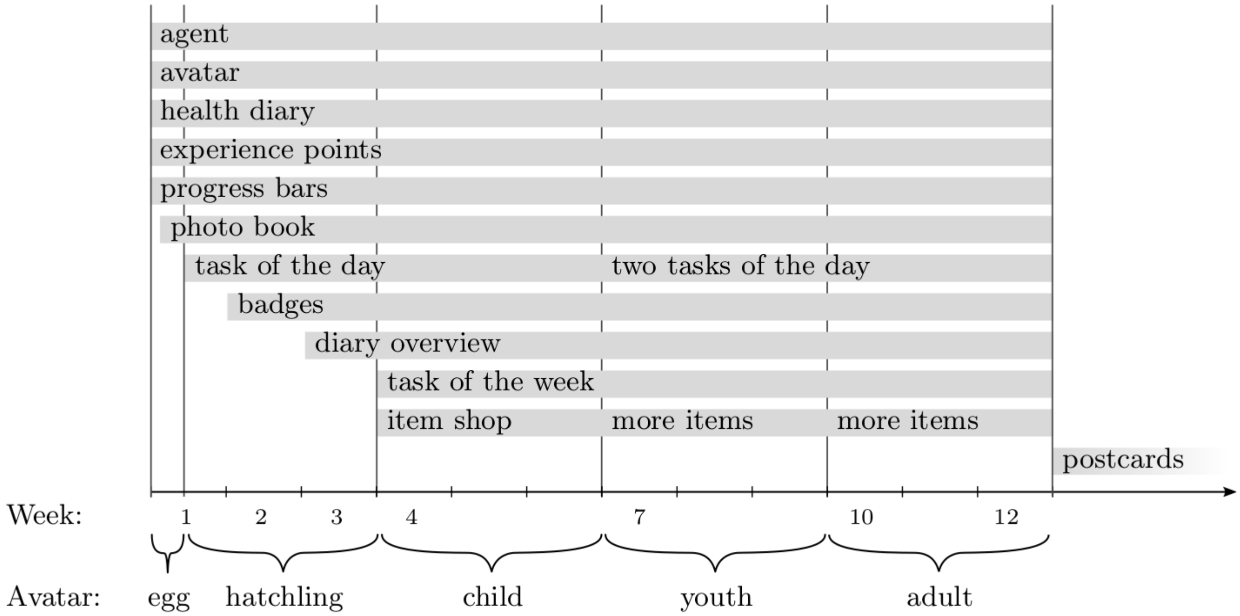

Figure 15: App features timeline.

2.4.2 "Stress-Mentor's" Concept

2.4.2.1 Diary

Based on the MVP, the final version of Stress-Mentor includes a health diary, with which the user can track a variety of stress-relevant behaviors, as well as aspects regarding his or her well-being in the past 24 hours. Previous studies show that such digital health diaries improve the user's compliance compared to paper and pencil approaches (Taylor et al., 2019). In accordance with the user feedback regarding the MVP and feedback from medical experts, who suggested adding the categories sleep quality, exercise intensity, digestion and step count, the diary was extended.

As a result, it now encompasses the following 14 categories: sleep duration (Alvarez and Ayas, 2004; Gallicchio and Kalesan, 2009), sleep quality (Pilcher et al., 1997), physical exercise duration (World Health Organization, 2011), physical exercise intensity (Shephard, 2001), positive events (Cohen and Hoberman, 1983), negative events (Kanner et al., 1981; DeLongis et al., 1982; Lu, 1991), subjective stress-level (Levenstein et al., 1993; Cohen et al., 1994), prevailing mood (DeLongis et al., 1988), caffeine consumption (Nawrot et al., 2003; European Food Information Council, 2018), fruit/vegetable consumption (Boeing et al., 2012; Oyebode et al., 2014), alcohol consumption (German Nutrition Society, 2010), consumption of water/unsweetened beverages (German Nutrition Society, 2017), digestion, and step count (Tudor-Locke and Bassett, 2004). The user is provided with direct feedback through a traffic light color coding (green, light green, yellow, orange, red), on whether the tracked behaviors meet general health recommendations. See Table 6 for details. This way the diary supports *self-monitoring* and *provides feedback on the user's performance* which enables the user to easily appraise his or her behavior. Another form of direct feedback is provided through an overview diagram (Christmann et al., 2017b) that summarizes the user's diary entries over time within one chart. Such graphs have been associated with increased task meaningfulness as well as need satisfaction (Sailer et al., 2017).

Table 6: Diary categories of Stress-Mentor's full version with entry evaluation. Sleep duration and alcohol are displayed in a simplified manner as the health recommendations depend on the user's age and gender respectively.

Diary Categories	Green Evaluation	Light Green Evaluation	Yellow Evaluation	Orange Evaluation	Red Evaluation
Sleep duration in hours	8	7, 9	6, 10	5, 11	≤ 5, ≥ 11
Sleep quality	good	-	medium	-	bad
Sport duration in minutes	≥ 30	-	10-20	-	0
Sport intensity	medium	easy-medium, medium-hard	easy, hard	very easy, very hard	-
Step count	≥ 12500	<12500	< 10000	< 7500	< 5000
Consumption of water/unsweetened beverages in liter	1.75, ≥2	1.5	1,1.25	0.75	0-0.5
Caffeine intake in 100mg units	0-2	3	4-5	6	≥ 7
Portions of fruits and vegetables	6, ≥7	5	3-4	2	0,1
Alcohol intake in glasses	0-1	-	1.5	-	2-≥4
Digestion	good	-	medium	-	bad
Emotional state	happy	content	neutral	-	sad, angry, worried
Stress level	0-1	2-3	4-5	7-8	9-10
Positive events	many	some	few	-	none
Negative events	none	few	some	-	many

Additionally, when the user presses the info button next to each category, he or she is presented with information regarding the health recommendations for the respective behaviors. Figure 16 displays screenshots of the diary and diary overview diagram.

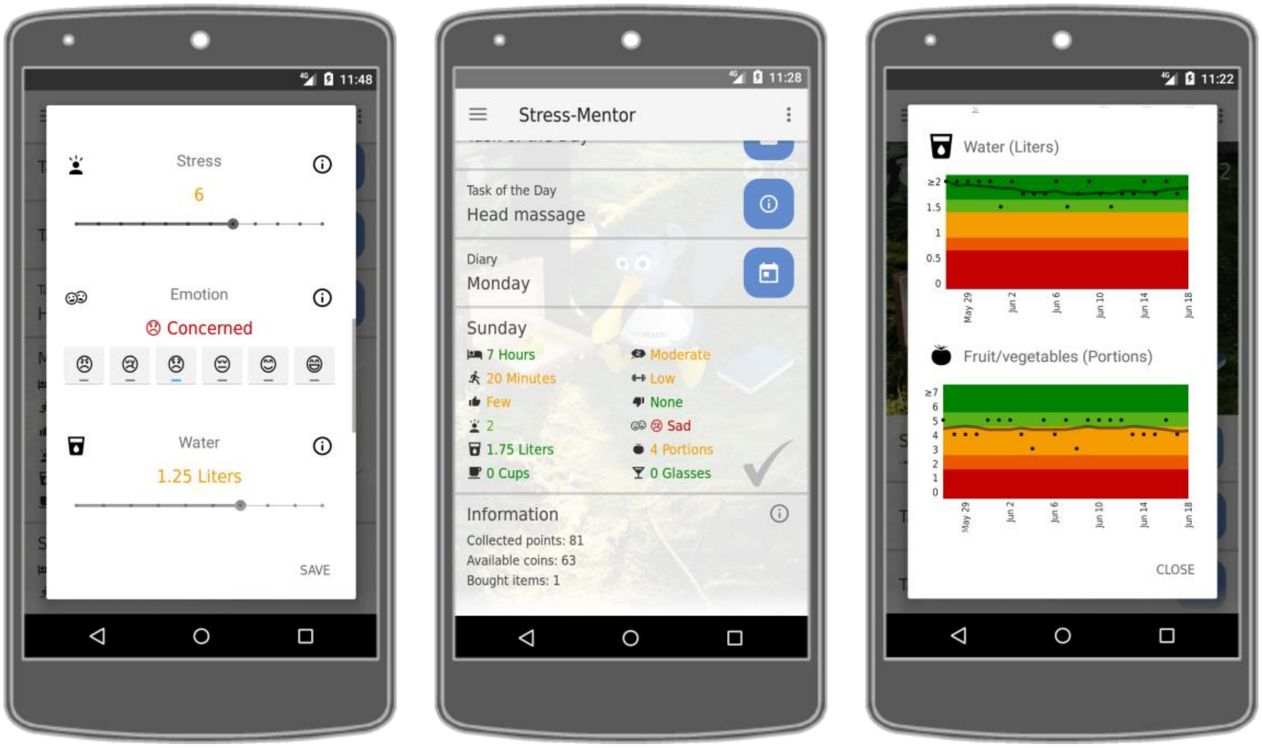

Figure 16: Left: screenshot of the diary. Middle: the app's main screen with a summary of the color-coded diary entries. Right: screenshot of the diary overview diagram.

2.4.2.2 Avatar

As was the case for the MVP, the avatar's (a bird-like cartoon animal, namely the Rhineland-Palatinate Elwetritsch) appearance is linked to the user's diary entries. Such approaches for vicarious reinforcement can affect the user's behavior by observing and interpreting the rewards and punishments experienced by others (e.g., Fox et al., 2009; Byrne et al., 2012; Christmann et al., 2018b). For this purpose, each diary category was linked to one feature of the avatar's appearance (e.g., less stress leads to a fuller plumage, see Table 1 for details). This *provides information about the consequences* of the user's behavior.

The avatar's overall condition mirrors to what extent the health recommendations in the diary were followed during the past week. This *provides contingent rewards*. This approach has been shown to support continuous usage behavior over a four-week interval, compared to a control condition with a static avatar (see Christmann et al., 2018b for details). These positive effects are reflected in an increase in user compliance when digital diaries are paired with gamification (Taylor et al., 2019). Moreover, the avatar's size mirrors the progress within the app, which in turn *provides feedback on the user's performance*. Positive feedback is a common gamification strategy that aims at reinforcing the desired behaviors (Alahäivälä and Oinas-Kukkonen, 2016). Figure 17 depicts both the avatar's adjustment according to the user's behavior and its development over time.

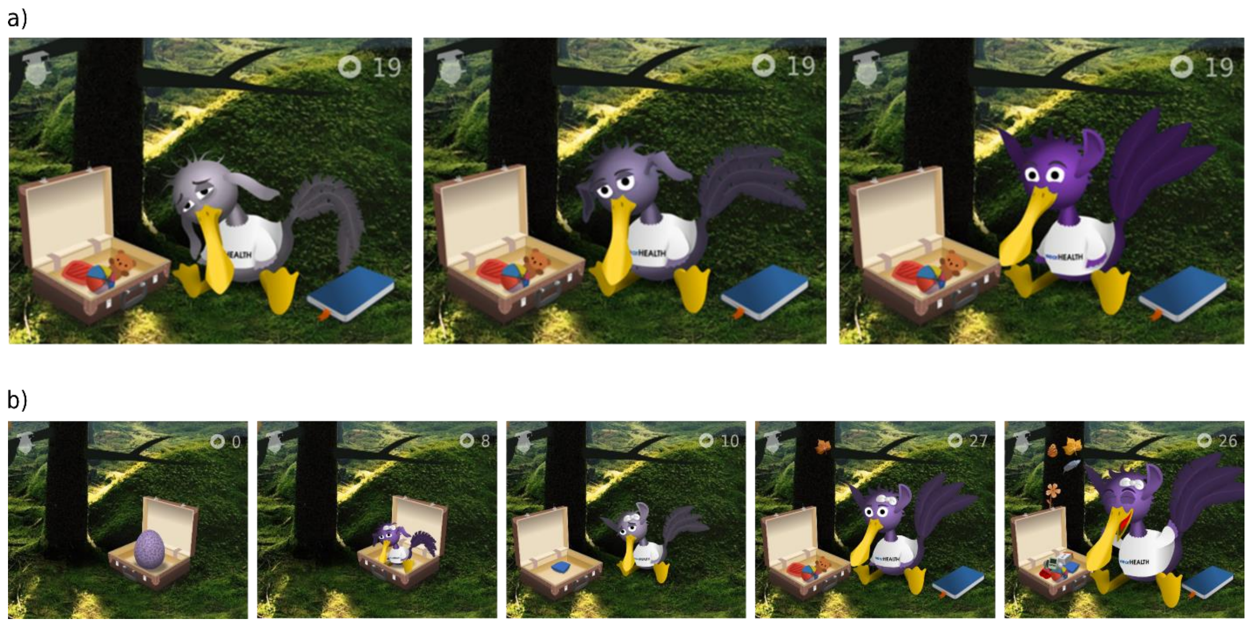

Figure 17: a): Changes in the avatar's looks dependent on the diary entries (left: red entries, middle: yellow entries, and right: green entries). b): Avatar's development over time, dependent on the user's progress (left to right: egg, infant, child, teenager, and adult).

To track the avatar's development and to counteract change blindness, the user is given the opportunity to make photos of the avatar, save them in a photo book (Figure 18) and to share these photos with friends (*provide opportunities for social comparison*). This further promotes social relatedness through the avatar (Sailer et al., 2017).

2.4.2.3 Agent

Another of Stress-Mentor's gamification features is the app's agent: a wise owl which serves as the user's mentor. Using agents is especially effective to deliver health related communication and health behavior change interventions (Bickmore et al., 2005) because they can have positive impact on the user's learning (Holmes, 2007). The owl instructs the app usage and introduces new app features (provide instruction). It also provides general information about stress management and about the consequences of behavior change in the form of short tips (Figure 19). These tips also target social support, time management, and the identification of barriers (e.g., "It is important to identify barriers in our daily life which prevent us from reaching our goals. When you are aware of them you can include them in your plans”). It also shows encouraging quotations from prominent figures (e.g., “Success is falling nine times and getting up ten” by Jon Bon Jovi) to *provide general encouragement*.

Figure 18: Screenshot of the photobook to track the avatar’s development.

Moreover, at the very first start of the app, the agent entrusts the care of the avatar to the user though a *behavioral contract*. In order to raise the avatar and prepare it for its future life, the user has to fulfill the tasks given in the application. The owl’s daily reminders to complete the tasks and to fill out the stress diary help to *prompt daily practice* and to *prompt the self-monitoring of behaviors*.

2.4.2.4 Task of Day/Week

In order to progress in the app and raise the avatar, the user has to complete a certain number of tasks which teach evidence-based stress management strategies. In line with multi-technique approaches (e.g., Mason, 2001; Davis et al., 2008), Stress-Mentor includes a large variety of emotion- and problem-focused stress management strategies. Furthermore, different behavior change techniques are realized in the tasks. Appendix 18 provides an overview on the behavior change techniques and their implementation within the app.

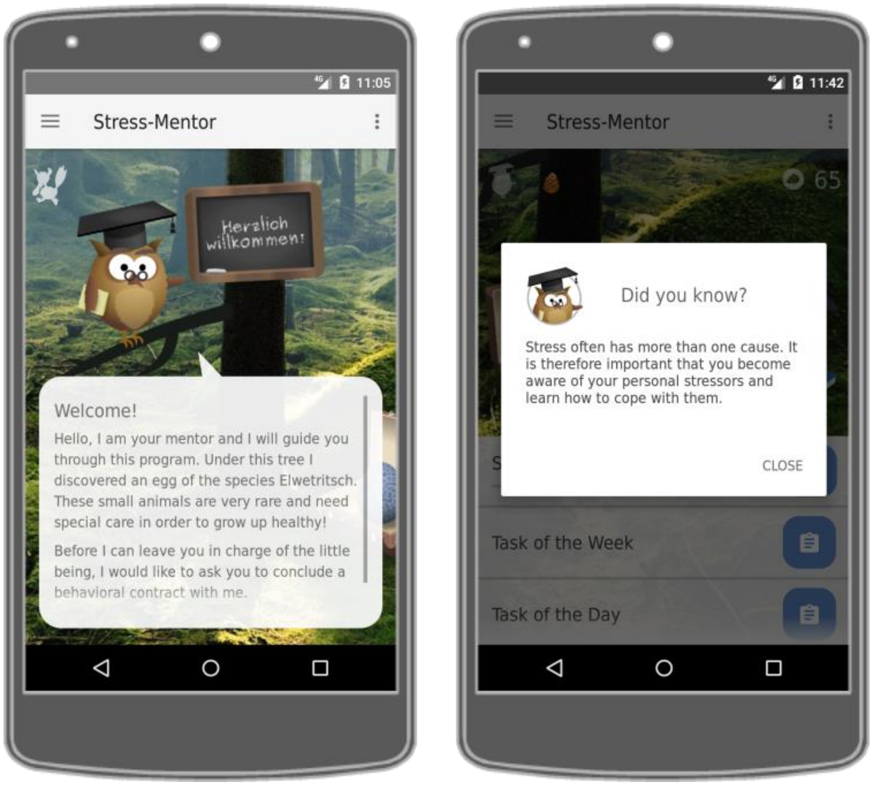

Figure 19: Left: the "wise owl" provides instructions and information about the app's usage. Right: the "wise owl" provides general information about stress management.

At the beginning the user can perform one "task of the day". As the user progresses, a second "task of the day" and tasks spanning up to seven days ("task of the week") are added to support further practice. The integration of the stress management exercises through a task feature supports the user in *setting attainable goals*. For each task, the user can choose between three stress management exercises.

Because it has been shown that not all coping techniques work equally well for every individual (Murphy, 1996; Koldijk et al., 2016), the order in which the exercises are presented to the user is based on a weekly stress checklist. In the stress checklist, the user can enter stress-related mental and physical complaints (i.e., fears and worries, sadness, anger, stress at work, stress in private life, muscle tension, head, neck, and back grievances caused by tension, digestive problems, and sleep problems). Each aspect is rated on a scale from 0 (none) to 10 (maximum). This approach was chosen because certain stress management exercises are especially effective for reducing specific stress-related complaints (Davis et al., 2008). As is the case for the diary, the entries in the stress checklist are color coded in accordance with an extended traffic light coloring system (0-1 = green, 2-3 = light green, 4-6 = yellow, 7-8 = orange, 9-10 = red). Tailoring the content to the user's personal demands in this manner (Atienza and Patrick, 2011) improves the adoption of self-management strategies. In addition to the task features, all exercises the user has already unlocked by executing them can be reviewed in the app's menu and repeated whenever and as often as the user wants.

2.4.2.5 Visualization of Progress

As recommended, the user's experience is reflected through a progressive, nonlinear point and leveling system that includes badges in order to document benchmarks in the user's progress (Zichermann and Cunningham, 2011). Each finished task rewards the user with experience points and virtual currency, which can later be exchanged for virtual items in a shop. These items are thematically fitting to the topic of preparing the avatar for its future life, such as a soft blanket, a calendar, or books. They support the *visualization of the user's progress* by being placed in a suitcase next to the avatar (see Figure 20). Besides the shop, progress bars help to visualize the user's advancement. Progress bars are commonly used to convey to the user how many points he or she has attained and how many more points are needed in order to achieve the next level. This motivates the app's continued use (Zichermann and Cunningham, 2011). Another *visualization of progress* is provided through *contingent rewards* in the form of badges. Badges can, for example, be earned through consistently keeping the diary and following the health recommendations (Figure 20). This supports both need satisfaction and perceived task meaningfulness (Sailer et al., 2017).

2.4.2.6 Additional Features

In addition, the aspect of discovery is realized through several methods that are distributed throughout the app (e.g., after reaching a new level the avatar grows, discounts on items in the shop, opportunity for taking photos is not fully predictable). Generating unpredictability in this manner is expected to encourage the user to return (Chou, 2015). After finishing the designated usage period, the avatar leaves, symbolizing freedom and autonomy. The user receives postcards from the avatar on a regular basis to provide follow-up prompts.

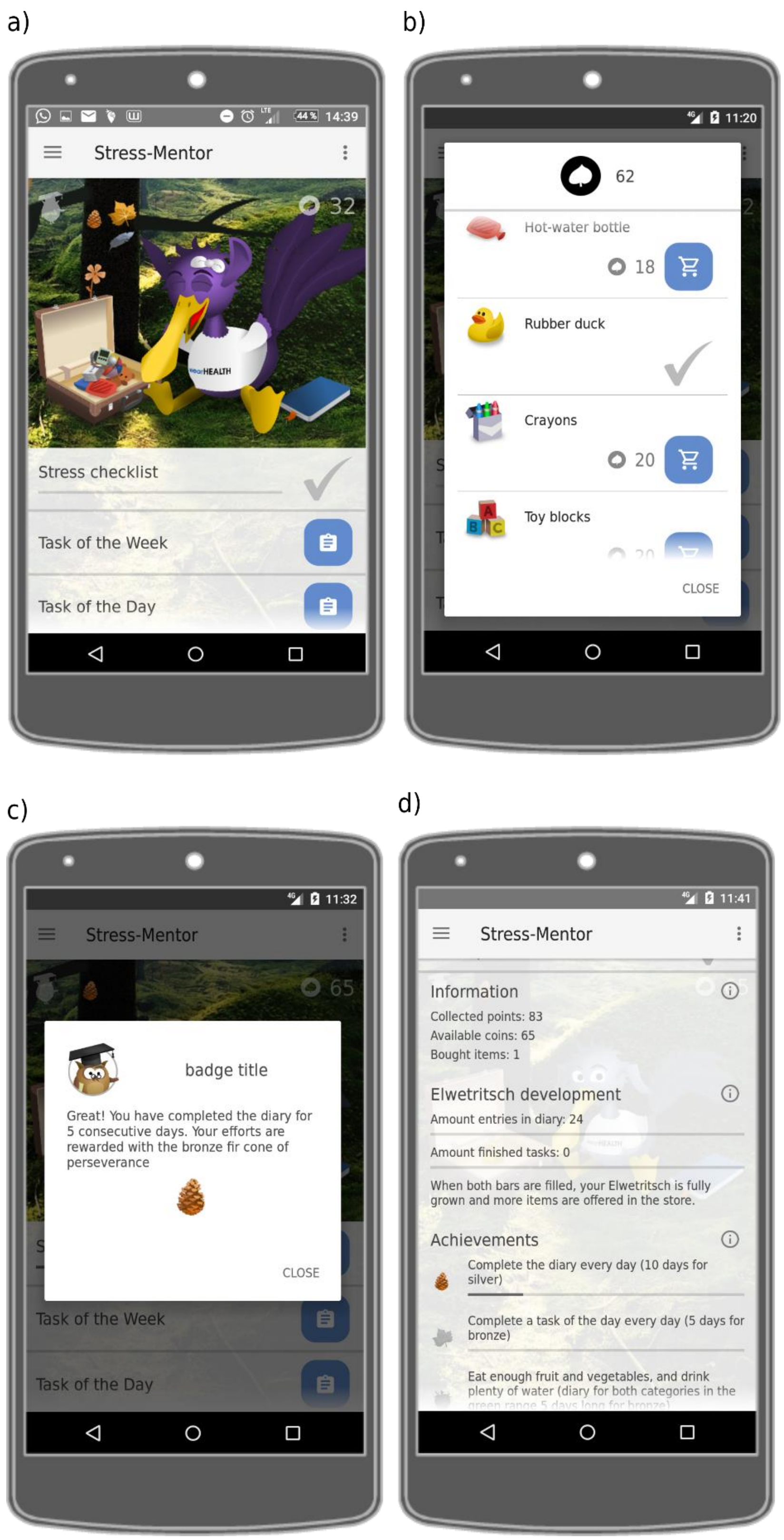

Figure 20: a) Screenshot of Stress-Mentor that shows the avatar and the suitcase which is filled with virtual items. The badges are placed on the tree trunk. The experience points are displayed in the upper right corner. b) Screenshot of the shop. c) The "wise owl" hands a new badge to the user, who has completed the tasks of the day on a regular basis. d) Progress bars depict the user's progress with regard to the next level and the badges.

Though Stress-Mentor has a limited usage period by design (if the user extensively uses the app on a daily basis, he or she is finished in three months), the app does not instill the pressure to be used. If the user does not manage to finish all available features, or if he or she forgets to use the app for a while, this does not have negative consequences. It simply leads to an extended usage period. Furthermore, even though the app's gamification features are removed after the user finishes the program, the app can still be used as a reference guide to look up information and repeat the stress management exercises.

2.4.3 Summary

Compared to the MVP, Stress-Mentor's full version encompasses a variety of gamification techniques, instead of just an avatar, as well as a large number of behavior change techniques and stress management exercises, that add to the MVP's diary.

With its comprehensive content, including behavior change, stress management and gamification theories, Stress-Mentor includes all three recommended aspects to ensure the app's effectiveness: predisposing factors, enabling factors, and reinforcing factors (Payne et al., 2016). It includes information regarding attitudes and knowledge, increases awareness, helps to evaluate beliefs and values, and promotes confidence, motivation, and self-efficacy. It also teaches a large number of stress management and behavior change skills and allows the tracking of stress-related behaviors through a diary. Moreover, it reinforces the desired behaviors through an extensive rewards and feedback system.

Chapter 3

A Systematic Comparison of the MVP with the Final Version of "Stress-Mentor" Regarding Perceived App Quality

Though the evaluation of the MVP showed that gamification in form of an avatar was met with a positive attitude by the users, there was still room for improvement. This was shown by the app's quality ratings, especially regarding the MVP's perceived impact and engagement (see Figure 4 for the uMARS ratings). This finding was confirmed by the participants' comments after using the MVP for four weeks. After integrating the users' feedback from this first study into Stress-Mentor's extended design (see Chapter 2.4 for details), the full version's general quality was again evaluated with the uMARS questionnaire in a four-week user study. The aim of Chapter 3 is to assess Stress-Mentor's quality and to reveal improvements in comparison to the MVP. This allows inferences on the success of the app's iterative development process. A slightly modified version of this chapter was previously published in Hoffmann et al. (2019).

3.1 Introduction

3.1.1 Background

Chronic stress and its negative impact on health are a growing problem in today's society (Wiegner et al., 2015). A solution for this problem that is met with increasing approval is the use of health-promoting smartphone applications (mHealth). There is already a large number of stress management apps available on the market (e.g., "Ovia", "Mevi", "DeStressify" and "myCompass"). First evidence that the use of such apps can have positive impact on a person's stress level has already been published (Ahtinen et al., 2013; Economides et al., 2018; Lee and Jung, 2018).

To ensure their effectivity, stress management apps should integrate evidence-based content. This includes well-established coping and relaxation methods, as well as behavior change techniques (Morris et al., 2010; Harrison et al., 2011; Chittaro and Sioni, 2014). However, current app reviews show that only few of the available stress management apps include a broad range of established methods (Coulon et al., 2016; Christmann et al., 2017a). Notwithstanding the importance of appropriate content, these methods alone have been suggested as insufficient to promote long-term behavior change through apps (Vandelanotte et al., 2007). The integration of gamification in mHealth products has been proposed as a possible solution to this problem.

3.1.2 Gamification in Health Apps

Gamification is defined as the use of game elements in non-game contexts (Deterding et al., 2011). It can have positive effects on user experience (Johnson et al., 2016) and usability (Zagel and Bodendorf, 2014), as well as on the user's motivation and engagement (Hamari, 2013). Moreover, the implementation of gamification can improve usage consistency and, thus, result in a greater exposure to the content of mHealth products (Vandelanotte et al., 2007). Nevertheless, it is critically discussed whether gamification should be used in stress management applications. So far, studies report positive as well as negative or neutral effects with respect to behavioral and cognitive aspects. In addition, the effects of gamification are often dependent on the context and aim of the application (Johnson et al., 2016). For example, a study in the context of smoking cessation suggests that the combination of game elements and behavior change techniques can be well received by users (Edwards et al., 2018). On the other hand, the developers of the stress management app "Ovia" report that users do not wish for an integration of game elements in this context. This shows that users might not accept game elements in the context of stress management. Users are clearly reluctant towards linking stress management with gamification (Ahtinen et al., 2013). This might be one reason why, despite its great potential, gamification is hardly found in current stress management applications (Hoffmann et al., 2017). Despite the users' reluctance toward the hypothetical use of gamification in stress management apps, little is known about how game elements are actually perceived in this context.

3.1.3 "Stress-Mentor's" Final Version vs. the MVP

To close the identified gaps, a first prototype of the stress management app Stress-Mentor was developed. As described in Chapter 2.1, this MVP (Lenarduzzi and Taibi, 2016) combined the self-monitoring of stress-relevant behaviors and events through a diary with vicarious reinforcement through the appearance of an avatar. The MVP's general quality was then assessed in a longitudinal study. The MVP was tested by 26 participants over a period of 4 weeks (Christmann et al., 2018b). After the four-week trial period the participants rated the MVP's quality with the user Mobile Application Rating Scale (uMARS, Stoyanov et al., 2016), which was applied as a semi-structured interview in order to receive suggestions for improvement. The results of this first study showed that linking the self-monitoring of stress-relevant behaviors with vicarious reinforcement through an avatar was well received by users. However, the study also revealed that the inclusion of more game elements might be needed to improve the usage behavior (Christmann et al., 2018b).

Including the user into the development process in this manner is an important method to ensure the functionality of health technologies. Thus, as is recommended in iterative app development (Stinson et al., 2010), the user feedback from this first study was used to create an extended app concept.

Based on the feedback, additional stress management, behavior change, and gamification features were incorporated in Stress-Mentor's full version. For example, in addition to self-monitoring through a diary and its long-term visualization (Christmann et al., 2017b), the full version includes daily stress management exercises (i.e., breathing, progressive muscle relaxation, meditation, guided imagery, euthymic methods, physical exercises, cognitive aspects, time management, setting priorities, and planning social support and change). Moreover, it links these stress management aspects with a number of game elements. Besides the avatar, the full version includes an agent that guides the user through the app by explaining new functions and providing tips on stress and stress management. When using the app for the first time, the agent hands over the responsibility of raising the avatar to the user through a behavioral contract, which provides a narrative context. This means, in addition to reflecting the user's diary entries through its appearance, the user's progress is visualized through the avatar's growth in the full version. The user can capture this progress in a photobook. The avatars pictures can also be shared with others. This provides a social component. Additionally, the user's progress is visualized through progress bars. The full version also supports goal setting by providing the stress management exercises in "tasks of the day" and "tasks of the week". Here, the user can choose one out of three suggested exercises he or she wants to accomplish each day. The app's consistent usage is rewarded through badges. Another reward system is the points the user receives for every task and diary entry. These points can be exchanged for items for the user's avatar in a shop. The addition of new gamification aspects over time is aimed at upholding the user's curiosity. This is further supported through random sales in the app's shop, randomly provided motivational quotes of famous people, and the randomly provided option to take pictures of the avatar.

The resulting full version of Stress-Mentor, thus, combines an extended gamification framework with evidence-based stress management methods and behavior change techniques. For more information regarding Stress-Mentor's concept, see Chapter 2.4 or Christmann et al. (2018a). Screenshots and content of both the MVP and the full version are displayed in Figure 21.

The full version's general quality was again evaluated in a four-week user study using uMARS. The aim of this study was to assess Stress-Mentor's quality and to reveal improvements compared to the MVP.

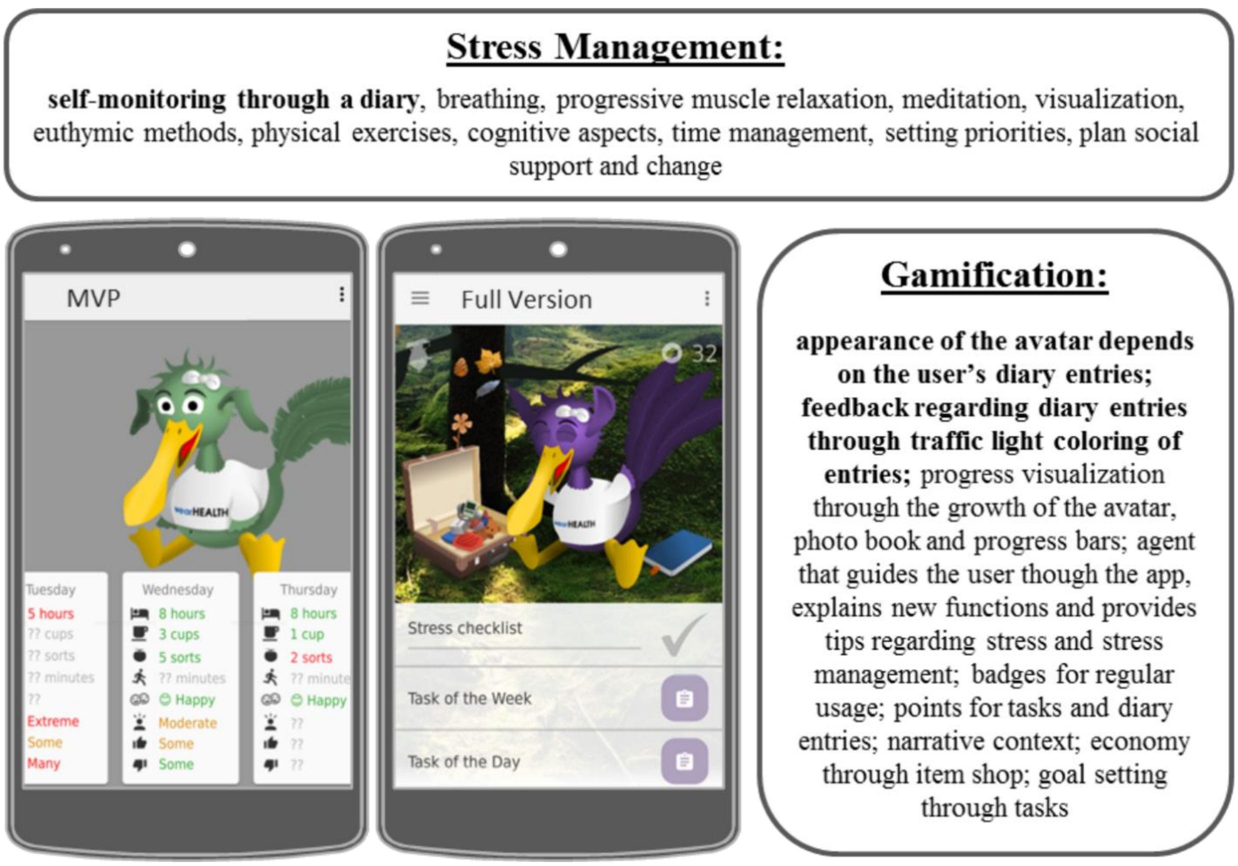

Figure 21: Screenshots and content of Stress-Mentor's first test version (minimal viable product, MVP, left screenshot) and its full version (right screenshot). The app combines evidence-based stress management techniques with gamification. The content already included in the MVP is highlighted in bold print.

3.2 Methods

3.2.1 Participants

Participant recruitment was carried out via an e-mail distribution list of the Technische Universität Kaiserslautern and associated research institutes, as well as an article in a local newspaper. All participants gave written consent to participate in accordance with the declaration of Helsinki.

In 2016, the MVP was tested by 26 participants (*M* age = 23.38 years, *SD* = 3.01 years, 11 female, 15 male) in a first user study (Christmann et al., 2018b). Of these participants, 6 were in employment and 20 were university students.

In 2018, Stress-Mentor's full version was again tested in a user study with 20 participants. However, one participant did not return for the final study appointment. This resulted in a total of 19 participants (*M* age = 33.6 years, *SD* = 8.8 years, 10 female, 9 male) that were included in the analysis. Of these participants, 12 were in employment and 7 were university students.

To ensure that the testing conditions were as realistic as possible, the app was installed on the users' own smartphone (Android version 4.4 or higher) and then used for four weeks in both studies. Therefore, the regular use (at least once a day) of a smartphone or tablet was a prerequisite for participation. In addition, participation in both studies required a minimum age of 18 years and fluency in German (all app instructions are only available in German).

3.2.2 Study Procedure

At the beginning, each participant was informed about the course and aim of the study, as well as about the collected data. Each participant gave his or her written consent to participate. Subsequently, the participant's demographic data was collected and the lead investigator installed the full version of the stress management app on the participant's smartphone or tablet. After the installation, a brief introduction was given by the lead investigator. First, the participant had to adjust the settings to his or her gender and age, as well as his or her preferences (tracking of alcohol and coffee consumption, reminder, color schemes and text size). Then, he or she made an entry in the app's health diary (sleep duration and quality, duration and intensity of sport, positive and negative events, general degree of stress, prevailing mood, digestion, consumption of water, vegetables and fruit, caffeinated drinks and alcohol). This was followed by a four-week usage period, during which participants should use the app daily.

The study was completed with a second meeting that took place at least one day and a maximum of 14 days after the end of the four-week usage period. As was done with the MVP, the participants then again rated the full version's quality using uMARS (Stoyanov et al., 2016).

3.2.3 App Quality

The user version of the mobile application rating scale was applied to assess the quality of both the MVP and the app's full version. uMARS was specifically designed to assess the quality of health apps from the user perspective (Stoyanov et al., 2016).

Participants rated each of the questionnaire's items on a 5-point Likert-scale (from 1 inadequate to 5 excellent). uMARS was applied as a semi-structured interview. This means that, after each rating, participants had the opportunity to explain their answer and provide feedback and suggestions for changes in the app (open response format). Presenting questionnaires as semi-structured interviews provides deeper insight into the reasoning for the ratings, as well as suggestions for possible improvements (e.g., Anderson et al., 2016a).

A total of three questions were removed from the questionnaire. One question from the category information (16. "Does the app come from a credible source?") was removed because the participants were informed of the app's source in detail prior to participation. Furthermore, two questions from the category subjective quality were removed: 18. "How many times do you think you would use this app in the next 12 months if it was relevant to you?" because this question was not included in the MVP questionnaire due to a shorter targeted usage period and 19. "Would you pay for this app?" because the app is available for free. The term "health behavior" in the questions of the app specific section was replaced with "stress management".

3.2.4 Analysis

In order to check whether the addition of established stress management methods and other gamification techniques had a positive effect on the perceived quality of the app, the results from both studies were tested for differences. Because the data was not normally distributed, (all $p < .05$) non-parametrical tests (namely Mann-Whitney U tests for two independent samples) were used to identify differences in the central tendencies of the uMARS ratings of the full version and the MVP (Field, 2009). Bonferroni correction was used to adjust the statistical results for an accumulation of alpha errors (new alpha-level $p \leq .013$). The effect size (r) according to Cohen (1992) was calculated. Effects r = 0. 10 – 0.30 correspond to a small effect, r = 0.30 - 0.50 a moderate effect and $r \geq 0.50$ a large effect.

3.3 Results

The analysis revealed significantly higher ratings for the full version (Median (*Mdn*) = 4.55) compared to the MVP (*Mdn* = 4.03) regarding the app's general quality; $U = 71.00$, $p \leq .001$, $r = 1.54$. In detail, improvements in the full version (*Mdn* = 4.20) compared to the MVP (*Mdn* = 2.80) could be observed for the category engagement; $U = 39.00$, $p \leq .001$, $r = 2.05$. Moreover, higher ratings were observed in all questions of this category (i.e., entertainment, interest, customization, interactivity and target group). With regard to aesthetics, the full version (*Mdn* = 4.67) also received better ratings than the MVP (*Mdn* = 4.33); $U = 113.00$, $p = .001$, $r = 1.07$. However, here only the question of visual appeal showed a significant improvement, while no difference could be found between the rating regarding Stress-Mentor's layout and graphics. An improvement of the app's full version (*Mdn* = 4.67) in comparison with the MVP (*Mdn* = 4.00) could also be detected regarding the category information, $U = 142.50$, $p = .009$, $r = 0.81$. The ratings for both the quality and quantity of the included information increased. No difference between MVP and full version was identified regarding the app's visual information. Even though the app is now much more complex and has a lot more features, ratings for the app's functionality did not decrease and its functionality was assessed as excellent for both the MVP (*Mdn* = 4.75) and full version (*Mdn* = 4.75); $U = 216.50$, $p = .34$, $r = 0.27$. This is also reflected in the rating of each of the questions regarding the app's functionality (i.e., performance, ease of use, navigation, and gestural design). There was no difference between MVP and full version with respect to these aspects. In contrast, MVP (*Mdn* = 2.83) and full version (*Mdn* = 3.83) differed with respect to their perceived impact; $U = 105.00$, $p = .001$, $r = 1.15$. The app's full version received higher ratings for all questions in this category (i.e., awareness, knowledge, attitudes, intention to change, help seeking, and behavior change). In addition, the full version (*Mdn* = 4.00) received a better overall star rating than the MVP (*Mdn* = 3.00); $U = 135.00$, $p = .004$, $r = 0.87$. Also, participants were more likely to recommend the full version (*Mdn* = 5.00) of Stress-Mentor to others in comparison to the MVP (*Mdn* = 3.00); $U = 96$, $p < .001$, $r = 1.24$. The corresponding medians for the three main uMARS categories (general quality, subjective quality, and perceived impact) of both studies are displayed in Figure 22. Means, standard deviations, and statistical values for each item are found in Table 7.

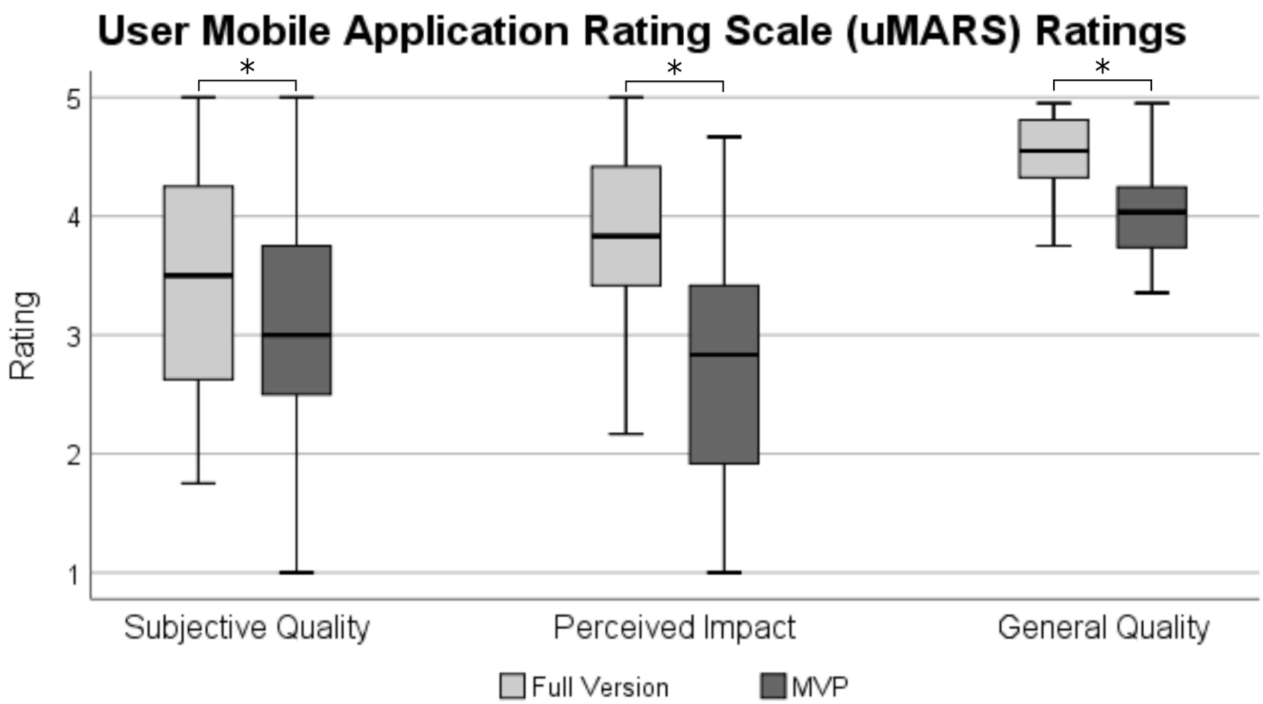

Figure 22: uMARS (user Mobile Application Rating Scale) ratings of the MVP (minimal viable product) and the full version of Stress-Mentor are displayed. Depicted are the medians, maximum and minimum values, as well as first and third quartiles. Significant differences between the groups are marked with an asterisk (*).

Table 7: Means (*M*), standard deviations (*SD*), and statistical values of Mann-Whitney-U tests for uMARS ratings of the MVP and the full version of Stress-Mentor by item. The new alpha level is set at .025 according to the Bonferroni correction method.

Category	Item	*M* MVP	*M* Full-Version	*SD* MVP	*SD* Full-Version	*U*	*p*	*r*
Engagement	Entertainment	4.11	3.00	0.19	0.18	99.00	≤ .001	0.56

Category	**Item**	***M* MVP**	***M* Full-Version**	***SD* MVP**	***SD* Full-Version**	***U***	***p***	***r***
	Interest	4.42	2.89	0.14	0.20	62.00	≤ .001	0.66
	Customization	4.11	2.7	0.24	0.24	108.50	.001	0.50
	Interactivity	3.84	2.22	0.19	0.17	53.00	≤ .001	0.69
	Target Group	4.63	3.81	0.16	0.16	115.00	.001	0.50
Functionality	Performance	4.79	4.74	0.12	0.09	234.00	.48	0.10
	Ease of Use	4.58	4.81	0.14	0.08	207.00	.15	0.21
	Navigation	4.53	4.81	0.12	0.08	182.50	.04	0.31
	Gestural Design	4.95	4.85	0.05	0.07	232.00	.31	0.15
Aesthetics	Layout	4.79	4.59	0.10	0.10	206.00	.17	0.20
	Graphics	4.79	4.44	0.12	0.12	177.00	.03	0.31
	Visual Appeal	4.42	3.67	0.12	0.16	126.00	.002	0.47
Information	Quality of Information	4.42	3.56	0.14	0.25	150.50	.012	0.37
	Quantity of Information	4.63	3.85	0.11	0.18	121.00	.001	0.49
	Visual Information	4.32	4.59	0.27	0.11	229.00	.48	0.10
Subjective Quality	Recommendation to Others	4.32	2.89	0.22	0.24	96.00	≤ .001	0.54
	Overall Star Rating	4.21	3.33	0.16	0.21	135.00	.004	0.42
Perceived Impact	Awareness	3.74	3.19	0.27	0.23	182.50	.09	0.25
	Knowledge	3.84	2.67	0.25	0.28	136.00	.006	0.41

Category	Item	*M* MVP	*M* Full-Version	*SD* MVP	*SD* Full-Version	*U*	*p*	*r*
	Attitudes	3.58	2.52	0.35	0.25	152.00	.02	0.35
	Intention to Change	4.11	3.00	0.26	0.29	142.50	.009	0.39
	Help Seeking	3.37	1.93	0.29	0.18	96.00	≤ .001	0.54
	Behavior Change	3.95	2.59	0.24	0.29	126.50	.003	0.44

3.4 Discussion

In summary, although not all aspects showed significant improvements between the MVP and the full version, there is a general trend towards enhancing the overall quality of the full version compared to the MVP. The positive ratings with regard to entertainment, interest, customization, interactivity and target group show that the use of gamification can affect perceived user engagement in a positive manner. This is in agreement with the results of previous studies showing that gamification can have positive impact on motivation and engagement (Hamari, 2013). Increasing user engagement in this manner is important to improve usage consistency and can, thus, increase an intervention's effectivity (Zagel and Bodendorf, 2014). This is supported by the technology acceptance model (TAM3) which identified joy of use as one major predictor for the intention to use technologies (Venkatesh and Bala, 2008).

Besides the gamified context, the integration of an extensive number of stress management techniques and information regarding stress and stress relevant behaviors can explain why the participants are more likely to recommend the full version than the MVP. It is also very likely the cause for the improved overall star rating. MVP users commented that they wished for the integration of more stress management methods (Christmann et al., 2018b). The integration of an extensive number of stress management and behavior change techniques in response to this feedback can also explain the more positive evaluation of the full version's perceived impact. It also explains the increase of the ratings with regard to the app's quality and quantity of information. This emphasizes the importance of integrating evidence-based exercises and behavior change techniques into mHealth products (Vandelanotte et al., 2007). Nonetheless, some users suggested including more graphics, pictures and videos for the stress management exercises.

Moreover, references and links supporting the user in the search for professional help were added to Stress-Mentor's full version. These links were included in both a menu item and in specific tasks. This likely resulted in a more positive rating regarding the likelihood of the user to seek further help.

Although the full version is much more complex than the MVP, there was no significant deterioration in Stress-Mentor's navigation and visual information. The functionality of both app versions was rated as excellent. This indicates that the use of gamification could positively affect the system's usability (Zagel and Bodendorf, 2014).

The positive evaluations regarding the general quality of the full version suggests that the combination of established stress management methods and gamification (Christmann et al., 2018a) that was applied in Stress-Mentor was well received by the users. Although some users of stress management apps explicitly oppose the use of gamification (Ahtinen et al., 2013) the positive evaluation of Stress-Mentor shows that this is not necessarily the case. The users' comments emphasized that the embedding of stress-related aspects into a gamified context was perceived in an overall positive way. In fact, the participants found the app interesting to use and enjoyed exploring its contents. This is in line with previous results that indicate that the combination of gamification with evidence-based content from stress management and behavior change could have positive effects on the user's engagement and, thus, potentially make health apps more effective (Edwards et al., 2018).

However, it is apparent that users are not aware of the effect of gamification (Thorpe and Roper, 2017). Some of the participants said that the gamification concept would not have been necessary in their opinion. However, they also described that they tried to adjust their behavior so that they could make positive entries in their diary in order to make their avatar look healthier. This observation coincides with previous results on the effectiveness of vicarious reinforcement through avatars (e.g., Fox et al., 2009). It further underpins the potential usefulness of gamification in the context of stress management.

This demonstrates that it is not sufficient to ask potential users whether gamification is desired in a particular context. Rather, studies must examine if the implementation of game elements in a specific context appeals to the user and how their use influences the effectiveness of the product (Cafazzo et al., 2012; González et al., 2016). When integrating gamification, the context and goal of the app should therefore always be taken into account (Johnson et al., 2016). The positive response to combining gamification with stress management in Stress-Mentor is probably due to the fact that the integrated gamification elements support the actual goal of the app, namely the learning of stress management methods through daily and weekly tasks. Nonetheless, little is known about the effect of gamification on actual usage behavior. This aspect is the focus of Chapter 4, where the usage behavior of a gamified version of Stress-Mentor and a not gamified version of the app was investigated.

3.5 Conclusion

In summary, an overall trend of improvement could be observed for Stress-Mentor's full version. The positive ratings confirm that Stress-Mentor is of good general quality, speaking for the app's successful, iterative development. This study, therefore, shows that an iterative development process involving the user can lead to an improvement in the product's overall quality.

Even though some users of stress management apps were opposed to the use of gamification (Ahtinen et al., 2013), this study demonstrates that the combination of stress management methods with gamification was, in fact, well received. This highlights the potential usefulness of gamification in the context of stress management.

Chapter 4

Evaluation of “Stress-Mentor’s” Gamified Concept

The previous chapter showed that Stress-Mentor’s iterative development resulted in improvements of the app’s quality and indicated that the app is well-received by users. This suggests that gamification in the context of stress management is met with a positive user attitude. Furthermore, previous studies have shown that the gamification of health apps can increase their usage (Taylor et al., 2019). However, so far little is known as to how gamification affects the usage of stress management apps. This chapter, therefore, focuses on evaluating the effect of gamification on the usage behavior of Stress-Mentor, namely on the number of performed stress management exercises and the amount of time the users spend with the app. In addition to this aspect, this chapter deals with the question of how age mediates the app’s usage. For this purpose, the final version of Stress-Mentor (which was introduced in Chapter 2.4) was compared to a non-gamified version of the app in a four-week user study with a broad sample of young (18-35 years), middle aged (36-50 years), and old users (51-65 years).

4.1 Introduction

4.1.1 Background

Evidence is growing that chronic stress can have negative impact on a person’s physical, as well as mental, well-being (Stansfeld and Candy, 2006; Cohen et al., 2007). However, a person’s well-being does not solely depend on his or her exposure to stress. Rather, it depends on the way a person copes with stress (Cooper, 1994). Lazarus distinguished two types of coping mechanisms with regard to stress management, namely problem-focused (e.g., time management) and emotion-focused coping (e.g., meditation) (Lazarus, 1985). These coping strategies are usually taught in single or group therapy sessions.

In recent years, the use of mHealth technologies, such as smartphone apps, has gained popularity in the context of personal health management (Harrison et al., 2011). Such apps aim at affecting the user’s education, motivation and adherence (Handel, 2011; Ahtinen et al., 2013) and pose an alternative to expensive conventional treatments (Proudfoot et al., 2010; Luxton et al., 2011). “Ovia”, “Mevi”, “DeStressify”, “myCompass”, “Stress-Eraser”, and “AEON", are only some examples of currently available stress management apps. First evidence shows that the use of such apps can have positive impact on a person’s stress level (Ahtinen et al., 2013; Economides et al., 2018; Lee and Jung, 2018). However, in order to ensure their effectiveness (Harrison et al., 2011), they have to include evidence-based content from behavior change theory (Abraham and Michie, 2008) and stress management theory (Williams and Kemper, 2010). Furthermore, it is important to ensure adequate user engagement and motivation. These aspects have great influence on the user’s exposure to the app’s content (Webber et al., 2010) and are, therefore, directly linked to an app’s effectiveness (Vandelanotte et al., 2007). Both motivation and engagement can be increased through the integration of gamification (Cafazzo et al., 2012).

Gamification is aimed at making interventions more enjoyable, motivating, and engaging (Deterding et al., 2011). This goes a long way to ensure the user’s interest (Oinas-Kukkonen and Harjumaa, 2009) and increase his or her exposure to the evidence based content of the intervention (Davies et al., 2012). Empiric evidence supports that the use of gamification techniques can, indeed, increase the use of a service (Hamari, 2013). For example, Taylor et al. (2019) demonstrated that a gamified health diary increased the users’ compliance compared to a not gamified version of the app.

Moreover, it was already shown that the gamification in form of rewards through points for diabetes patients (Cafazzo et al., 2012) and the combination of different gamification techniques for weight management in children (e.g., points, badges, leaderboards, time pressure) can be effective in promoting behavior change through apps (González et al., 2016). Another study that focused on an app for smoking cessation showed that the combination of game elements (e.g., challenges, social pressure, social support, and digital rewards such as trophies) and behavior change techniques (e.g., providing information about health consequences, goal setting, and self-monitoring through a digital diary) can be well received by users (Edwards et al., 2018).

One aspect that needs to be considered when designing gamified health apps, however, is that the acceptance of gamification appears to be affected by the user’s age. Younger adults are more likely to download gamified health apps than older adults (Goyal et al., 2016). A reason for this could be that older adults are more opposed towards gamification

(Thiel et al., 2016). Also, older adults show a lower self-efficacy and more anxiety with respect to the use of technologies, such as health apps (Czaja et al., 2006). Gamification could further increase these existing problems. It has, therefore, been suggested that alternatives without gamification might be preferable for older users (Hierhammer and Herrmanny, 2013). On the other side, older users also show greater sustained engagement and usage as a result of gamification than younger ones (Brauner et al., 2013; Goyal et al., 2016). This shows that, regardless of the users' age, combining gamification with behavior change techniques and stress management methods is a promising approach to ensure the effectiveness of stress management apps.

Nonetheless, app reviews show that few apps on the market make use of these concepts. The available apps vary largely with respect to the amount of included behavior change techniques and stress management methods (Coulon et al., 2016; Christmann et al., 2017a). While some apps do not make use of any evidence-based techniques, those that do mostly focus on one specific coping method. However, there are also exceptions that combine several behavior change and stress management techniques within one app (e.g., "Mevi") (Christmann et al., 2017a). The implementation of gamification in stress management apps, on the other hand, remains an exception. Moreover, app designers do not combine gamification with evidence-based content from behavior change or stress management theory in order to improve usage duration and behaviors (Hoffmann et al., 2017).

One reason why gamification might not be met with enthusiasm in this context could be the ongoing discussion about whether gamification should be used in stress management applications. While many studies revealed positive effects of gamification on behavioral and cognitive aspects, there have also been reports on negative, or neutral effects (Johnson et al., 2016). Furthermore, users of the stress management app "Ovia" did not wish for an integration of game elements in this context (Ahtinen et al., 2013). This suggests that users might neither want, nor accept, the integration of game elements into stress management apps. In contrast, a recently published study with Stress-Mentor's MVP suggests that the combination of an avatar with stress-focused self-monitoring through a digital diary, was met with a positive reaction by users (Christmann et al., 2018b). This was further confirmed by the request of participants from a study that investigated the usability of the neurofeedback-based meditation app MUSE (see Chapter 2.3). Here, the users requested more gamification features in order to keep their sustained interest. Nevertheless, both the context and aim of an application are decisive for the effects of gamification (Johnson et al., 2016). Though the previous results in this thesis suggest a positive user attitude towards gamification, at this point, little is known about how gamification is perceived compared to a non-gamified app version and whether it improves usage consistency in the context of stress management.

4.1.2 Short Summary of "Stress-Mentor's" Concept

To close the identified gaps in research, the stress management app Stress-Mentor was developed in an agile development process (see Chapter 2 for details). In a first step, an MVP (Lenarduzzi and Taibi, 2016) was created that combined self-monitoring through a digital diary with an avatar. This MVP was then tested in a first user study. The results showed that, while the avatar was perceived positively, it might not be sufficient to improve frequency of use (Christmann et al., 2018b). Based on this study, an extended app framework was developed that includes additional stress management, behavior change and gamification features. Besides self-monitoring through a diary and its long-term visualization (see Chapter 2.4 or Christmann et al., 2017b), the full version includes a broad range of daily stress management exercises (i.e., breathing, progressive muscle relaxation, meditation, guided imagery, euthymic methods, physical exercises, cognitive aspects, time management, setting priorities, and planning social support and change).

Stress-Mentor links these stress management aspects with a number of game elements. In addition to the avatar, Stress-Mentor now includes an agent that guides the user through the app, explains new functions, and provides tips on stress and stress management. When the app is used for the first time, the agent hands the responsibility of raising the avatar over to the user. This is done through a behavioral contract and provides a narrative context. The user's progress is now visualized through the avatar's growth, in addition to reflecting the user's diary entries through its appearance. The user can capture changes in his avatar in a photobook. The taken pictures can also be shared with others. This counteracts change blindness and provides a social component without instilling social pressure. Along with the avatar's appearance, the user's advances are visualized through progress bars. Moreover, Stress-Mentor supports goal setting by providing the stress management methods in "tasks of the day" and "tasks of the week". Here, the user can choose one out of three suggested exercises on a daily basis. The app allows for the user to perform more and more tasks as he or she progresses, starting with one and rising up to three per day as he or she advances. The available tasks are based on a weekly stress checklist. Here, the user can enter stress-related ailments (i.e., fears and worries, sadness, anger, stress at work, stress in private life, muscle tension, head, neck and back grievances caused by tension, digestive problems, and sleep problems). Stress-Mentor rewards the consistent usage of the app through badges and points that the user receives, e.g. for completed tasks and diary entries. In a shop, the gained points can then be exchanged for items for the user's avatar. To uphold the user's curiosity, new gamification aspects are added over time. This aspect is furthered through three additional elements: randomly provided motivating quotes of famous persons, the randomly provided option to take pictures of the avatar, and random sales in the app's shop. These elements support unpredictability and curiosity, which play an important role to keep the user interested (Chou, 2015). As a result, Stress-Mentor combines an extended gamification framework with evidence-based techniques from stress management and behavior change theory. See Figure 23 for a screenshot and Christmann et al. (2018a) for further information regarding Stress-Mentor's concept.

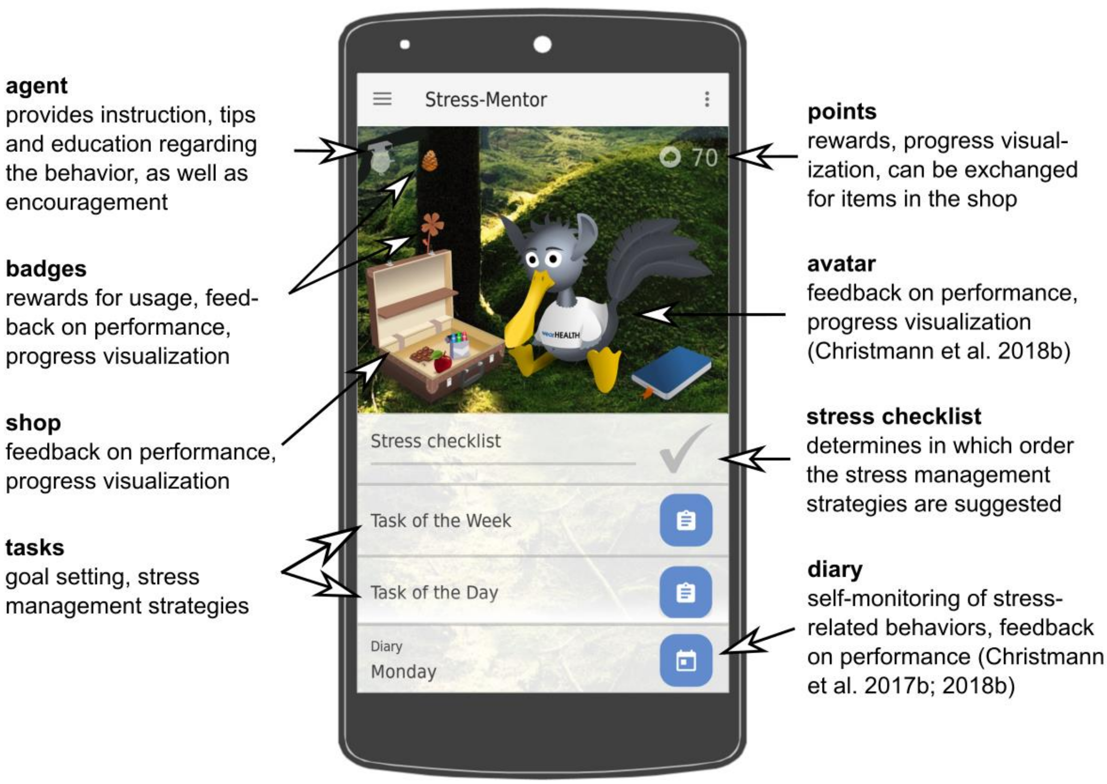

Figure 23: Screenshot of Stress-Mentor.

4.1.3 Study Aim

Previous studies suggested that gamification can have positive effects on user experience (Johnson et al., 2016) and usability (Zagel and Bodendorf, 2014), as well as usage consistency (Vandelanotte et al., 2007). Notwithstanding gamification's positive impact, as of yet, little is known about how the implementation of game elements in the context of stress management apps actually affects these aspects. The aim of this study was therefore to investigate the impact of gamification on user experience, usability, perceived app quality, and the usage behavior in the context of stress management apps. For this purpose, a four-week randomized control study was conducted. The experimental group (EG) received a full version of Stress-Mentor with gamified aspects and the control group (CG) received the app without gamification.

4.2 Methods

4.2.1 Participants

Participants were recruited via an e-mail distribution list of the Technische Universität Kaiserslautern and associated research institutes, as well as an article in a local newspaper. All participants gave written consent to participate in accordance with the declaration of Helsinki. The study procedure was approved by the ethics committee of the Department of Social Sciences at the Technische Universität Kaiserslautern.

Because the app was tested over a period of four weeks on the participants' own smartphones, the regular use (at least once a day) of a smartphone or tablet was a requirement for participation. Also, participants had to be of legal age (18 years) and fluent in German, as all app instructions are only available in German.

A total of 64 participants were randomly assigned to an EG, who tested Stress-Mentor with all game elements, and a CG, who received an app version with the same stress-management techniques but without gamification (Figure 24). The data of seven participants had to be excluded from the analysis (three did not return for the final interview, two had technical problems, and two dropped out during the study for personal reasons). This left N = 57 participants to be included in the analysis (EG (N = 29; 15 female; *M* age = 41.30 years, *SD* = 13.69), CG (N = 28; 16 female; *M* age = 41.23 years; *SD* = 12.95)). χ^2-tests revealed no differences between the two groups regarding gender, profession, use of health apps, use of stress management apps, and the application of stress management methods in everyday life; a t-test revealed no differences between the two groups with regard to age (all $p > .05$).

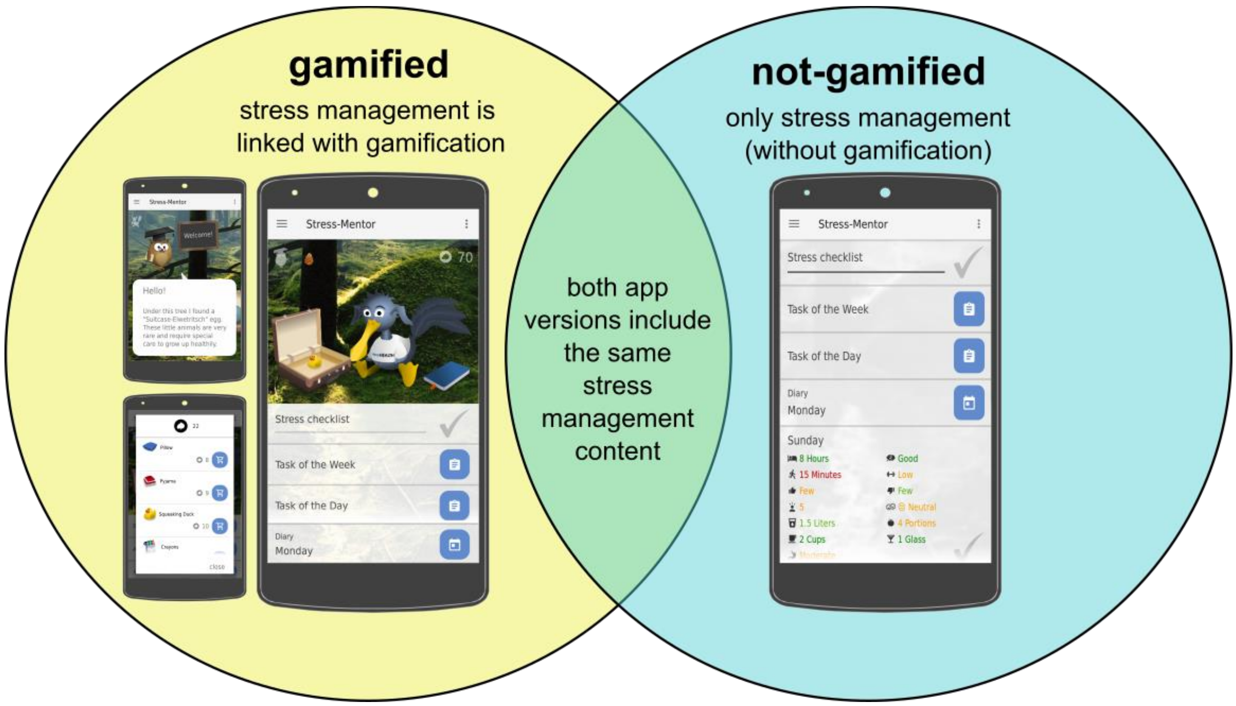

Figure 24: Differences and similarities between the gamified and the not gamified version of Stress-Mentor.

4.2.2 Study Procedure

During the first study appointment, the lead investigator installed a shortened version of Stress-Mentor (all of the app's features could be unlocked and experienced within the study period) on the participant's own smartphone (Android version 4.4 or higher). The participant then used the app for the first time in the presence of the lead investigator following a specific protocol. First, he or she was asked to adjust the app settings according to his or her individual preferences (age, gender, tracking of caffeine and alcohol consumption, step counter, reminder alarms, color scheme, font size). Next, the participant was instructed to fill out the stress checklist (fears and worries, sadness, anger, stress at work, stress in private life, muscle tensions, tension related head, neck and back pains, digestive problems, sleep problems), followed by a diary entry (sleep duration, sleep quality, physical exercise duration, sport intensity, occurrences of positive and negative events, stress level, prevailing mood, digestion, water consumption, fruit/vegetable consumption, and optional: caffeine consumption, alcohol consumption, and step count).

Then, he or she selected one of the available stress management tasks and read the instruction to get an impression of what the tasks entailed. Afterwards the participant used the app for a period of four weeks in his or her daily life, to ensure testing conditions were as realistic as possible. The participant was told to use the app at least once a day, if possible. However, it was left up to the participant how he or she used the app and how long he or she used it. After the four-week trial period the participant returned for a final questionnaire on the app's usability and user experience, as well as a semi-structured interview aimed at rating the general quality of Stress-Mentor and receiving open ended feedback on how the app could be further improved. See Figure 25 for an overview of the study procedure.

4.2.3 App Usage

During the four-week test period, the participant's usage data was recorded. This included the number of tasks that were performed and the time spent using the app.

4.2.4 Usability and User Experience

The usability of the app was measured with the widely used System Usability Scale (SUS, Brooke, 1996), because this questionnaire provides reliable results even for small samples (Sauro, 2011). The SUS consists of 10 questions, which are evaluated on a five-point Likert scale from 1 (strongly disagree) to 5 (strongly agree).

The User Experience Questionnaire (UEQ) was used to assess the user experience. The UEQ covers six aspects (attractiveness, perspicuity, dependability, efficiency, stimulation, and novelty) in 26 pairs of opposites (Laugwitz et al., 2008). The respondents rate the app on a seven-level scale, by indicating whether they tend to assign the app to one adjective or another.

4.2.5 App Quality

The Mobile Application Rating Scale (MARS) was developed specifically for determining the quality of health apps (Stoyanov et al., 2015). In addition to the original version for health and IT experts, a user version of the MARS (uMARS) with simplified wording was published in 2016 (Stoyanov et al., 2016). Like the MARS, the uMARS comprises three categories: subjective quality, perceived impact, and general quality. Questions from the category "perceived impact" were adapted to the topic of stress, i.e. the term "health behavior" was replaced by "stress management".

Participants rated each of the 26 questions on a five-point Likert scale from 1 (inadequate) to 5 (excellent). In order to enable differentiated user feedback, the uMARS was applied as a semi-structured interview (i.e., after each question the participant had the opportunity to explain his or her answer and report on the experiences with the app).

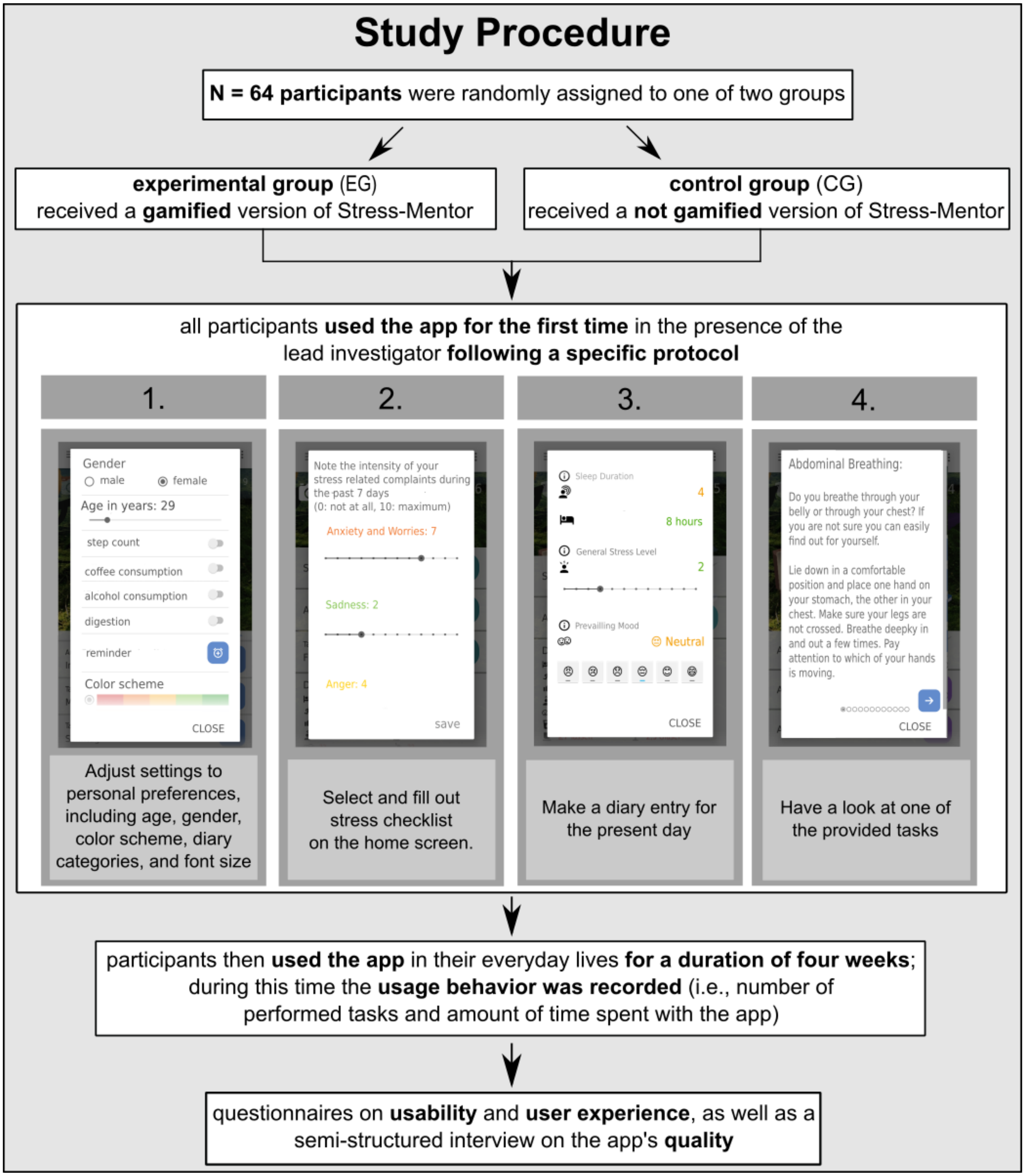

Figure 25: Diagram of the study procedure.

4.2.6 Analysis

Even though the Shapiro-Wilk test revealed that the data was not normally distributed, MANOVA was used for analysis. MANOVA with repeated measures is considered to be robust again violations of normal distribution, provided the sample size of both groups is larger than 20 (Pagano, 2010; Wilcox, 2013). Thus, a MANOVA with two between-subject factors (group (EG and CG) and age (young users = 18-35 years, middle aged users = 36-50 years, and old users = 51-65 years), and the within-subject factor week, was used to detect differences in the groups' usage behavior over time. The usage behavior was operationalized by number of performed tasks and amount of time spent with the app, which were moderately correlated, $r = .60$, $p \leq .001$. Additional ANOVAs and post-hoc tests were applied to determine were the revealed differences originated. The Greenhouse-Geisser adjustment was applied to correct for violations of sphericity. Because the questionnaires' data was not normally distributed (all $p < .05$) non-parametric tests (namely Mann-Whitney U tests for two independent samples (Bortz, 2016) were used to see if the two groups differed in their ratings of Stress-Mentor's usability, user experience and quality. The effect size (*r*) according to Cohen (1992) was calculated.

4.3 Results

4.3.1 Usage Behavior

The MANOVA revealed main effects of group, age, and week. An interaction was detected between week*age. No further interactions were found (see Table 8 for details).

Table 8: MANOVA results for main effects and interactions.

Factor	λ	df 1	df 2	F	p	Part. η^2
group	0.77	2	50	7.43	.001	.23
age	0.80	4	100	3.02	.02	.11
week	0.21	6	46	28.89	≤ .001	.71
group*age	0.91	4	100	1.23	.30	.05
week*group	0.77	6	46	2.25	.06	.23
week*age	0.64	12	94	1.90	.04	.20
week*group*age	0.75	12	92	1.20	.29	.14

Comparable results were revealed by the ANOVAs, with two exceptions. There was a significant interaction between week*group for the number of performed tasks. Furthermore, the interaction between week*age was only apparent for the number of tasks, but not for the amount of time spent with the app (see Table 9 for details).

Table 9: ANOVA results for main effects and interactions for number of performed tasks and amount of time spent with the app.

Factor	Dependent Variable	df 1	df 2	F	p	Part. η^2
group	number of performed tasks	1	106	10.70	.002	.17
	amount of time spent with app	1	86	11.38	.001	.18
age	number of performed tasks	2	106	5.18	.009	.17
	amount of time spent with app	2	86	3.47	.04	.12
week	number of performed tasks	3	106	40.57	≤ .001	.44
	amount of time spent with app	3	86	16.14	≤ .001	.24
group*age	number of performed tasks	2	106	0.86	.43	.03
	amount of time spent with app	2	86	2.28	.11	.08
week*group	number of performed tasks	3	106	5.94	.003	.10
	amount of time spent with app	3	86	0.34	.68	.01
week*age	number of performed tasks	6	106	2.97	.02	.10
	amount of time spent with app	6	86	0.32	.84	.01
week*group *age	number of performed tasks	6	106	1.19	.32	.04
	amount of time spent with app	6	86	1.03	.41	.04

T-tests revealed that the EG spent more time with the app ($t(34) = 2.80$, $p = .01$) and performed more tasks ($t(55) = 3.02$, $p = .01$) compared to the CG (see Figure 26 for details). No differences were detected between the usage behaviors of young and middle aged, as well as middle aged and old users. However, old users performed more tasks and spent more time using the app than young users (Table 10).

Table 10: Post-hoc t-test results for differences between young, middle aged and old users regarding the usage behaviors.

Dependent Variable	Age Groups	df	t	p	d
number of performed tasks	young - middle aged	38	-1.86	.07	0.60
	young - old	35	-2.70	.01	0.91
	middle aged - old	35	-0.98	.33	0.33
amount of time spent using the app	young - middle aged	38	-1.92	.06	0.62
	young - old	26	-3.00	.01	1.18
	middle aged - old	35	0.15	.88	0.05

The number of performed tasks increased in both groups over time, with no further increase between week 3 and 4 (see Table 11). This trend was more distinct in the EG (Figure 26).

Additionally, the EG performed more tasks compared to the CG in each week (all $p \leq .05$, see Table 12). The amount of time the users spent using the app decreased after the first week, then stayed constant throughout weeks 2 and 3 (see Table 13 and Figure 27).

Table 11: T-test results for within-groups comparisons regarding the number of performed tasks per week.

Group	Weeks	df	*t*	*p*	*d*
experimental group	week 1 – week 2	28	-5.99	≤ .001	2.26
	week 1 – week 3	28	-6.40	≤ .001	2.42
	week 1 – week 4	28	-6.28	≤ .001	2.37
	week 2 – week 3	28	-4.68	≤ .001	1.77
	week 2 – week 4	28	-3.83	.001	1.45
	week 3 – week 4	28	-0.16	.87	0.06
control group	week 1 – week 2	27	-4.76	≤ .001	1.83
	week 1 – week 3	27	-4.58	≤ .001	1.76
	week 1 – week 4	27	-3.07	.005	1.18
	week 2 – week 3	27	-2.30	.03	0.89
	week 2 – week 4	27	-0.54	.59	0.21
	week 3 – week 4	27	1.56	.13	0.60

Table 12: Post-hoc t-test results for differences between the experimental and control group with regard to the number of performed tasks per week.

Dependent Variable	Week	df	*t*	*p*	*d*
number of performed tasks	week 1	47	2.45	.02	0.72
	week 2	55	2.05	.046	0.55
	week 3	55	2.41	.02	0.65
	week 4	55	2.97	.004	0.80

Table 13: T-test results regarding differences between the weeks regarding the amount of time users spent with the app.

Dependent Variable	Weeks	df	*t*	*p*	*d*
amount of time spent using the app	week 1 – week 2	56	6.57	≤ .001	1.76
	week 1 – week 3	56	3.45	.001	0.92
	week 1 – week 4	56	6.27	≤ .001	1.68
	week 2 – week 3	56	-0.25	.81	0.07
	week 2 – week 4	56	1.89	.06	0.51
	week 3 – week 4	56	1.95	.06	0.52

The number of performed tasks increased over time in all three age groups. However, this increase was more distinct with increasing age (see Table 14 and Figure 26). The young users performed less tasks than the middle aged users in week 2 and less than the old users in week 2, 3 and 4. Middle aged and old users did not differ in the number of tasks they performed each week (see Table 15).

Table 14: T-test results for within-groups comparisons for young, middle aged, and old users, regarding the number of performed tasks per week.

Group	Weeks	df	*t*	*p*	*d*
young	week 1 – week 2	19	-1.84	.08	0.84
	week 1 – week 3	19	-2.27	.04	1.04
	week 1 – week 4	19	-2.26	.04	1.04
	week 2 – week 3	19	-2.01	.06	0.92
	week 2 – week 4	19	-1.90	.07	0.87
	week 3 – week 4	19	-0.06	.95	0.03
middle aged	week 1 – week 2	19	-6.59	≤ .001	3.02
	week 1 – week 3	19	-5.83	≤ .001	2.68
	week 1 – week 4	19	-4.00	.001	1.84
	week 2 – week 3	19	-2.61	.02	1.20
	week 2 – week 4	19	-0.86	.40	0.40
	week 3 – week 4	19	1.22	.24	0.56
old	week 1 – week 2	16	-6.30	≤ .001	3.15
	week 1 – week 3	16	-6.82	≤ .001	3.41
	week 1 – week 4	16	-5.49	≤ .001	2.75
	week 2 – week 3	16	-4.27	.001	2.14
	week 2 – week 4	16	-2.49	.02	1.25
	week 3 – week 4	16	0.29	.78	0.15

Table 15: Post-hoc t-test results for differences between young, middle aged, and old users regarding the number of performed tasks per week.

Week	Age Groups	df	*t*	*p*	*d*
week 1	young - middle aged	38	-1.67	.11	0.54
	young - old	35	-1.34	.20	0.45
	middle aged - old	35	0.06	.96	0.02
week 2	young - middle aged	38	-2.88	.01	0.93
	young - old	35	-4.16	.001	1.41
	middle aged - old	35	-1.64	.12	0.55
week 3	young - middle aged	38	-1.59	.13	0.52
	young - old	35	-3.08	.007	1.04
	middle aged - old	35	-1.36	.19	0.46
week 4	young - middle aged	38	-1.91	.07	0.62
	young - old	35	-2.39	.03	0.81
	middle aged - old	35	-0.67	.51	0.23

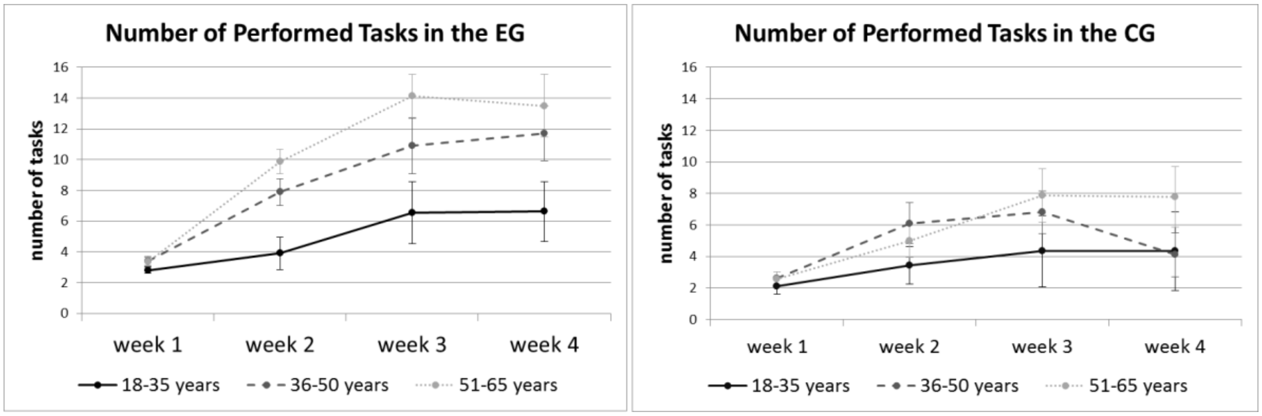

Figure 26: Means and standard errors of the number of performed tasks for the experimental group (EG) who received the gamified version of Stress-Mentor, and the control group (CG) who received a non-gamified version of the app, for young (18-35 years), middle aged (36-50 years), and old users (51-65 years).

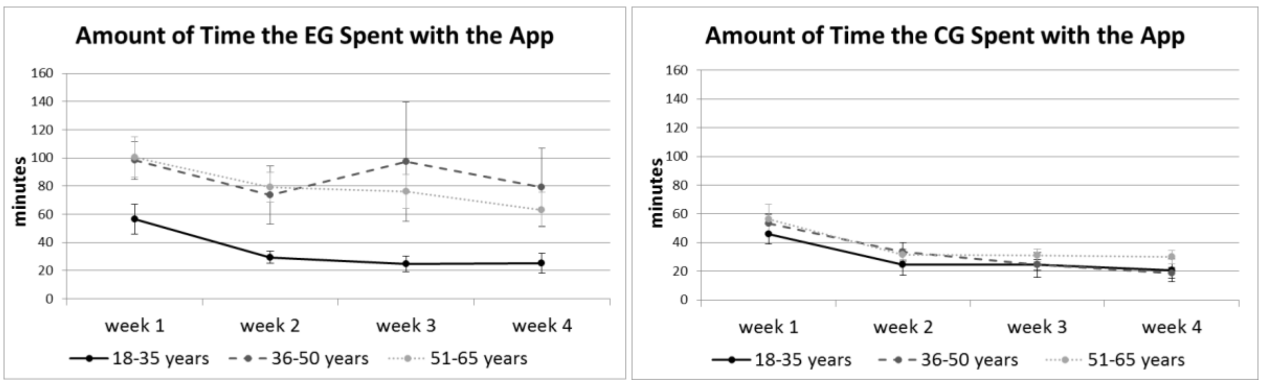

Figure 27: Means and standard errors of the amount of time users spent with the app for the experimental group (EG) who received the gamified version of Stress-Mentor, and the control group (CG) who received a non-gamified version of the app, for young (18-35 years), middle aged (36-50 years), and old users (51-65 years).

4.3.2 Usability and User Experience

The SUS scores were rated to be above the average for both the gamified and not gamified version of the app ((EG) *Mdn* = 95.00; (CG) *Mdn* = 90.00). In fact, the ratings are within the top 10% and therefore correspond to a grade of "A" (Sauro, 2011).

The UEQ ratings (see Table 16) showed that the user experience for both the gamified and not gamified version of Stress-Mentor was evaluated positively (all mean values higher than 0.8, Schrepp, 2015). Mann-Whitney U tests revealed no differences between the SUS and UEQ ratings of the EG and CG (all $p > .05$).

4.3.3 App Quality

The ratings regarding the app's perceived impact ((EG) *Mdn* = 3.83; (CG) *Mdn* = 3.75; U = 330, $p = .22$, $r = 0.02$) and subjective quality ((EG) M = *Mdn* = 4.25; (CG) *Mdn* = 3.88; U =

290.5, *p* = .06, *r* = 0.03) did not differ between the two groups. The app's general quality received higher ratings in the EG (*Mdn* = 4.57) compared to the CG (*Mdn* = 4.32, *U* = 254.5, *p* = .016, *r* = 0.04; see also Figure 28 and Appendix 19 and 20).

Table 16: Means (*M*), standard deviations (*SD*), and statistical values of Mann-Whitney-U tests for ratings of the User Experience Questionnaire of the MVP and the full version of Stress-Mentor.

UEQ Scale	*M* CG	*M* EG	*SD* CG	*SD* EG	*U*	*p*	*r*
attractiveness	1.89	1.52	0.89	0.71	341.00	.30	0.02
perspicuity	1.87	2.15	0.84	0.68	290.50	.06	0.03
efficiency	2.02	1.68	0.66	0.61	332.00	.23	0.02
dependability	1.31	1.45	0.68	0.74	346.00	.34	0.02
stimulation	1.27	0.91	1.03	0.93	354.00	.41	0.01
novelty	1.41	0.91	0.84	0.89	298.00	.08	0.03

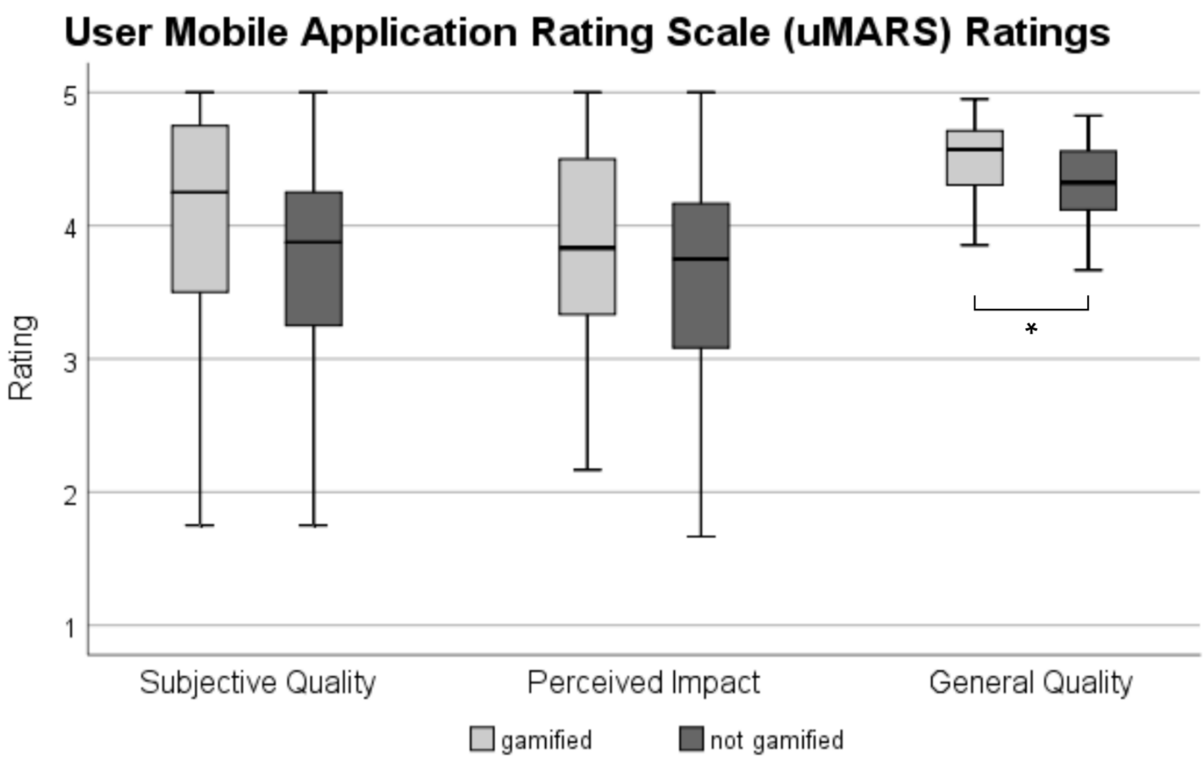

Figure 28: User Mobile Application Rating Scale ratings for the experimental group (EG) who received the gamified version of Stress-Mentor, and the control group (CG) who received a non-gamified version of the app. Depicted are the medians, maximum and minimum values, as well as first and third quartiles. Significant differences between the ratings of both groups are marked with an asterisk (*).

4.4 Discussion

4.4.1 Usage Behavior

The results revealed that users who received the gamified version of Stress-Mentor practiced more stress management techniques and spent more time using the app. This shows that gamification had a positive impact on the usage behavior of Stress-Mentor. Thus, it increased the contact with the stress management and behavior change content. This goes in line with the results of previous studies (Davies et al., 2012) and suggests that gamification could increase the app's effectiveness (Taylor et al., 2019).

In agreement with the app's design (which allows for the user to perform more and more tasks, starting with one and rising up to three per day as he or she progresses), both groups showed an increase in the number of stress management exercises that were performed over time. Even though no interaction between week*group was detected through the multivariate tests, this outcome contradicts the results of the univariate tests, where the group modifies the effect of week on the number of performed tasks. This is supported by the post hoc-tests, which showed that the increase in performed tasks was strongest in the EG. Stress-Mentor's gamified concept (see Chapter 2.4 for details), therefore, led to increased exposure to the app's stress management content. Taylor et al. (2019) reported a similar effect. In their study, gamification led to a significantly higher number of diary entries compared to a non-gamified control condition. This supports previous results about gamification and its positive impact to the user's exposure to the evidence-based content of health apps (Davies et al., 2012; Hamari, 2013).

In contrast to the increase in performed tasks, the time participants spent with the app decreased after the first week of the usage period. This initial drop is likely due to the increased interest at the beginning of using a service (Koivisto and Hamari, 2019). This effect has been previously observed for health apps in other contexts, where the usage also peaked during the first week of the usage period and then declined (Katule et al., 2016). Despite the observed novelty effect, the results of the presented study reveal that gamification sustained its positive impact on the usage behaviors over the trial period (Bamidis et al., 2016).

Another important aspect that affected the usage behavior was the participants' age. Old users performed more tasks and spent more time using the app than young users. Also, the increase in the number of performed tasks over time was stronger for middle aged and old users, compared to young users. This coincides with previous studies that showed an increase in sustained engagement in older adults compared to younger adults (Brauner et al., 2013). A reason for these findings could be that conscientiousness to use such interventions increases with age (Roberts et al., 2006). Previous research shows that conscientiousness can positively affect a person's health behaviors (Bogg and Roberts, 2004). It suggests that stress management apps could become more effective with increasing age, as older users are more likely to maintain the use of such interventions (Goyal et al., 2016; Koivisto and Hamari, 2019). Though the effect of week was modified by age, age did not modify the effect of group.

4.4.2 Usability, User Experience and Quality Ratings

Previous studies suggest that gamification can also positively affect an app's usability (Zagel and Bodendorf, 2014). This could not be confirmed in the present study as both app versions' usability was assessed to be very good (Sauro, 2011). Gamification has also been connected to an increase in user experience ratings. Higher ratings would especially be expected regarding novelty and stimulation, as these aspects examine a system's hedonic quality (Schrepp, 2015). Hedonic quality is highly related to the construct "joy of use" from the technology acceptance model. High ratings in these categories would therefore raise the intention to use the app (Venkatesh and Bala, 2008). Because users in the EG used the app more consistently than users in the CG, it was expected that the gamified version of Stress-Mentor received higher ratings on the novelty and stimulation scales. Surprisingly, this was not the case. However, both app versions were evaluated positively with respect to all user experience aspects (Schrepp, 2015).

Also, both groups rated Stress-Mentor's perceived impact and subjective quality to be good. Only the overall quality was rated as excellent in the EG, compared to good in the CG. This was due to the gamified version of Stress-Mentor being assessed as more aesthetic. As both app versions received positive ratings in all tests, ceiling effects could be the reason that no group differences were detected. Another reason could be that the sample size was too small to uncover effects. Additionally, the non-gamified version of Stress-Mentor already included several visual and interactive aspects (e.g., tasks of the day/week, diary and diary overview). Therefore, the design of the control version might have led to the lack of differences regarding the usability and user experience ratings. This is supported by comments from users of the CG who stated that they thought the broad number and range of stress management tasks was perceived as interesting and engaging.

4.4.3 Participants' Feedback to "Stress-Mentor"

Users of the CG mentioned that they would have liked game elements, such as badges and points, to be included. They said this would have motivated them into using the app more intently. This shows that users of health apps expect and demand mHealth contents to be gamified, in order to present the content in an interesting manner (Oinas-Kukkonen and Harjumaa, 2009) and increase both the user's engagement and motivation (Deterding et al., 2011). This is further supported by the fact that the EG perceived the included game elements in a positive manner. This goes in line with previous findings that support the use of gamification in the context of stress management (Christmann et al., 2018b) and further supports the app's gamified concept (Christmann et al., 2018a). Many participants in the EG shared the conclusion that the app's gamification had a motivating effect on them and reported that they enjoyed using the app. This positive reception of gamification contradicts the suggestion that the use of game elements in stress management apps is not wanted by users (Ahtinen et al., 2013). Two older participants said that they would not have needed gamification. Nevertheless, they also reported paying more attention to their stress-related behaviors and that they tried to meet the health recommendations in the diary in order to positively influence their avatar's appearance. This goes in line with previous studies on vicarious reinforcement which showed that linking an avatar's appearance with the user's performance positively affects health behaviors such as physical exercise (Fox et al., 2009). However, it also indicates that older adults might be more critical towards gamification (Thiel et al., 2016). Besides this, the participants suggested allowing the user to choose between different avatars (e.g., other animals or mythical creatures, raising a tree, or building a house). These avatars could move through different scenes and further reflect the user's behaviors through their actions in addition to their appearance (King et al., 2013). Further personalization could increase the relatedness between the user and his or her avatar and thus lead to the user putting even more effort into achieving the app's health recommendations (van Vugt et al., 2009).

Another aspect that revealed need for improvement was the app's shop. Most participants reported that they did not know what items to buy, as they were unsure what their avatar needed. This shows that the shop should support the overall goal of the app (Anderson et al., 2016a). In response, the app's shop was adjusted. It now includes items that match the app's context, such as books on stress management, items that are linked to the app's diary categories (e.g., running shoes) or stress management exercises (e.g., chocolate for euthymic methods). Each item now also includes a small description of its purpose.

Though the badges were received positively, the users would like to know from the beginning that these rewards exist and how they can be obtained (Zichermann and Cunningham, 2011). Even though unpredictability is said to play an important role in keeping the user's interest (Chou, 2015), this shows that a certain degree of transparency is required. In response, progress bars and a short description for each badge were added to Stress-Mentor.

The notification users could set to remind them to use the app was also well received. Nevertheless, additional reminders (especially for the tasks) were requested (Ruzic and Sanford, 2018). Still, app designers should be careful so the app's reminders are not perceived as irritating. This can lead to the users uninstalling the app (Dennison et al., 2013). If health apps include more than one aspect that requires the user's attention, individualized, easily set, and turned off notifications or alarms should be included.

Overall, the participants from both groups reported that they liked that the app provided them with a large variety of different stress management techniques. However, they thought that the exercises they could choose from were sometimes too long to practice in their current situation. This could pose a problem, as user's display very little patience for using health app's if they are perceived as time consuming (Dennison et al., 2013). This could be avoided by making sure the app always provides tasks of varying lengths. However, this could result in the users' only accomplishing short tasks, even though lengthier exercises might be more suitable for their symptoms (Davis et al., 2008). Furthermore, participants liked that they could repeat the tasks they previously accomplished as often as they wanted. However, such a feature should be communicated in an obvious manner. This further supports that transparency might be preferred over unpredictability in the context of health apps (Chou, 2015).

Besides the app's content, data security gains importance to users of health technologies (Wilkowska and Ziefle, 2012). Though Stress-Mentor saves the user's data locally on his or her smartphone in encrypted form, 58 % (33 of 57) of the users would have liked to be able to set a password to restrict the access to the app. In addition to local storage and encryption, app developers should, therefore, provide the user the option to set a password, as forcing the user to set a password may dissuade them from using the app (Dennison et al., 2013).

4.4.4 Limitations

This study only focused on the usage of Stress-Mentor over four weeks. Though the results show that gamification can positively influence the usage behavior, this study does not provide insight into how long this effect might last. Longer studies should further investigate the long-term effects of gamification on usage consistency.

Moreover, though usage consistency is linked with effectivity (Vandelanotte et al., 2007), future studies should research whether gamification in the context of stress management does improve the effectivity of an intervention. The present study only looked at the number of performed tasks and time spent with the app, thus no assumptions can be made on whether the app is actually effective in reducing stress. Previous studies have shown positive, as well as neutral and negative effects of gamification on cognitive and behavioral outcomes (Johnson et al., 2016). Future studies should, therefore, also focus on investigating if gamified stress management apps are effective in reducing stress and increasing the use of evidence-based coping strategies.

4.5 Conclusions

Though previous studies suggested that the gamification of stress management apps might not be desired by users (Ahtinen et al., 2013), this study clearly shows that this is not necessarily true. Notwithstanding the identified areas of improvement, Stress-Mentor was evaluated positively with respect to usability, user experience, and quality. While the questionnaires did not show an impact of gamification on the app's hedonic quality or perceived engagement, it did affect the app's usage behavior. Gamification resulted in the accomplishment of more stress management tasks and more time spent using the app. This shows that gamification can have positive effects on the usage of mHealth interventions. In line with previous results (e.g., Taylor et al., 2019), this suggests that gamification increases user compliance. These outcomes are further supported by the users' positive perception of Stress-Mentor's gamified concept. Overall, this confirms that if gamification is designed to support the content and overall goal of an intervention, it is well received and effective in promoting the usage of stress management apps.

Chapter 5

Clinical Outlook: "Pain-Mentor" – Applying the Concept of "Stress-Mentor" in the Context of Chronic Pain Management

Though Stress-Mentor was designed specifically to prevent chronic stress in healthy adults, its gamified concept can be applied in other health contexts. This chapter shows how the app was modified for the application in the context of chronic pain management, resulting in the pain management app Pain-Mentor. The acceptance of Pain-Mentor was then evaluated in an expert study, by assessing the app's quality in the eyes of experts with a background in chonic pain management. In addition, the experts' feedback was used to determine areas that require improvement and to further adapt the app's content to the context of chronic pain management.

5.1 Introduction

5.1.1 Background

Besides chronic stress, chronic pain is a major factor that impacts the health and well-being of today's society. Approximately 1/3 of the American and European population suffers from chronic pain (Breivik et al., 2006; American Psychological Association, 2013). This makes chronic pain a major health care problem that needs to be taken more seriously (Breivik et al., 2006). With profound negative consequences regarding psychological, social, physical, and economic aspects for those affected, chronic pain can have a serious negative impact on a person's overall quality of life (Becker et al., 1997; Carmona, 2001; Pérez et al., 2017; Molander et al., 2018).

Next to medical treatments (e.g., medication, surgery, rehabilitation, and physical therapy), psychological treatments are an important aspect of pain management (American Psychological Association, 2013). In fact, the combination of five theory-based functionalities, namely pain-related education, self-monitoring, goal setting, social support, and the training of self-care strategies, have been suggested to promote self-management of chronic pain (Wantland et al., 2004; Murray et al., 2005; Stinson et al., 2013; Stinson et al., 2014; Alexander and Joshi, 2016). Because patients need to understand and manage the thoughts, emotions, and behaviors that often accompany chronic pain (American Psychological Association, 2013), the mediated self-care strategies should include stress coping skills such as relaxation techniques, problem solving, and communication skills training (Jensen et al., 2003; Macea et al., 2010; Irvine et al., 2015). Multimodal approaches that integrate these aspects can better improve overall quality of life in chronic pain patients when compared to treatments that are strictly medication focused (Flor et al., 1992; Elgar and McGrath, 2003; Barlow and Ellard, 2006).

The integration of multimodal approaches into routine primary and tertiary care has, however, been slow (Breivik et al., 2006). Major barriers, such as poor accessibility due to geographical reasons, limited availability of trained professionals, and large therapy-related costs, keep patients from accessing pain-specific education and psychological treatment (Peng et al., 2007a; Peng et al., 2007b; Stinson et al., 2013). As a result, the majority of patients never receive the required education or skills training to promote pain self-management (Peng et al., 2007a; Peng et al., 2007b; Lynch et al., 2008).

However, the care for chronic pain is no longer strictly limited to medical environments and clinician-guided telehealth due to the rising number of mHealth products (Demiris et al., 2008). mHealth describes the use of mobile technologies in order to improve health (Harrison et al., 2011) by affecting the user's education, motivation, and adherence (Handel, 2011; Ahtinen et al., 2013). It has already been applied to support mental, as well as physical, health programs (World Health Organization, 2011). As such, mHealth can enhance the self-management of chronic conditions (Anderson et al., 2016b). Indeed, preliminary evidence suggests that pain management apps have great potential to support chronic pain treatment and are well received by patients (Irvine et al., 2015; Jamison et al., 2018). Because such apps are available anywhere and anytime (Blackburne et al., 2016), they can reduce the frequency and cost of face-to-face interventions. As a result, they have the potential to make healthcare systems more effective (Institute for Healthcare Informatics, 2013).

In order to ensure their effectiveness, apps for chronic pain management must be based on evidence-based content (i.e., pain-related education, self-monitoring, goal-setting, social support, and the training of self-care strategies including stress management) (Morris et al., 2010; Harrison et al., 2011; Chittaro and Sioni, 2014). Even though these aspects are easy enough to implement, app reviews show that existing pain management apps have limited content. Rather than providing evidence-based behavior change programs, the reviewed pain apps lack the combination of evidence-based functionalities (Rosser and Eccleston, 2011; Wallace and Dhingra, 2014; Lalloo et al., 2015). Indeed, most apps only include one of the five suggested components (Rosser and Eccleston, 2011; Wallace and Dhingra, 2014). Apps mostly focus on supplying information (Rosser and Eccleston, 2011; Wallace and Dhingra, 2014; Lalloo et al., 2015). They seldom help to achieve social support and often lack evidence-based self-care skills and the tracking of the multidimensional experience of pain. Most apps only allow the tracking of pain intensity (Rosser and Eccleston, 2011; De La Vega and Miró, 2014; Lalloo et al., 2015; Portelli and Eldred, 2016) (e.g., FitBack, Irvine et al., 2015). Some apps also allow the user to track other pain related aspects such as pain location, medication, and pain source (e.g., Pain Squat, Stinson et al., 2013). However, only a minority of apps also allow the assessment of emotional and cognitive aspects. In addition, the educational content that is included is often of poor quality. An exception is the pain diary app PainTracker, which includes three of the five suggested functionalities. It allows the tracking of a number of pain related aspects, as well as goal setting and provides informational content (Jamison et al., 2018). Notwithstanding this exception, comprehensive, evidence-based, and clinically informed smartphone apps for pain self-management are highly needed (Lalloo et al., 2015).

While important, the use of evidence-based content alone has been considered as insufficient to ensure adequate user engagement and motivation (Vandelanotte et al., 2007), two aspects that have great influence on the usage of an intervention program (Webber et al., 2010). In fact, further improvement is needed to make pain apps more engaging and entertaining (Jamison et al., 2018).

One way to increase user engagement and motivation is gamification. Gamification, the use of game elements in nongame contexts (Deterding et al., 2011), aims to make interventions such as health apps more enjoyable, motivating, and engaging (AlMarshedi et al., 2016). However, the use of gamification in health apps has been critically discussed, as its effects depend on the context and goal of the application (Johnson et al., 2016). While users do not always want the implementation of gamification in health apps (Ahtinen et al., 2013), it could be shown that its use in health apps can have positive effects on both health and wellness (Johnson et al., 2016).

Indeed, gamification has already been shown to positively influence user self-management (AlMarshedi et al., 2016), lifestyle (Kamel Boulos et al., 2015), health and behavioral outcomes (Allam et al., 2015), as well the retention of desired user behaviors (Kuo and Chuang, 2016). This confirms that the implementation of gamification can be effective in promoting behavior change through apps (González et al., 2016).

Even so, few pain apps make use of this concept. Two pain apps for adolescent cancer patients have included gamification in the form of a virtual rewards system and ranks that were linked to the users' pain diaries and adherence (Stinson et al., 2013; Jibb et al., 2018). However, so far there is no chronic pain management app with an extensive gamification framework for adults.

5.1.2 "Pain-Mentor"

Because current pain management apps lack in their use of both evidence-based self-management skills (Lalloo et al., 2015) and gamification (Jamison et al., 2018), Pain-Mentor was developed to close this gap in research. Pain-Mentor is based on the concept of the gamified stress management app Stress-Mentor (see Chapter 2.4 or Christmann et al. (2018a) for details regarding the app's concept). Stress-Mentor includes all five suggested theory-based functionalities. For one, it realizes self-monitoring though a diary that allows the tracking of up to 14 diary categories (i.e., sleep duration and quality, sport duration and intensity, daily uplifts and daily hassles, stress level, mood, digestion, consumption of water, fruit/vegetables, coffee and alcohol, as well as step-count). In addition, the app teaches different self-help skills through daily and weekly tasks. In these tasks, the user can choose one out of three suggested skills he or she wants to practice. Which techniques are offered by the app depends on the user's entries into a stress checklist. In the stress checklist, the user can enter stress-related aspects (i.e., fears and worries, sadness, anger, stress at work, stress in private life, muscle tension, head, neck and back grievances caused by tension, digestive problems, and sleep problems) on a scale from 0 to 10 on a weekly basis. This concept of personalized tasks encourages the user to set daily and weekly goals and supports the repetition of exercises. The mediated skills include relaxation exercises (i.e., abdominal breathing, meditation, mindfulness, progressive muscle relaxation, guided imagery, stretching exercises, and self-massage), problem solving (i.e., time management, goal setting, planning of social support and social change, and barrier identification), cognitive aspects (i.e., assertiveness training, refuting irrational ideas, appraisal of stress, and stressful situations, and avoiding perfectionism) and the transfer of educational information about stress. Moreover, the app provides tips on stress, based on the user's documented stress-level.

In addition to self-monitoring, stress management skills, and educational information, Stress-Mentor includes several other behavior change techniques (see Appendix 18 for a list) to support long-term behavior change. The included behavior change techniques are linked to an extensive gamification concept aimed at motivating and engaging the user (Christmann et al., 2018a). As such, the app includes an avatar (a bird-like cartoon animal) that provides feedback by reflecting both the user's diary entries (vicarious reinforcement) (Bandura et al., 1963; Christmann et al., 2018b) and progress. Another aspect is the app's agent (a wise owl), who poses as a mentor that entrusts the care of the avatar to the user via a behavioral contract, provides instructions on app functions, and provides general encouragement, as well as the educational tips on stress. The user can collect "woodland coins" that can later be exchanged for items for the user's avatar. Moreover, the app provides feedback on the user's performance through progress bars, a diary overview diagram (Christmann et al., 2017b), badges, and the visual development of the avatar and its surroundings. A detailed description of the implemented behavior change techniques and how they were linked with gamification can be found in Chapter 2.4.

Because stress management, as well as cognitive and behavioral aspects, play an important role in the treatment of chronic pain (American Psychological Association, 2013), Stress-Mentor's concept was adopted for the pain management app (Pain-Mentor) that was evaluated in this study. While all stress related content remained, additions were made to the existing diary, tips, stress checklist, and daily tasks to further adapt Pain-Mentor to the context of chronic pain treatment. The following adjustments were made to better suit Pain-Mentor to its designated usage context: For one, the stress checklist was renamed to symptoms checklist. The symptoms checklist was then extended with a numerical rating scale for pain that is commonly used in therapy. It allows the user to enter his or her pain level on a scale from 0 (no pain) to 10 (worst possible pain) (Farrar et al., 2001). The diary was also extended by this scale. This provides the user with the opportunity to track the trend of his or her pain on a daily basis (Hawker et al., 2011). Eight pain specific daily tasks were added to the task pool: one each to develop a plan for setbacks, planning social support for pain management and planning a dropped activity, as well as five physical exercises for muscle strengthening and stretching. Moreover, the tips given by the app's mentor were extended by additional information on chronic pain and pain management. A screenshot of the app is displayed in Figure 29.

All in all, Pain-Mentor differs from other pain management apps regarding one important property: It includes all five suggested self-management functionalities (i.e., educational information on pain and stress, a total of 87 pain and stress specific self-help skills, goal setting through tasks of the day/week, multidimensional self-monitoring through a digital diary, and social support through tailored exercises) and combines this content with gamification in order to motivate and engage the user. This poses great potential for supporting in-person therapy and reducing therapy costs (Bender et al., 2011).

Figure 29: Screenshot of Pain-Mentor.

5.1.3 Motivation

In contrast to general recommendation, most app designers have neither included experts from pain management in the development of their apps (Portelli and Eldred, 2016), nor have they used expert reviews to assess their apps' quality (Rosser and Eccleston, 2011). Contrary to this, Pain-Mentor's contents were extended in consultation with a physician who is specialized in chronic pain treatment. The involvement of health care professionals in the development of health apps, as was done for Pain-Mentor, is important for health apps to contribute value to health care and chronic disease management (Kao and Liebovitz, 2017). With many apps promising effective pain treatment, patients face a large array of possible apps to choose from, with little guidance regarding their quality (Institute for Healthcare Informatics, 2013). To ensure quality, functionality, and relevance of the content, health apps need to be tested in scientific trials and involve health care professionals, not only during their development, but also in the evaluation process (Rivera et al., 2016).

In general, health apps must be acceptable to both the user, who must decide whether the program is usable and can provide benefit in an operational environment, and health professionals, who determine whether the app does what it is supposed to do (Stead et al., 1994). In contrast to the users' goal, the experts' primary goal is to assess the quality of a health app in order to identify apps to recommend to their patients (Baptista et al., 2017). They focus on different aspects and provide different feedback than users and developers (Wang et al., 2017). Therefore, even though few pain management apps have been scientifically evaluated by experts (Lalloo et al., 2015), testing the quality of health apps through expert evaluations is essential to assess key app features (Gibbons et al., 2018).

This study, therefore, conducted an expert evaluation of the newly developed pain management app Pain-Mentor that combines a multimodal approach to pain self-management with an extensive gamification framework. The aim was to evaluate the app's general quality in the eyes of health professionals and to assess their opinion on combining pain management with gamification.

5.2 Methods

5.2.1 Participants

To assess the app's quality in the eyes of health professionals, experts with a background in chronic pain management were recruited (Figure 30). For this purpose, based on an internet search, physicians that are specialized in pain treatment and general psychotherapists in a 100 km radius of the Technische Universität Kaiserslautern were contacted via email. Of 94 contacts, 8 were willing to participate in this study. An additional three experts learned about the study from one of their colleagues and volunteered to participate as a result. In the end N = 11 experts (5 physicians and 1 nurse with a background in pain management, as well as 5 psychotherapists) participated in this study. Previous research suggested the use of at least 2-4 experts (Nielsen, 1994; Barnum, 2011; Masterson Creber et al., 2016), though a larger sample size increases the percentage of identified problems in the apps. Because as many as N = 9 participants enable the identification of about 95% of all problems (Virzi, 1992), it can be concluded that the sample size of this study provides a good insight into the app's quality and enables to identify most of the concerns arising from experts for chronic pain treatment.

All participants had specific experience in the field of pain and pain management. They had neither prior experience with the tested app, nor comprehensive experience in handling and testing mobile devices.

5.2.2 Study Procedure

The whole procedure was approved by the local ethics committee from the department of social sciences of the Technische Universität Kaiserslautern. To ensure a standardized approach the following study procedure was predefined (see Figure 30). At the beginning, the health professional was informed about the procedure, aim and data collected in the study. Each participant gave written consent according to the declaration of Helsinki.

Afterwards, Pain-Mentor and its contents were presented to the professional in detail in a PowerPoint presentation that explained which pain and stress management methods, behavior change techniques, and gamification aspects were included and how they were interconnected. After the presentation, the expert tested the app on a tablet (Lenovo TB-4706F). For this purpose, the app was set to a specific default setting in order to ensure that all participants were exposed to the same content and features. Testing also followed a predefined process that took 15-25 minutes for each participant (see Figure 31 for a detailed description).

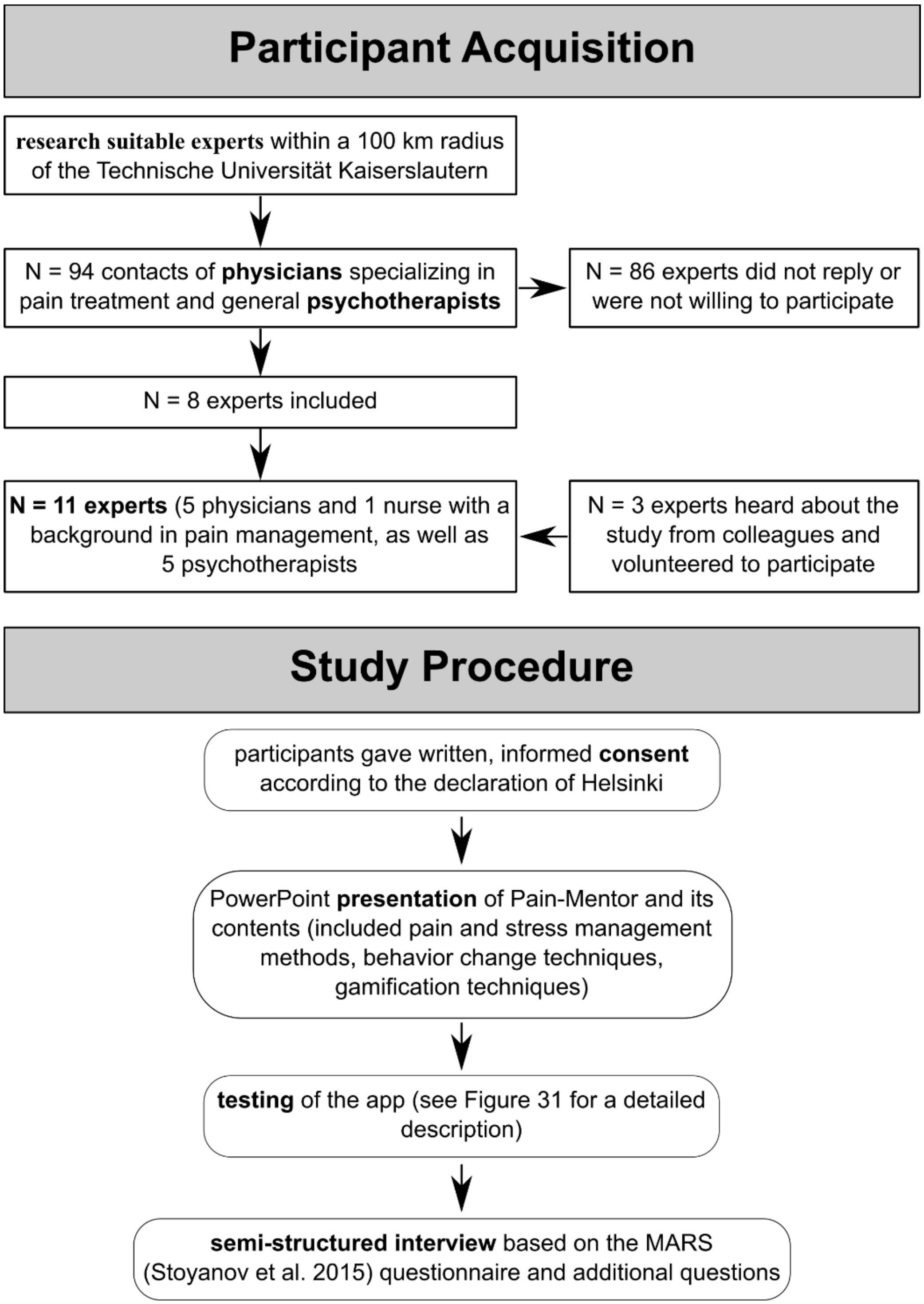

Figure 30: Depiction of the participant acquisition and study procedure.

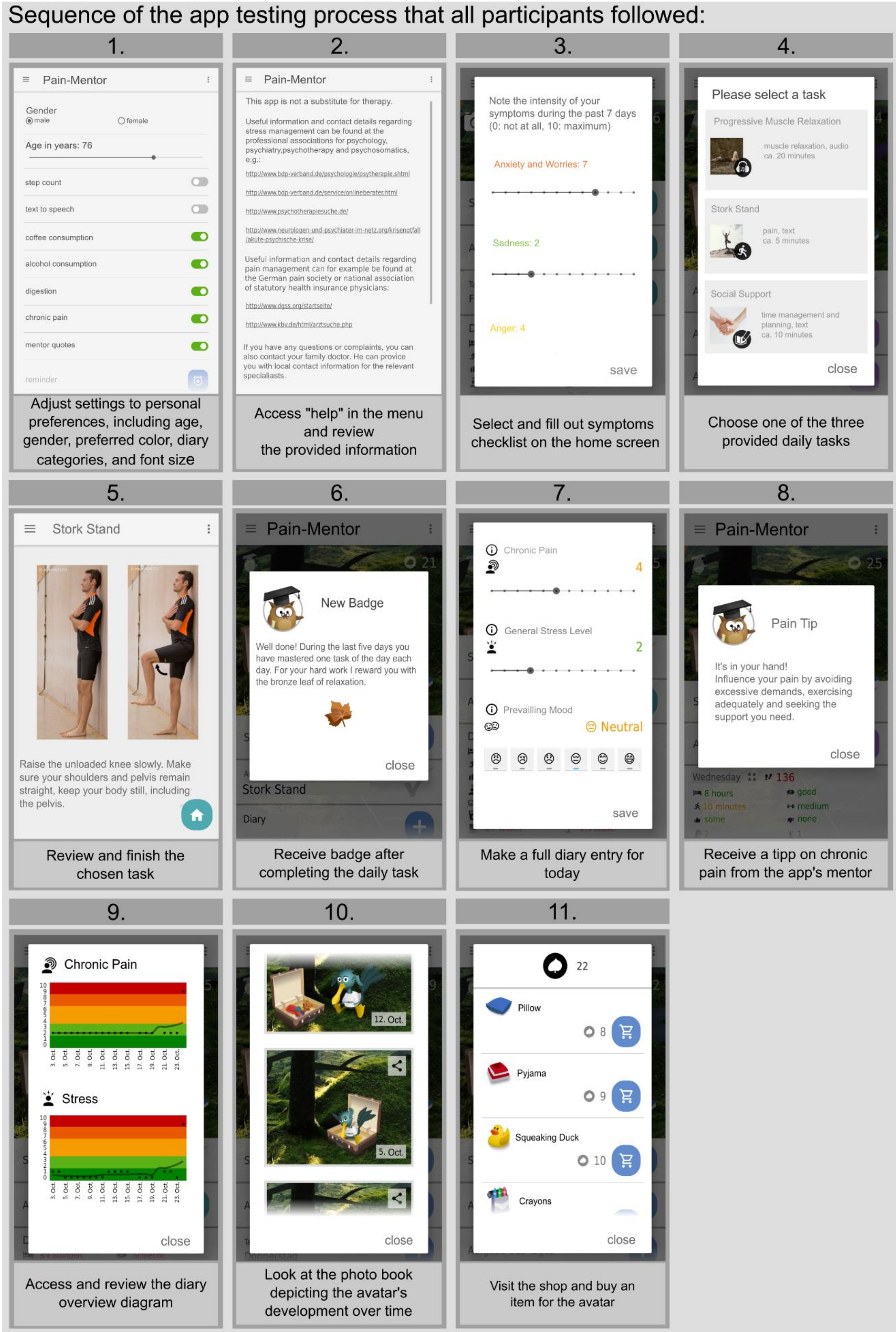

Figure 31: Sequential order of the app testing process that all experts followed.

5.2.3 App Quality

After testing Pain-Mentor, the expert rated the quality of the app and was asked for feedback. For this purpose, the mobile application rating scale (MARS) was applied. MARS was specifically designed to assess the quality of health apps, with the help of experts from health and IT (Stoyanov et al., 2015). MARS consists of six subscales. Four of those scales (engagement, functionality, aesthetics, and information quality) assess the general app quality. The subjective quality section evaluates the user's overall satisfaction, while the app specific section assesses the perceived impact of the app on the user's knowledge, attitudes, intentions to change, as well as the likelihood of actual change in the targeted health behavior.

Participants rated each of the 23 MARS items on a 5-point Likert-scale (from 1 = inadequate to 5 = excellent). To allow for differentiated user feedback, MARS was applied as a semi-structured interview. This means that after each rating, participants had the opportunity to explain their answer and give suggestions regarding further improvement of the app (open response format). Presenting MARS and other questionnaires as semi-structured interviews has been done in previous studies and promises deeper insight into the raters reasoning and possible improvements (e.g., Anderson et al., 2016a). As the experts only spent a limited time testing the app (approx. 20-30 minutes), an additional answer option was added to each question, namely "I cannot assess this". This ensured that experts were not forced to answer if they felt that they had not spent enough time with the app to assess an aspect.

Three questions from the subscale "subjective quality" were removed from the questionnaire: 18. "Does the app come from a credible source?" because the source of the app was explained to the participants in detail, 19. "Has the app been tested?" because an evaluation regarding the app's effectiveness has not been conducted so far and 22. "Would you pay for this app?" because health experts are not the target user audience of the app. In addition, question 20. was adapted to the context and changed into 20. "Would you recommend this app to patients who might benefit from it?" The questions from the "app specific" section were adapted to the context, i.e. the term "health behavior" was replaced with "stress and pain management".

5.2.4 Additional App Specific Questions

In addition to MARS, the participants answered seven additional questions on a 5-point Likert-scale (from 1 = inadequate to 5 = excellent) and one open response question with regard to the app's expected appeal for patients and specific app features. As with MARS, participants received the opportunity to explain their answers. See Appendix 21 for a list of all additional questions.

5.3 Results

5.3.1 Quality of "Pain-Mentor"

Overall, experts rated Pain-Mentor to be of excellent quality (M = 4.51, SD = 0.54). The subjective app quality was appraised as excellent (M = 4.51, SD = 0.31). The app specific questions were rated as good with a mean of 4.27 (SD = 0.76). The MARS quality subsections engagement and aesthetics were rated as good, functionality and information were rated as excellent (see Appendix 22 for mean values and standard deviations).

Looking at each question of these subsections in detail, the results showed means ranging from 3.80 (SD = 0.63) for customization to 4.73 for entertainment (SD = 0.647) and interest (SD = 0.47) for the subsection engagement. Functionality showed mean ratings between 4.55 (SD = 0.69) for ease of use and 4.83 (SD = 0.41) for gestural design. The app's aesthetics were assessed to be good, with mean values of 4.36 for layout (SD = 0.65) and visual appeal (SD = 0.51), as well as 4.64 (SD = 0.51) for graphics. The information communicated in Pain-Mentor also showed good to excellent ratings with means ranging from 4.09 (SD = 0.70) for the quality of the information to 4.90 (SD = 0.32) for information quantity. The subjective quality of the app showed ratings of 4.27 (SD = 0.47) regarding the recommendation of the app to patients and the overall star rating and 5.00 (SD = 0.00) for usage duration. The app specific questions of MARS showed values between 3.67 (SD = 0.92) with regard to encouraging patients to seek help outside the app and 4.8 (SD = 0.45) regarding the app's potential to promote behavior change. A detailed visualization of the results is included in Figure 32.

Occasionally, experts did not feel able to answer a question (see Appendix 23 for details). Participants gave several reasons for being unable to assess these questions, which will be discussed.

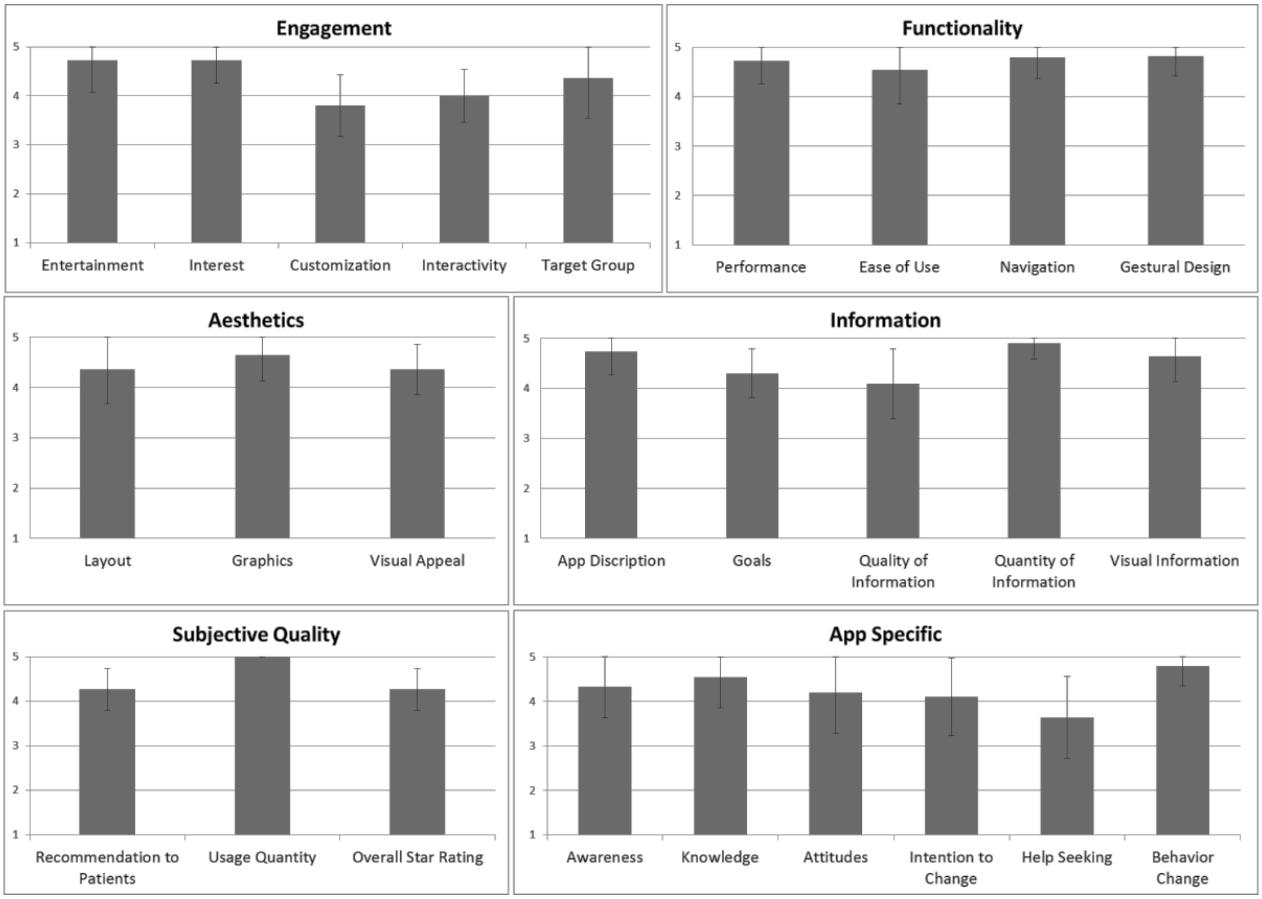

Figure 32: Experts mean ratings with standard deviations for the MARS (Mobile Application Rating Scale) regarding Pain-Mentor.

5.3.2 Additional Questions Outcomes

Good to excellent mean ratings could be observed for all additional questions (see Table 17). Experts thought it was very likely that Pain-Mentor would appeal to patients (*M* = 4.45, *SD* = 0.46). The app's gamification concept was rated as good (*M* = 4.27, *SD* = 0.65). The experts' expectations of Pain-Mentor were mostly fulfilled (*M* = 4.18, *SD* = 0.60). They rated the implemented diary (*M* = 4.64, *SD* = 0.67), daily tasks (*M* = 4.64, *SD* = 0.67), and symptoms checklist (*M* = 4.73, *SD* = 0.47) as sensible and useful. Moreover, they believed that using Pain-Mentor to support in-person therapy would be beneficial (*M* = 4.55, *SD* = 0.52). No expert was unable to assess an additional question.

5.4 Discussion

5.4.1 Principal Results

Overall, experts rated the app to be of excellent overall and subjective quality. The app specific section of MARS was rated as good. These positive results reflect that the app and its contents (diary, daily tasks, information, symptoms checklist, as well as its gamification concept) met the experts' expectations. As anticipated, these results show that combining the five suggested self-management functionalities (i.e., educational information, self-care skills, self-monitoring, goal setting, and social support) (Wantland et al., 2004; Murray et al., 2005; Stinson et al., 2013; Stinson et al., 2014; Alexander and Joshi, 2016) in an app targeting chronic pain management is approved by health experts.

Table 17: Experts' mean (*M*) ratings and standard deviations (*SD*) for the additional questions regarding Pain-Mentor.

Question	*M*	*SD*
Do you think the app would appeal to patients?	4.45	0.52
Did you like the gamification concept?	4.27	0.65
Did the app meet your expectations?	4.18	0,60
How useful is the diary?	4.64	0.67
How useful is the symptoms checklist?	4.64	0.67
How useful is the concept of the daily exercises?	4.73	0.47
How useful is it to apply the app in addition to therapy?	4.55	0.52

In addition, the experts thought that the app's gamification concept, especially the avatar, made the program engaging and interesting to use. This is mirrored by the experts' comments, e.g. E4: *"I think the app motivates those patients who want to be proactive and do something to manage their pain."* This implies that the use of gamification could pose a solution to the lack of engagement and entertainment in pain management apps mentioned by Jamison et al. (2018).

However, two experts expressed concerns that the existing gamification concept, especially the choice in avatar, might be too childlike and not suitable for the elderly. While most users prefer human avatars that match their own gender (Nowak and Rauh, 2005), it is very subjective which avatar appeals to a user. To solve this problem, the participants suggested providing the user the choice between different avatars. This reflects the comments from the user study with Stress-Mentor that was presented in Chapter 4. Some experts expressed skepticism regarding the avatar's suitability for the elderly. Chapter 4 implicates that these concerns might be unwarranted, as older users were even more compliant than younger ones when testing a gamified stress management app. This agrees with the health professionals' opinion that Pain-Mentor was well suited for the target group (adults with chronic pain) and that it was very likely that the app would appeal to patients. This further supports the combination of evidence-based content and gamification (Allam et al., 2015; Kamel Boulos et al., 2015; AlMarshedi et al., 2016; Kuo and Chuang, 2016).

In addition, gamification most likely had positive impact on the app's aesthetics. Experts especially liked that they could choose the avatar's color and that the avatar's appearance was linked to both the diary entries and progress (e.g., *E9: "It's nice that the user can pick the avatar's color. Especially the visual elements invite you to explore and play around a little.")* and the app's "simple visual design" that allowed them to use the app intuitively. This supports the effects of both personalization (Gay and Leijdekkers, 2012) and the use of vicarious reinforcement through avatars in health apps (Christmann et al., 2018b).

While the experts were mostly satisfied with Pain-Mentor's customizability, they also suggested including more reminders to help the user remember to practice throughout the day. This shows that one reminder is the minimum, while the inclusion of more appears to be preferable. Similar comments were encountered by Stress-Mentor's users. Five experts also suggested the addition of new diary aspects. However, there was no consensus between these experts on which aspects should be added (they suggested weight, additional dietary aspects and notation of additional exercises). Thus, it cannot be concluded which categories would make the most sense to add based on the feedback gained from this study. Besides, the addition of aspects could easily overwhelm the user and lead to a decline in the app's usability and usage (Thompson et al., 2005). One expert also commented that not all aspects are equally important for every patient. To solve this issue, developers could add a notes section that leaves room for further patient specific entries, as was implemented in other pain management apps (e.g., Chronic Pain Tracker, FitBack) (Rosser and Eccleston, 2011; Irvine et al., 2015). Another approach could be to allow the addition of individual scales in the diary (Koskinen and Salminen, 2007). Both solutions would leave the addition of further diary categories up to the user and his or her treating health professional.

Regardless of the suggested extensions, the diary was perceived as very useful by the experts. One participant expressly mentioned that she especially liked that the diary wasn't overly focused on pain, but rather allowed tracking the patient's overall well-being, including emotional (stress level and mood) and cognitive aspects (daily hassles and daily uplifts). This goes in line with Rosser and Eccleston (2011) and Lalloo et al. (2015), who criticized that most pain apps focused only on tracking medication and pain levels.

Though the app's interactivity was assessed as good, one expert commented that it could be further improved by giving advice based on the user's diary entries. Consequently, developers should think about further personalizing their apps through linking user entries (e.g., from a diary) to suitable health information and tasks. For example, the app MyBehavior automatically provides personalized suggestions based on a health diary (Rabbi et al., 2015).

Participants thought that the information imparted in the app was generally well formulated and of high quality. Moreover, the experts would recommend the app to many of their patients. This shows that the experts thought the app to be a good therapy supplement. Whether they would recommend the app to a patient depended on whether he or she would profit from using the app. This depends largely on patients' age, disease pattern, and current state. However, age does not necessarily impact compliance or satisfaction with a pain app (Jamison et al., 2018). Moreover, the combination of visual features (due to the gamification concept) and content (Christmann et al., 2018a) was perceived very positively (e.g., *E9: "It (the app) has a good balance of simple visual design and good content-related information.")*. This further supports the use of gamification in the context of pain management.

There was general approval of the self-management skills that are imparted in the app. Still, three experts suggested including additional tasks, such as more stretching exercises, more tasks specifically aimed at dealing with pain, and exercises aimed at distracting patients from their pain. This emphasizes that experts see the potential of using apps to teach the user a large number of stress-related self-help skills, including relaxation, problem solving, and cognitive aspects (Macea et al., 2010). However, it should be supplemented by more pain specific aspects to provide maximal suitability.

The experts also mentioned that they would like to be able to review the app's data with their patients on a computer in order to monitor and discuss their patients' progress. However, an automatic data transmission to the treating health professionals was seen as problematic due to the experts' limited available time. To avoid this problem, an optional sharing function could be added that allows patients to share their data with health professionals on a voluntary basis (Rosser and Eccleston, 2011). Such functions provide health professionals with the opportunity to gather data on a patient's behavior (Luxton et al., 2011). My Pain Diary, for example, offers the export of data to a computer (Jamison et al., 2018).

The experts' positive ratings regarding the app's ability to positively influence patients' awareness, knowledge, attitude, intention to change, help seeking, and behavior change align with the fact that health apps can enhance users' self-management of chronic conditions (Anderson et al., 2016b). However, it was emphasized that the individual played an important role regarding these aspects. Nonetheless, the experts thought that it would be very useful to employ the app to supplement in-person therapy. This supports the great potential of health apps to support regular treatment (World Health Organization, 2011).

In addition to using the app for therapy support, one expert suggested to use Pain-Mentor to bridge the time until patients can receive in-person therapy. Because waiting times for therapy are often long, applying health apps in this manner could further increase their potential to improve health care (Demiris et al., 2008) and diminish the number of patients who do not receive adequate treatment (Lynch et al., 2008). This area of application should therefore be the focus of future research.

When asked which aspects they thought would keep patients from using the app, experts mostly mentioned a lack of motivation for people to change. However, they also brought up a lack of familiarity with mobile devices/apps and data security.

This goes in line with previous studies that have shown that technical affinity (Czaja et al., 2006), data privacy, and security (Wilkowska and Ziefle, 2012) are important aspects for choosing and using health technologies. This aspect was also pointed out as important by participants who tested a neurofeedback-based meditation exercise with the wearable EEG MUSE and its companion app (see Chapter 2.3). Developers should, therefore, make sure to pay special attention to data security when developing health apps (Luxton et al., 2011).

While the experts' comments showed that some adjustments, such as adding more pain specific exercises, diary categories, and reminders could further improve the app, they still rated Pain-Mentor to be of excellent overall quality. This shows that minor adjustments of suitable health apps (e.g., from stress management) can make them useful tools that can be applied in different contexts (e.g., chronic pain management). Nonetheless, whether an app will be assessed as useful or not depends on both the app's content and the suggested context of use (Lewis and Wyatt, 2014). Not every health app should be applied in or adjusted for other contexts. Underlining the importance of expert evaluations (Rivera et al., 2016), this study showed that involving health experts who have a background in the target context helps to determine an app's suitability and to identify necessary adjustments.

5.4.2 Limitations

The experts approved of the app's gamification concept and thought that using gamification in this manner could improve patients' motivation and engagement. However, from this study it cannot be concluded how the use of gamification does, in fact, affect these aspects. Although Chapter 4 indicates that the app's gamified concept can positively affect user compliance, further research is needed in order to determine its effects in the context of chronic pain treatment. Moreover, while experts approved of the idea to use the app to support therapy, randomized controlled trials are needed to actually determine Pain-Mentor's effectiveness as a therapy-support tool.

Because all experts were given a detailed introduction into Pain-Mentor and spent approximately 20-30 minutes using the app, they received detailed insight into how the app worked. Nonetheless, experts had trouble assessing the app's gestural design and the app's potential to positively affect behavior change. Future studies could avoid this problem through longer trial periods.

5.5 Conclusions

This study supports the use of gamification in a pain management app. The participating health experts approved of the app's gamification concept and described it as a good way to enhance user motivation and engagement. Moreover, the app received positive ratings with regard to general and subjective quality, as well as app specific aspects. This shows that the use of gamification did not have a negative impact on the app's credibility and integrity and that the combination of gamification with the five recommended self-management functionalities (i.e., pain-related education, self-monitoring, goal setting, social support, and the training of self-care strategies) led to an overall positive evaluation of Pain-Mentor. This study also shows that applying MARS, in combination with additional, more app specific questions in a semi-structured interview, can provide insight into user ratings and disclose possible areas for improvement. Hence, this approach helps to adjust health apps for a specific target audience and to identify further application scenarios.

Moreover, the study showed that Stress-Mentor's concept is not limited to an application in the stress management context. It can, in fact, be successfully applied in other mHealth contexts that are aimed at supporting mental and physical health. The behavior change techniques which are associated with the agent, the avatar, the photo book, the experience points, the acquisition of items for the avatar, and the badges can be transferred into many other health contexts. Furthermore, because their content can be adapted to the respective context, the "task of the day" and "task of the week" features provide a universal tool to teach self-management skills and knowledge with respect to different health behaviors. Stress-Mentor's concept can therefore be useful for a broad range of interventions, including, but not limited to, chronic pain management.

Chapter 6

General Discussion

6.1 Principal Results

As mentioned in the beginning, the gamification of stress management apps is a critically discussed topic. Previous research suggested that users of stress management apps might be opposed to the use of gamification in this context (Ahtinen et al., 2013). However, it remained unclear how gamification is actually perceived by the users. Furthermore, it was unknown how gamification affects the usage behavior of gamified stress management apps (Johnson et al., 2016). To fill this gap in research, the focus of this thesis was put on the iterative development and evaluation of the gamified stress management app Stress-Mentor. Such user-driven development approaches help to ensure qualities such as the app's acceptance and usability, both of which are important correlates of an app's usage intention and social impact (Venkatesh and Bala, 2008).

In Chapter 1, several requirements for a gamified stress management app were identified. One was that it should combine evidence-based content from behavior change theory, stress management theory, and gamification in order to ensure the users' adherence and long-term effectiveness (e.g., Davis et al., 2008; González et al., 2016; Payne et al., 2016). In a first approach to meet this demand, a MVP was developed that included a health diary for self-monitoring and vicarious reinforcement through an avatar (Christmann et al., 2018b). The feedback from a user study with this MVP was used to improve the app, including the integration of a diary overview chart (Christmann et al., 2017b), as well as the extension of the gamification framework (see Chapter 2.4 and Christmann et al., 2018a for details). In addition, the participants stated that they wanted instructions on how to perform effective stress management exercises to be included in the app.

This supports the idea that users need detailed instructions on how to perform suitable exercises, because they have trouble to come up with appropriate coping strategies on their own (Kocielnik and Sidorova, 2015). Previous research suggests that multi-technique approaches are more effective than single-technique interventions (Murphy, 1996; Jones and Johnston, 2000). Also, they can help the user to identify the coping strategies that are most suitable to them (Barabasz and Perez, 2007). However, such approaches are still rare in currently available stress management apps (Coulon et al., 2016; Christmann et al., 2017a).

Stress-Mentor closes this gap by including a multi-technique approach that is based on a variety of different audio- and text-based stress management exercises and combines these techniques with aspects from behavior change theory and gamification theory (see Chapter 2.4 for details). This ensures that the three prerequisites for effective stress management presented by Payne et al. (2016) (i.e., predisposition, engagement, and reinforcement) are met. Chapter 3 revealed that this approach led to improvements of Stress-Mentor's full version compared to the MVP, especially regarding its perceived engagement, information quality and quantity, and perceived impact.

In addition to content, users put emphasis on preserving their autonomy (Sax et al., 2018). Stress-Mentor meets this demand with two design features. For one, the task of the day/week feature provides the user with the choice between three different stress management exercises every day. This leaves him or her in control of which task to perform. The second feature is Stress-Mentor's limited usage period by design. Stress-Mentor's concept allows the user to identify which stress management techniques work best for him or her and how the exercises can be integrated into daily life. Moreover, the graded tasks let the user become increasingly independent with regard to performing the exercises. This leads to Stress-Mentor making itself superfluous over time.

Another important user requirement is the personalization of mHealth interventions (Gay and Leijdekkers, 2012). Stress-Mentor satisfies this requirement by adjusting its content to the user's personal information. This includes the modification of the assessment of the diary categories sleep and alcohol consumption according to the user's age and gender, respectively. Moreover, the user can decide which behaviors he or she wants to track in the diary and adjust the apps color scheme to his or her own preference. In accordance with the user's feedback from the conducted studies they can set several reminders and decide whether they want to be presented with motivational quotes. Furthermore, the presentation order of the stress management techniques is based on the user's stress checklist. Though throughout the intervention the user can perform every exercise, the ones that are most relevant to the user's self-reported grievances are presented first. This ensures that the app adjusts to the user's personal demands (Atienza and Patrick, 2011), which can improve the adoption of the presented stress management strategies (Gibbons et al., 2018).

To ensure that the app is widely accessible and reaches as many potential users as possible, Stress-Mentor is available for free. It was previously suggested that health apps should be low cost in order to ensure their user acceptance (Krebs and Duncan, 2015). This was confirmed in the presented studies, which revealed that the participants did not want to pay for Stress-Mentor. The reason they gave for this was that they generally did not pay to use apps. This indicates that free health apps have the potential to be more widely distributed, because free apps are generally downloaded more often (AppBrain, 2019). Another aspect is that the app's content is always available, because once installed, the content's presentation is independent of the internet. However, such design decisions can come at the cost of memory space, which could potentially be perceived as a disadvantage by users (Dennison et al., 2013). Nonetheless, Stress-Mentor only has a size of 104 MB.

The evaluation of Stress-Mentor in user and expert studies showed that the iterative development approach, under consideration of the aforementioned user demands, led to a high quality product that is accepted by users and experts alike (Chapter 3, 4, and 5). Including users as well as experts has been suggested as an important approach, because they provide feedback from different perspectives (Wang et al., 2017).

The results also showed that the concern of the users' negative attitude of gamification in stress management apps (Ahtinen et al., 2013) did not hold true. The users of Stress-Mentor approved of its gamified concept and commented positively on the integrated game elements, especially the avatar. Similarly positive user attitudes were reported for combining gamification with behavior change content in a smoking cessation app (Edwards et al., 2018). This supports previous reports that showed that users are willing to adopt gamification into their lives (Spil et al., 2017). The positive view on the app's concept is mirrored by user comments from the Google Play Store after Stress-Mentor was published:

"Your app stands out from the 'usual' offers and is absolutely brilliant! It is based on a holistic approach and combines many different instruments for self-perception and stress management. In addition, it is attractively designed and cleverly combined with a playful and challenging character.

The app helps me to regularly pay closer attention to my stress management and to keep my personal 'health diary'. And it is also fun to care for and raise the Elwetritsch (as one should care for one's inner child) and to reward oneself with nice gimmicks. I have already recommended the app many times and received only very positive feedback so far."

The gamified concept of the app was further supported by the results of the study presented in Chapter 4, which investigated Stress-Mentor's usage behavior. Though previous research showed that gamification can have positive effects on user compliance of health apps (Taylor et al., 2019), the presented study was the first investigating how gamification affects the usage behavior of stress management apps. The results revealed that gamification increased the amount of time the users spent with the app, as well as the number of performed stress management exercises. This indicates that gamification not only increases the use of a service, it can also increase the exposure to an app's evidence based content (Webber et al., 2010). This goes a long way to ensure the app's effectivity (Cafazzo et al., 2012).

Overall, the applied experimental approach to the iterative development of Stress-Mentor, allowed for the continuous improvement of the app, as well as for adjustments due to user demands and expert suggestions. It also enabled the ability to reveal improvements compared to the MVP (see Chapter 3), to investigate the users' perception of gamification in this context, and to study the effects of gamification on the usage behavior (see Chapter 4). Moreover, the application of the app's gamified concept in the context of chronic pain management tested its applicability in other health contexts (see Chapter 5).

6.2 Ethical, Legal and Social Implications

Notwithstanding Stress-Mentor's positive evaluation, health app developers need to consider which other factors might influence the usage intention of their applications. For example, an online survey with Stress-Mentor suggested that intrinsic motivational factors (e.g., subjective stress level and the subjective importance of stress management) have great impact on both the app's perceived usefulness and usage intention (Faust-Christmann et al., 2018). This shows that people adopt gamified health apps more easily when they already have the right attitude towards the change and then use such apps for additional motivation in order to reinforce the desired behaviors (Alahäivälä and Oinas-Kukkonen, 2016). This results in a so-called "prevention dilemma", because risk groups who are not aware of the importance of stress management for their own health and well-being are not willing to use such an app (Faust-Christmann et al., 2018). Extrinsic motivators, such as loyalty programs from health insurance companies, have been suggested as possible incentives and could lead to the usage of health apps for monetary reasons (Friedel and Trautvetter, 2011). However, this does not appear to apply in the context of stress management apps (Faust-Christmann et al., 2018).

Aside from their potential, health apps also generate new security and privacy concerns (Meingast et al., 2006). These issues pose obstacles to the use of health apps (Rasche et al., 2018). This is not surprising, considering that health data is often highly sensitive. This makes the protection of this data all the more important (Kotz, 2011) in order to ensure the systems acceptability (Guo et al., 2016). This was supported by the comments of users who tested the MUSE companion app (Chapter 2.3.2). The users made it clear that having to disclose their personal information online in order to use the app would keep them from downloading it onto their own device. While only the minority of available health apps ensure the protection of the user's health data (e.g., He et al., 2014), Stress-Mentor meets these demands by only storing the data locally on the smartphone in encrypted form and not passing the user's data on to third parties. This is specifically important when applying the app's gamified concept in medical contexts (Meingast et al., 2006; Albrecht and Fangerau, 2015; Chan et al., 2015).

On the other hand, patient monitoring remains an important factor for the application and recommendation of medical apps. However, while experts wish to review their patients' data in order to monitor their progress, external monitoring also remains an ethical and legal issue, because it is often unclear who has access to the obtained information (Olff, 2015). Also, the professionals fear an overload of work due to time limitations (Gagnon et al., 2016), while the patients need to be in control of the collection, recording, distribution, and access to their mHealth data (Kotz, 2011). Leaving the choice of sharing personal health data up to the user, instead of automatically sharing the data with medical professionals, could solve these issues (Rosser and Eccleston, 2011).

Besides data privacy, other ethical and legal concerns must also be considered when developing and testing health apps in user trials. Depending on the context, medical product laws (MPG, 2019) and general requirements with regard to medical research (World Medical Association, 2013) may have to be met. So far, which codes are followed largely depends on the researchers' profession. However, it is questionable to what extent these codes cover the constantly changing field of mobile technologies. Therefore, guidelines and codes should be developed that are adapted to the specific requirements of mHealth solutions (e.g., Albrecht and Fangerau, 2015).

Because Stress-Mentor strictly targets prevention in healthy users and aims at promoting a healthy lifestyle, it does not fall under the medical product laws in Germany (MPG, 2019). However, these laws apply to medical apps that are, for example, designed to support patients in their dealing with a chronic disease in everyday life. This includes Pain-Mentor, which aims at promoting self-efficacy for chronic pain management patients. If an app has a primarily medical purpose (e.g., it supports the diagnosis or therapy of diseases) it needs to be certified as a medical product. For this, it must undergo a conformity process. This process does not, however, include the testing of the quality of the app's content, its benefit for the user, its usability, or the extent of the app's data security concept (MPG, 2019).

Thus, even though reliability, quality, correctness of the content, and minimal error rate are minimum requirements, there are many apps on the market that do not meet them (De La Vega and Miró, 2014). As a result, medical professionals and patients have difficulty to identify secure, high quality apps (Institute for Healthcare Informatics, 2013; Boudreaux et al., 2014). A standardized evaluation tool could help to solve this issue (Chan et al., 2015). The MARS and uMARS scales (Stoyanov et al., 2015; Stoyanov et al., 2016), which were used to assess Stress-Mentor's quality in this thesis, are one solution approach that is more and more often applied. In addition to providing easily interpretable quality scores, the application of these scales also allowed the observation of improvements between the development stages and helped to identify required adjustments in the presented studies.

6.3 Recommendations for the Development of Gamified Health and Medical Apps

Aside from the requirements discussed in the Principal Results section and ethical, legal, and social aspects, a health apps' accuracy, legitimacy, and security were identified as important factors to ensure their acceptance (Dennison et al., 2013). Thus, developers should ensure that the content of their interventions is evidence-based (Abraham and Michie, 2008; Cafazzo et al., 2012). This includes the utilization of behavior change techniques, gamification techniques, and context specific aspects (Payne et al., 2016). Here, not only the number of techniques that are included is important, but also that the chosen techniques match the app's context (Johnson et al., 2016).

With regard to gamification, for example, several aspects need to be kept in mind. For one, the effects of gamification are diverse. Thus, the choice of techniques might differ depending on the intervention's target group. While leaderboards, badges, and performance graphs positively affect need satisfaction and perceived task meaningfulness, avatars and narrative context can promote social relatedness (Sailer et al., 2017). Moreover, not all game elements are equally well suited depending on their application context. For example, further burdening the user by including time or social pressure is likely counterproductive in the context of stress management (Hoffmann et al., 2017).

Also, in order to ensure the acceptance and usage of interventions, developers should clearly communicate the content and aim of their apps and provide the user with an easy way to identify the quality of the intervention (Chan et al., 2015). This includes research studies targeting the effectiveness and acceptance (e.g., Lee and Jung, 2018), as well as quality ratings of health apps (e.g., Stoyanov et al., 2016). Moreover, developers should not take the user's theoretical opposition to gamification in specific mHealth contexts at face value. Even though previous research indicated that users did not want gamification in stress management apps (Ahtinen et al., 2013), it was well received in the user studies conducted with Stress-Mentor (e.g., Christmann et al. (2018b) and Chapter 4). This shows that, if implemented in a way that supports the overall goal of the app (Anderson et al., 2016a), gamification can be accepted and positively influence the users' compliance.

In line with this, it should be kept in mind that, even though potential consumers often show high interest in linking gamification with mHealth, they are often unaware of the value of gamification in the respective contexts (Thorpe and Roper, 2017). App developers should, therefore, make the functionality and relevance of gamification in health apps much more clear (Spil et al., 2017). This includes the conductance of user studies targeting the effects of gamification on usage behaviors and behavior change (Cafazzo et al., 2012;

González et al., 2016; Taylor et al., 2019). The research presented here is a first step to communicate the effects of gamification in the context of stress management. It shows that the gamification of Stress-Mentor is perceived in a generally positive manner and positively affects the app's usage. One way to inform the user of the uncovered results is by addressing them in the app description in the corresponding app store.

6.4 Recommendations for the Use of Wearables in the Stress Management Context

More and more health apps utilize wearables to achieve their objectives (e.g., Kaushik et al., 2006). The research presented here indicates that the users of Stress-Mentor were not generally opposed to the integration of wearables into the app. In fact, many stated that they want to automatically record their step count with their own smartwatches. This indicates that the usage of multifunctional wearables and ones that are already in the users' possession is likely to be accepted. However, the presented results also revealed that this might not be the case for devices the user has to acquire specifically to promote stress management. The participants who tested MUSE were unwilling to purchase the device (see Chapter 2.3.2). A reason for this could be its cost-benefit ratio. Other wearables, such as fitness bands for example, allow for a multitude of services (e.g., step counting, activity monitoring, distance and speed tracking, heart rate monitoring, calorie counting, detection of sleep patterns, body temperature and hydration level) (Spil et al., 2017), that are often applicable in various contexts. This suggests that, if wearables are integrated in a stress management app, they should be low cost (Hairston and Lawhern, 2015) and support a variety of functionalities. Moreover, they should be adapted to the context and environment of their application, because both convenience and comfort determine whether a wearable will target niche groups or establish itself as a mainstream solution (Mihajlovic et al., 2015).

Also, the participants preferred test- and audio-instructed relaxation exercises over those that rely on neurofeedback. The cause of this could be that the integration of wearables in order to support stress management exercises through biofeedback creates a dependence. This disagrees with the users' need for autonomy (Sax et al., 2018). It is, therefore, questionable how many potential users can be reached with apps that are dependent on one specific wearable that users will have to purchase particularly for using the app. In addition, the participants found that the online neurofeedback distracted them from the original meditation task. This agrees with previous findings that indicate that stress management exercises with biofeedback have no immediate effects on subjective relaxation level in novices (Faust-Christmann et al., 2019). Future studies should, therefore, further investigate how wearables can support active relaxation and to what extent these approaches are accepted by users.

Furthermore, it was already mentioned that accuracy, legitimacy, and security are important to ensure the user's trust and encourage mHealth usage (Dennison et al., 2013). Hence, when integrating wearables, developers should keep in mind that, though more convenient and easer to come by, the data of commercially available devices can possibly be intercepted or exposed to third parties (Alahäivälä and Oinas-Kukkonen, 2016). Moreover, they should not blindly trust the provided data if there is no proof of the system's reliability and validity, as this can prevent the usage of the intervention (Dennison et al., 2013). For example, Chapter 2.3.2 suggests that the reliability and validity of the wearable EEG MUSE depends on its application context and the utilized frequency bands. Similar observations were made in a previous study (Badcock et al., 2013).

Overall, the results of the conducted studies support the previous assumption that there are obstacles to the adoption of wearables for health purposes (Mihajlovic et al., 2015). When deciding which wearables to integrate into health apps, developers should make sure that the utilized technologies are accessible to people from a multitude of backgrounds and not just a certain age group or a group of enthusiasts (Alahäivälä and Oinas-Kukkonen, 2016). Also, developers should keep in mind that applying wearables in daily life situations might require the monitoring of contextual information (Mihajlovic et al., 2015), an aspect that is viewed critically by users (Dennison et al., 2013). In order to ensure the device's acceptance by users, it might also be necessary that the wearable is transparent or even invisible (McDowell et al., 2013). To identify and meet the demands that drive the acceptance and adoption of wearables into stress management interventions it is necessary to include users and other stakeholders into the development process (Mosconi et al., 2019).

6.5 Limitations and Outlook

Though this thesis shows that gamification in the context of stress management is accepted by users and has positive impact on the usage behavior, it does not address the app's effects on stress. The app includes important content, such as stress management exercises and behavior change techniques that should theoretically ensure its effectiveness. However, so far it remains unclear whether Stress-Mentor does in fact reduce stress or whether it positively affects the users' application of stress coping strategies in everyday life. Future studies should investigate this aspect in longitudinal randomized controlled trials. Moreover, it would be interesting to see if and to what extent gamification might mediate these effects, as gamification often only elicits short-term engagement (Sardi et al., 2017).

In addition, this thesis shows that Stress-Mentor's gamification concept can be easily adapted for the use in other health contexts. However, it has to be kept in mind that the context of an application has great impact on the effects and acceptance of such technologies (Johnson et al., 2016). Because the requirements differ (Carter et al., 2015), the opinions of experts and users should be obtained in order to disclose areas of improvement and adaptation when applying Stress-Mentor's concept in other contexts.

Another aspect that should be taken into account is that, so far, Stress-Mentor is only available for Android devices. This could lead to a bias in the collected data, because differences between Android and iOS users have been reported, e.g. with respect to income levels (Smith, 2013). However, with a market share of 78 % in Germany and 76 % worldwide for Android based devices (Statista, 2019b), the results and implications of the presented studies should be somewhat generalizable. Still, it limits Stress-Mentor's accessibility to some extent. To increase the app's coverage, it would be essential to extend the service to iOS and Windows platforms in a next step.

6.6 Conclusion

It has been shown in previous research that gamification can positively affect the usage of mHealth technologies (e.g., Taylor et al., 2019). However, it remained unclear whether this held true in the context of stress management apps and how such an approach would be perceived by users. This thesis shows that the combination of evidence-based content from behavior change theory, stress management theory, and gamification theory in the stress management app Stress-Mentor, is well received by users and experts alike. The quality, usability, and user experience ratings suggest that the app is accepted. The results also showed that gamification led to an increased usage of Stress-Mentor, compared to a control group who received a non-gamified version of the app. Furthermore, the increased number of performed tasks indicates that the app's gamified concept positively affects the users' compliance.

Though gamification resulted in an overall positive outcome, mHealth developers should critically scrutinize whether they want to link their apps to wearables. Even though users appear to be open to the usage of such devices, they are not always willing to acquire them. This appeared to be especially the case if the device was linked to relaxation aimed online feedback. In fact, the participants stated that they preferred audio- and text-based instructions for stress management exercises.

Overall, the applied iterative development process of Stress-Mentor resulted in a high-quality product that is readily adopted. In addition, the results imply that the use of gamification in Stress-Mentor could increase the app's effectivity by increasing its usage and, thus, the user's contact with the evidence based stress management strategies and behavior change techniques.

The presented research therefore extends the existing literature on the effects of gamification and supports the integration of gamification into stress management apps. Furthermore, it provides insight into important considerations that need to be made when developing health apps, as well as possible solutions to encountered challenges.

References

Abraham, C., and Michie, S. (2008). A taxonomy of behavior change techniques used in interventions. *Health psychology: official journal of the Division of Health Psychology, American Psychological Association* 27, 379–387. doi: 10.1037/0278-6133.27.3.379

Abujelala, M., Abellanoza, C., Sharma, A., and Makedon, F. (2016). "Brain-EE: Brain Enjoyment Evaluation using Commercial EEG Headband," in *Proceedings of the 9th ACM International Conference on Pervasive Technologies Related to Assistive Environments 2016* (Corfu, Greece: ACM Press), 1–5

Acharya, S. D., Elci, O. U., Sereika, S. M., Styn, M. A., and Burke, L. E. (2011). Using a personal digital assistant for self-monitoring influences diet quality in comparison to a standard paper record among overweight/obese adults. *Journal of the American Dietetic Association* 111, 583–588. doi: 10.1016/j.jada.2011.01.009

Ahtinen, A., Mattila, E., Välkkynen, P., Kaipainen, K., Vanhala, T., Ermes, M., et al. (2013). Mobile mental wellness training for stress management: feasibility and design implications based on a one-month field study. *JMIR mHealth and uHealth* 1, e11. doi: 10.2196/mhealth.2596

Akter, S., Ray, P., and D'Ambra, J. (2013). Continuance of mHealth services at the bottom of the pyramid: the roles of service quality and trust. *Electron Markets* 23, 29–47. doi: 10.1007/s12525-012-0091-5

Alahäivälä, T., and Oinas-Kukkonen, H. (2016). Understanding persuasion contexts in health gamification: A systematic analysis of gamified health behavior change support systems literature. *International journal of medical informatics* 96, 62–70. doi: 10.1016/j.ijmedinf.2016.02.006

Albrecht, U.-V., and Fangerau, H. (2015). Do Ethics Need to Be Adapted to mHealth? A Plea for Developing a Consistent Framework. *Journal of medical Internet research* 2. doi: 10.3233/978-1-61499-538-8-219

Alexander, J. C., and Joshi, G. P. (2016). Smartphone applications for chronic pain management: a critical appraisal. *Journal of pain research* 9, 731–734. doi: 10.2147/JPR.S119966

Allam, A., Kostova, Z., Nakamoto, K., and Schulz, P. J. (2015). The Effect of Social Support Features and Gamification on a Web-Based Intervention for Rheumatoid Arthritis Patients: Randomized Controlled Trial. *Journal of medical Internet research* 17, e14. doi: 10.2196/jmir.3510

AlMarshedi, A., Wills, G., and Ranchhod, A. (2016). Gamifying Self-Management of Chronic Illnesses: A Mixed-Methods Study. *JMIR serious games* 4, e14. doi: 10.2196/games.5943

Alvarez, G. G., and Ayas, N. T. (2004). The Impact of Daily Sleep Duration on Health: A Review of the Literature. *Progress in cardiovascular nursing* 19, 56–59. doi: 10.1111/j.0889-7204.2004.02422.x

American Psychological Association (2012). *Stress in America™: Our Health at Risk*, https://www.apa.org/news/press/releases/stress/2011/final-2011.pdf

American Psychological Association (2013). *Managing chronic pain: How psychologists help with pain management*, http://www.apa.org/helpcenter/pain-management.pdf

American Psychological Association (2015). *Stress Report*, https://www.apa.org/news/press/releases/stress/2014/stress-report.pdf

Anderson, K., Burford, O., and Emmerton, L. (2016a). App Chronic Disease Checklist: Protocol to Evaluate Mobile Apps for Chronic Disease Self-Management. *JMIR research protocols* 5, e204. doi: 10.2196/resprot.6194

Anderson, K., Burford, O., and Emmerton, L. (2016b). Mobile Health Apps to Facilitate Self-Care: A Qualitative Study of User Experiences. *PloS one* 11, e0156164. doi: 10.1371/journal.pone.0156164

Anglada-Martinez, H., Riu-Viladoms, G., Martin-Conde, M., Rovira-Illamola, M., Sotoca-Momblona, J. M., and Codina-Jane, C. (2015). Does mHealth increase adherence to medication? Results of a systematic review. *International journal of clinical practice* 69, 9–32. doi: 10.1111/ijcp.12582

AppBrain (2019). *Android app download statistics on Google Play*, https://www.appbrain.com/stats/android-app-downloads

Atienza, A. A., and Patrick, K. (2011). Mobile health: the killer app for cyberinfrastructure and consumer health. *American journal of preventive medicine* 40, 3. doi: 10.1016/j.amepre.2011.01.008

Badcock, N. A., Mousikou, P., Mahajan, Y., de Lissa, P., Thie, J., and McArthur, G. (2013). Validation of the Emotiv EPOC EEG gaming system for measuring research quality auditory ERPs. *PeerJ* 1, e38. doi: 10.7717/peerj.38

Bamidis, P. D., Gabarron, E., Hors-Fraile, S., Konstantinidis, E., Konstantinidis, S., and Rivera, O. (2016). "Gamification and Behavioral Change: Techniques for Health Social Media," in *Participatory health through social media*, eds. S. Syed-Abdul, E. Gabarron, and A. Y. S. Lau (London, UK: Elsevier), 112–135

Bandura, A. (1998). Health promotion from the perspective of social cognitive theory. *Psychology & Health* 13, 623–649. doi: 10.1080/08870449808407422

Bandura, A., Ross, D., and Ross, S. A. (1963). Vicarious reinforcement and imitative learning. *The Journal of Abnormal and Social Psychology* 67, 601–607. doi: 10.1037/h0045550

Bangor, A., Kortum, P., and Miller, J. (2009). Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *Journal of Usability Studies* 4, 114–123

Baptista, S., Oldenburg, B., and O'Neil, A. (2017). Response to "Development and Validation of the User Version of the Mobile Application Rating Scale (uMARS)". *JMIR mHealth and uHealth* 5, e16. doi: 10.2196/mhealth.6419

Barabasz, A., and Perez, N. (2007). Salient findings: hypnotizability as core construct and the clinical utility of hypnosis. *The International journal of clinical and experimental hypnosis* 55, 372–379. doi: 10.1080/00207140701339793

Barlow, J. H., and Ellard, D. R. (2006). The psychosocial well-being of children with chronic disease, their parents and siblings: an overview of the research evidence base. *Child: care, health and development* 32, 19–31. doi: 10.1111/j.1365-2214.2006.00591.x

Barnum, C. M. (2011). *Usability testing essentials: Ready, set - test!* Amsterdam: Elsevier

Bashivan, P., Rish, I., and Heisig, S. (2016). "Mental State Recognition via Wearable EEG," in *Proceedings of the 5th NIPS workshop on Machine Learning and Interpretation in Neuroimaging 2016*

Baum, A., and Posluszny, D. M. (1999). Health psychology: mapping biobehavioral contributions to health and illness. *Annual review of psychology* 50, 137–163. doi: 10.1146/annurev.psych.50.1.137

Beatty, J. (1972). Similar effects of feedback signals and instructional information on EEG activity. *Physiology & Behavior* 9, 151–154. doi: 10.1016/0031-9384(72)90227-2

Becker, N., Bondegaard Thomsen, A., Olsen, A. K., Sjøgren, P., Bech, P., and Eriksen, J. (1997). Pain epidemiology and health related quality of life in chronic non-malignant pain patients referred to a Danish multidisciplinary pain center. *Pain* 73, 393–400

Bender, J. L., Radhakrishnan, A., Diorio, C., Englesakis, M., and Jadad, A. R. (2011). Can pain be managed through the Internet? A systematic review of randomized controlled trials. *Pain* 152, 1740–1750. doi: 10.1016/j.pain.2011.02.012

Berger, H. (1929). Über das Elektrenkephalogramm des Menschen. *Archiv für Psychiatrie* 87, 527–570. doi: 10.1007/BF01797193

Bickmore, T. W., Caruso, L., and Clough-Gorr, K. (2005). "Acceptance and usability of a relational agent interface by urban older adults," in *CHI. Extended abstracts on Human factors in computing systems 2005* (New York, New York, USA: ACM Press), 1212–1215

Birbaumer, N., and Schmidt, R. F. (2006). *Biologische Psychologie*. Berlin/Heidelberg: Springer-Verlag

Blackburne, T., Rodriguez, A., and Johnstone, S. J. (2016). A Serious Game to Increase Healthy Food Consumption in Overweight or Obese Adults: Randomized Controlled Trial. *JMIR serious games* 4, e10. doi: 10.2196/games.5708

Blankertz, B., Tangermann, M., Vidaurre, C., Fazli, S., Sannelli, C., Haufe, S., et al. (2010). The Berlin Brain-Computer Interface: Non-Medical Uses of BCI Technology. *Frontiers in neuroscience* 4, 198. doi: 10.3389/fnins.2010.00198

Bleichner, M. G., Mirkovic, B., and Debener, S. (2016). Identifying auditory attention with ear-EEG: cEEGrid versus high-density cap-EEG comparison. *Journal of neural engineering* 13, 66004. doi: 10.1088/1741-2560/13/6/066004

Boeing, H., Bechthold, A., Bub, A., Ellinger, S., Haller, D., Kroke, A., et al. (2012). *Gemüse und Obst in der Prävention ausgewählter chronischer Krankheiten*, https://www.dge.de/fileadmin/public/doc/ws/stellungnahme/DGE-Stellungnahme-Gemuese-Obst-2012.pdf

Bogg, T., and Roberts, B. W. (2004). Conscientiousness and health-related behaviors: a meta-analysis of the leading behavioral contributors to mortality. *Psychological bulletin* 130, 887–919. doi: 10.1037/0033-2909.130.6.887

Bortz, J. (2016). *Statistik für human- und sozialwissenschaftler*. Springer

Bostock, S., Crosswell, A. D., Prather, A. A., and Steptoe, A. (2019). Mindfulness on-the-go: Effects of a mindfulness meditation app on work stress and well-being. *Journal of occupational health psychology* 24, 127–138. doi: 10.1037/ocp0000118

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* 4, 363–371. doi: 10.1007/s13142-014-0293-9

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, 49–59. doi: 10.1016/0005-7916(94)90063-9

Brauner, P., Calero Valdez, A., Schroeder, U., and Ziefle, M. (2013). "Increase Physical Fitness and Create Health Awareness through Exergames and Gamification," in *Human Factors in Computing and Informatics 2013*, eds. D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, et al. (Berlin, Heidelberg: Springer), 349–362

Breivik, H., Collett, B., Ventafridda, V., Cohen, R., and Gallacher, D. (2006). Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. *European journal of pain* 10, 287–333. doi: 10.1016/j.ejpain.2005.06.009

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, 523–529. doi: 10.1007/s13142-011-0076-5

Brooke, J. (1996). SUS - A quick and dirty usability scale. *Usability evaluation in industry*, 189–194

Byrne, S., Gay, G., Pollack, J. P., Gonzales, A., Retelny, D., Lee, T., et al. (2012). Caring for Mobile Phone-Based Virtual Pets can Influence Youth Eating Behaviors. *Journal of Children and Media* 6, 83–99. doi: 10.1080/17482798.2011.633410

Cafazzo, J. A., Casselman, M., Hamming, N., Katzman, D. K., and Palmert, M. R. (2012). Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. *Journal of medical Internet research* 14, e70. doi: 10.2196/jmir.2058

Carmona, L. (2001). The burden of musculoskeletal diseases in the general population of Spain: Results from a national survey. *Annals of the Rheumatic Diseases* 60, 1040–1045. doi: 10.1136/ard.60.11.1040

Carter, A., Liddle, J., Hall, W., and Chenery, H. (2015). Mobile Phones in Research and Treatment: Ethical Guidelines and Future Directions. *JMIR mHealth and uHealth* 3, e95. doi: 10.2196/mhealth.4538

Chan, S., Torous, J., Hinton, L., and Yellowlees, P. (2015). Towards a Framework for Evaluating Mobile Mental Health Apps. *Telemedicine journal and e-health: the official journal of the American Telemedicine Association* 21, 1038–1041. doi: 10.1089/tmj.2015.0002

Chi, Y. M., Jung, T.-P., and Cauwenberghs, G. (2010). Dry-contact and noncontact biopotential electrodes: methodological review. *IEEE reviews in biomedical engineering* 3, 106–119. doi: 10.1109/RBME.2010.2084078

Chittaro, L., and Sioni, R. (2014). Evaluating mobile apps for breathing training: The effectiveness of visualization. *Computers in Human Behavior* 40, 56–63. doi: 10.1016/j.chb.2014.07.049

Chou, Y.-k. (2015). *Actionable Gamification: Beyond Points, Badges, and Leaderboards*. Octalysis Media

Christmann, C. A., Hoffmann, A., and Bleser, G. (2017a). Stress Management Apps With Regard to Emotion-Focused Coping and Behavior Change Techniques: A Content Analysis. *JMIR mHealth and uHealth* 5, e22. doi: 10.2196/mhealth.6471

Christmann, C. A., Hoffmann, A., Zolynski, G., and Bleser, G. (2018a). "Stress-Mentor: Linking Gamification and Behavior Change Theory in a Stress Management Application.," in *HCI International. Posters' extended abstracts: 20th International Conference 2018*, ed. C. Stephanidis (Cham, Switzerland: Springer), 237–247

Christmann, C. A., Zolynski, G., Hoffmann, A., and Bleser, G. (2017b). "Effective Visualization of Long Term Health Data to Support Behavior Change," in *HCI International: Digital Human Modeling. Applications in Health, Safety, Ergonomics, and Risk Management: Ergonomics and Design: 8th International Conference 2017*, ed. V. G. Duffy (Cham: Springer), 237–247

Christmann, C. A., Zolynski, G., Hoffmann, A., and Bleser, G. (2018b). "Towards More Interactive Stress-Related Self-Monitoring Tools to Improve Quality of Life," in *Proceedings of the 20th Congress of the International Ergonomics Association. Volume I, Healthcare ergonomics 2018*, eds. S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, and Y. Fujita (Cham, Switzerland: Springer), 121–130

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale, New Jersey, USA: L. Erlbaum Associates

Cohen, J. (1992). Statistical Power Analysis. *Current Directions in Psychological Science* 1, 98–101. doi: 10.1111/1467-8721.ep10768783

Cohen, S., and Hoberman, H. M. (1983). Positive Events and Social Supports as Buffers of Life Change Stress. *Journal of Applied Social Psychology* 13, 99–125. doi: 10.1111/j.1559-1816.1983.tb02325.x

Cohen, S., Janicki-Deverts, D., and Miller, G. E. (2007). Psychological Stress and Disease. *Journal of American Medical Association* 298, 1685–1687. doi: 10.1001/jama.298.14.1685

Cohen, S., Kamarck, T., and Mermelstein, R. (1994). Perceived Stress Scale. *Measuring stress: A guide for health and social scientists* 10

Cooper, C. L., ed (1994). *Causes, coping and consequences of stress at work: Chapter 8: The determinants and consequences of coping with stress*. Chichester, New York, USA: J. Wiley

Coulon, S. M., Monroe, C. M., and West, D. S. (2016). A Systematic, Multi-domain Review of Mobile Smartphone Apps for Evidence-Based Stress Management. *American journal of preventive medicine*. doi: 10.1016/j.amepre.2016.01.026

Cutrell, E., and Tan, D. (2008). "BCI for passive input in HCI," in *Proceedings of CHI 2008* (Berlin: ACM Press)

Cuttone, A., Lehmann, S., and Larsen, J. E. (2013). "A mobile personal informatics system with interactive visualizations of mobility and social interactions," in *PDI. Proceedings of the 1st ACM international workshop on Personal data meets distributed multimedia 2013*, eds. V. K. Singh, T.-S. Chua, R. Jain, and A. S. Pentland (New York, New York, USA: ACM Press), 27–30

Cuttone, A., Petersen, M. K., and Larsen, J. E. (2014). "Four Data Visualization Heuristics to Facilitate Reflection in Personal Informatics," in *Universal Access in Human-Computer Interaction. Design for All and Accessibility Practice 2014*, eds. D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, A. Kobsa, F. Mattern, et al. (Cham: Springer), 541–552

Czaja, S. J., Charness, N., Fisk, A. D., Hertzog, C., Nair, S. N., Rogers, W. A., et al. (2006). Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement. *Psychology and aging* 21, 333–352. doi: 10.1037/0882-7974.21.2.333

Davies, C. A., Spence, J. C., Vandelanotte, C., Caperchione, C. M., and Mummery, W. K. (2012). Meta-analysis of internet-delivered interventions to increase physical activity levels. *The international journal of behavioral nutrition and physical activity* 9, 52. doi: 10.1186/1479-5868-9-52

Davis, M., Eshelman, E. R., and McKay, M. (2008). *The relaxation and stress reduction workbook: 6th edition*. Oakland, California, USA: New Harbinger

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, e101312. doi: 10.1371/journal.pone.0101312

De Vos, M., Kroesen, M., Emkes, R., and Debener, S. (2014). P300 speller BCI with a mobile EEG system: comparison to a traditional amplifier. *Journal of neural engineering* 11, 36008. doi: 10.1088/1741-2560/11/3/036008

DeLongis, A., Coyne, J. C., Dakof, G., Folkman, S., and Lazarus, R. S. (1982). Relationship of daily hassles, uplifts, and major life events to health status. *Health psychology: official journal of the Division of Health Psychology, American Psychological Association* 1, 119–136. doi: 10.1037/0278-6133.1.2.119

DeLongis, A., Folkman, S., and Lazarus, R. S. (1988). The impact of daily stress on health and mood: Psychological and social resources as mediators. *Journal of personality and social psychology* 54, 486–495. doi: 10.1037//0022-3514.54.3.486

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Demiris, G., Afrin, L. B., Speedie, S., Courtney, K. L., Sondhi, M., Vimarlund, V., et al. (2008). Patient-centered applications: use of information technology to promote disease management and wellness. A white paper by the AMIA knowledge in motion working group. *Journal of the American Medical Informatics Association* 15, 8–13. doi: 10.1197/jamia.M2492

Dennison, L., Morrison, L., Conway, G., and Yardley, L. (2013). Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. *Journal of medical Internet research* 15, e86. doi: 10.2196/jmir.2583

Deterding, S., Khaled, R., Nacke, L. E., and Dixon, D. (2011). "Gamification: Toward a Definition," in *CHI. Proceedings of the annual conference extended abstracts on Human factors in computing systems 2011* (New York, New York, USA: ACM Press)

DiClemente, C. C., Marinilli, A. S., Singh, M., and Bellino, L. E. (2001). The Role of Feedback in the Process of Health Behavior Change. *American journal of health behavior* 25, 217–227. doi: 10.5993/AJHB.25.3.8

Ebben, M. R., Kurbatov, V., and Pollak, C. P. (2009). Moderating laboratory adaptation with the use of a heart-rate variability biofeedback device (StressEraser). *Applied psychophysiology and biofeedback* 34, 245–249. doi: 10.1007/s10484-009-9086-1

Economides, M., Martman, J., Bell, M. J., and Sanderson, B. (2018). Improvements in Stress, Affect, and Irritability Following Brief Use of a Mindfulness-based Smartphone App: A Randomized Controlled Trial. *Mindfulness* 49, 55. doi: 10.1007/s12671-018-0905-4

Edwards, E. A., Caton, H., Lumsden, J., Rivas, C., Steed, L., Pirunsarn, Y., et al. (2018). Creating a Theoretically Grounded, Gamified Health App: Lessons From Developing the Cigbreak Smoking Cessation Mobile Phone Game. *JMIR serious games* 6, e10252. doi: 10.2196/10252

Eikey, E., Poole, E., and Reddy, M. (2015). "Information Presentation in Health Apps and Devices: The Effect of Color, Distance to Goal, Weight Perception, and Interest on Users' Self-Efficacy for Accomplishing Goals," in *iConference Proceedings 2015* (Newport Beach, California, USA: iSchools)

Ekandem, J. I., Davis, T. A., Alvarez, I., James, M. T., and Gilbert, J. E. (2012). Evaluating the ergonomics of BCI devices for research and experimentation. *Ergonomics* 55, 592–598. doi: 10.1080/00140139.2012.662527

Elgar, F. J., and McGrath, P. J. (2003). Self-administered psychosocial treatments for children and families. *Journal of clinical psychology* 59, 321–339. doi: 10.1002/jclp.10132

Estepp, J. R., Christensen, J. C., Monnin, J. W., Davis, I. M., and Wilson, G. F. (2009). "Validation of a Dry Electrode System for EEG," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2009* (Los Angeles, California, USA: SAGE Publications), 1171–1175

European Food Information Council (2018). *Koffein und Gesundheit*, http://www.eufic.org/article/de/artid/Koffein-Gesundheit/

Farrar, J. T., Young, J. P., LaMoreaux, L., Werth, J. L., and Poole, M. R. (2001). Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain* 94, 149–158. doi: 10.1016/S0304-3959(01)00349-9

Faust-Christmann, C. A., Spilski, J., Mayerl, J., and Bleser, G. (2018). "Einflussfaktoren auf die Nutzungsintention einer Stress-Management-App: Weder positive noch negative Effekte von Empfehlungen und Bonusprogrammen durch Krankenkassen.," in *Dritte Transdiziplinäre Konferenz "Technische Unterstützungssysteme, die die Menschen wirklich wollen" 2018* (Hamburg, Germany: Helmut-Schmidt-Universität)

Faust-Christmann, C. A., Taetz, B., Zolynski, G., Zimmermann, T., and Bleser, G. (2019). "Breathing-Mentor": A pilot study with a biofeedback app to instruct abdominal breathing. *JMIR mHealth uHealth* in print. doi: 10.2196/13703

Fiedler, P., Haueisen, J., Jannek, D., Griebel, S., Zentner, L., Vaz, F., et al. (2014). Comparison of three types of dry electrodes for electroencephalography. *ACTA IMEKO* 3, 33. doi: 10.21014/acta_imeko.v3i3.94

Field, A. P. (2009). *Discovering statistics using SPSS: (and sex, drugs and rock 'n' roll)*. Los Angeles: SAGE Publications

Fleck, R., and Fitzpatrick, G. (2010). "Reflecting on reflection," in *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction 2010*, eds. M. Brereton, S. Viller, and B. Kraal (New York, New York, USA: ACM Press), 216

Flor, H., Fydrich, T., and Turk, D. C. (1992). Efficacy of multidisciplinary pain treatment centers: A meta-analytic review. *Pain* 49, 221–230. doi: 10.1016/0304-3959(92)90145-2

Fogg, B. J. (2002). Persuasive technology: using computers to change what we think and do. *Ubiquity* 2, 5. doi: 10.1145/764008.763957

Fox, J., Bailenson, J., and Binney, J. (2009). Virtual Experiences, Physical Behaviors: The Effect of Presence on Imitation of an Eating Avatar. *Presence: Teleoperators and Virtual Environments* 18, 294–303. doi: 10.1162/pres.18.4.294

Frederick, J. A. (2012). Psychophysics of EEG alpha state discrimination. *Consciousness and cognition* 21, 1345–1354. doi: 10.1016/j.concog.2012.06.009

Friedel, H., and Trautvetter, D. (2011). Gesundheitsbezogene Verhaltensmodifikationen durch Bonusprogramme. *Prävention und Gesundheitsförderung* 6, 292–298. doi: 10.1007/s11553-011-0302-5

Gagnon, M.-P., Ngangue, P., Payne-Gagnon, J., and Desmartis, M. (2016). m-Health adoption by healthcare professionals: a systematic review. *Journal of the American Medical Informatics Association* 23, 212–220. doi: 10.1093/jamia/ocv052

Gallicchio, L., and Kalesan, B. (2009). Sleep duration and mortality: a systematic review and meta-analysis. *Journal of sleep research* 18, 148–158. doi: 10.1111/j.1365-2869.2008.00732.x

Gay, V., and Leijdekkers, P. (2012). Personalised mobile health and fitness apps: lessons learned from myFitnessCompanion®. *Studies in health technology and informatics* 177, 248–253. doi: 10.3233/978-1-61499-069-7-248

German Nutrition Society (2010). *Prävention durch moderaten Alkoholkonsum?* https://www.dge.de/up-loads/media/DGE-Pressemeldung-aktuell-02-2010_Alkohol-Karneval.pdf

German Nutrition Society (2017). *10 guidelines of the German Nutrition Society (DGE) for a wholesome diet*, https://www.dge.de/index.php?id=322

Gibbons, P., Boudreaux, E. D., and Haskins, B. L. (2018). Using practical search strategies to identify health apps: A case study with smoking cessation. *Journal of mobile technology in medicine* 7, 16–23. doi: 10.7309/jmtm.7.1.3

Glanz, K., Rimer, B. K., and Viswanath, K., eds (2015). *Health Behavior: Theory, Research, and Practice*. New York, New York, USA: John Wiley & Sons

González, C. S., Gómez, N., Navarro, V., Cairós, M., Quirce, C., Toledo, P., et al. (2016). Learning healthy lifestyles through active videogames, motor games and the gamification of educational activities. *Computers in Human Behavior* 55, 529–551. doi: 10.1016/j.chb.2015.08.052

Goyal, S., Morita, P. P., Picton, P., Seto, E., Zbib, A., and Cafazzo, J. A. (2016). Uptake of a Consumer-Focused mHealth Application for the Assessment and Prevention of Heart Disease: The <30 Days Study. *JMIR mHealth and uHealth* 4, e32. doi: 10.2196/mhealth.4730

Gramann, K., Gwin, J. T., Ferris, D. P., Oie, K., Jung, T.-P., Lin, C.-T., et al. (2011). Cognition in action: imaging brain/body dynamics in mobile humans. *Reviews in the neurosciences* 22, 593–608. doi: 10.1515/RNS.2011.047

Gramann, K., Jung, T.-P., Ferris, D. P., Lin, C.-T., and Makeig, S. (2014). Toward a new cognitive neuroscience: modeling natural brain dynamics. *Frontiers in human neuroscience* 8, 444. doi: 10.3389/fnhum.2014.00444

Grozea, C., Voinescu, C. D., and Fazli, S. (2011). Bristle-sensors – low-cost flexible passive dry EEG electrodes for neurofeedback and BCI applications. *Journal of neural engineering* 8, 25008. doi: 10.1088/ 1741-2560/8/2/025008

Guo, X., Zhang, X., and Sun, Y. (2016). The privacy – personalization paradox in mHealth services acceptance of different age groups. *Electronic Commerce Research and Applications* 16, 55–65. doi: 10.1016/j.elerap.2015.11.001

Hairston, W. D., and Lawhern, V. (2015). How Low Can You Go? Empirically Assessing Minimum Usable DAQ Performance for Highly Fieldable EEG Systems 9183, 221–231. doi: 10.1007/978-3-319-20816-9_22

Hamari, J. (2013). Transforming homo economicus into homo ludens: A field experiment on gamification in a utilitarian peer-to-peer trading service. *Electronic Commerce Research and Applications* 12, 236–245. doi: 10.1016/j.elerap.2013.01.004

Handel, M. J. (2011). mHealth (mobile health)-using Apps for health and wellness. *Explore* 7, 256–261. doi: 10.1016/j.explore.2011.04.011

Harland, C. J., Clark, T. D., and Prance, R. J. (2002). Remote detection of human electroencephalograms using ultrahigh input impedance electric potential sensors. *Applied Physics Letters* 81, 3284. doi: 10.1063/1.1516861

Harrison, V., Proudfoot, J., Wee, P. P., Parker, G., Pavlovic, D. H., and Manicavasagar, V. (2011). Mobile mental health: Review of the emerging field and proof of concept study. *Journal of Mental Health* 20, 509–524. doi: 10.3109/09638237.2011.608746

Hassenzahl, M. (2001). The Effect of Perceived Hedonic Quality on Product Appealingness. *International Journal of Human-Computer Interaction* 13, 481–499. doi: 10.1207/S15327590IJHC1304_07

Hassenzahl, M., Platz, A., Burmester, M., and Lehner, K. (2000). "Hedonic and ergonomic quality aspects determine a software's appeal," in *Proceedings of the SIGCHI conference on Human factors in computing systems 2000*, eds. T. Turner, and G. Szwillus (New York, New York, USA: ACM Press), 201–208

Hawker, G. A., Mian, S., Kendzerska, T., and French, M. (2011). Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). *Arthritis care & research* 63, 240–252. doi: 10.1002/acr.20543

He, D., Naveed, M., Gunter, C. A., and Nahrstedt, K. (2014). "Security Concerns in Android mHealth Apps," in *AMIA Annual Symposium 2014* (American Medical Informatics Association).

Hierhammer, C., and Herrmanny, K. (2013). "Gamification für ältere Menschen – Potenziale und Herausforderungen," in *Mensch & Computer - Workshopband. 13. fachübergreifende Konferenz für interaktive und kooperative Medien 2013*, eds. S. Boll-Westermann, S. Maaß, and R. Malaka (Berlin/Boston: De Gruyter Oldenbourg)

Hjelm, S., and Browall, C. (2000). "Brainball–using brain activity for cool competition," in *Proceedings of the 1st Nordic Conference on HumanComputer Interaction 2000* (New York, New York, USA: ACM Press), 177–188

Hoffmann, A., Christmann, C. A., and Bleser, G. (2017). Gamification in Stress Management Apps: A Critical App Review. *JMIR serious games* 5, e13. doi: 10.2196/games.7216

Hoffmann, A., Faust-Christmann, C. A., Zolynski, G., and Bleser, G. (2019). "Gamification of a Stress Management App: Results of a User Study," in *HCI International: Design, User Experience, and Usability. Application Domains 2019*, eds. A. Marcus, and W. Wang (Cham: Springer), 303–313

Holmes, J. (2007). Designing agents to support learning by explaining. *Computers & Education* 48, 523–547. doi: 10.1016/j.compedu.2005.02.007

Ijjada, M. S., Thapliyal, H., Caban-Holt, A., and Arabnia, H. R. (2015). "Evaluation of Wearable Head Set Devices in Older Adult Populations for Research," in *International Conference on Computational Science and Computational Intelligence 2015* (IEEE), 810–811

Institute for Healthcare Informatics (2013). *Patient Apps for Improved Healthcare: From Novelty to Mainstream*, http://moodle.univlille2.fr/pluginfile.php/215345/mod_resource/content/0/IIHI_Patient_Apps_Report.pdf

InteraXon Inc. *MUSE*, http://www.choosemuse.com/

IQVIA (2017). *The Growing Value of Digital Health: Evidence and Impact on Human Health and the Healthcare System*, https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/growing-value-of-digital-health-infographic.pdf?la=en&hash=996EEEEDD6EAA3B4ADA23BEAFC53B66483F665B9&_=1566396996401

Irvine, A. B., Russell, H., Manocchia, M., Mino, D. E., Cox Glassen, T., Morgan, R., et al. (2015). Mobile-Web app to self-manage low back pain: randomized controlled trial. *Journal of medical Internet research* 17, e1. doi: 10.2196/jmir.3130

Istepanian, R.S.H., and Lacal, J. C. (2003). "Emerging mobile communication technologies for health: some imperative notes on m-health," in *Proceedings of the 25th annual international conference of the IEEE Engineering in Medicine and Biology Society: A new beginning for human health 2003* (IEEE), 1414–1416

Jake-Schoffman, D. E., Silfee, V. J., Waring, M. E., Boudreaux, E. D., Sadasivam, R. S., Mullen, S. P., et al. (2017). Methods for Evaluating the Content, Usability, and Efficacy of Commercial Mobile Health Apps. *JMIR mHealth and uHealth* 5, e190. doi: 10.2196/mhealth.8758

Jamison, R. N., Mei, A., and Ross, E. L. (2018). Longitudinal trial of a smartphone pain application for chronic pain patients: Predictors of compliance and satisfaction. *Journal of telemedicine and telecare* 24, 93–100. doi: 10.1177/1357633X16679049

Jensen, M. P., Nielson, W. R., and Kerns, R. D. (2003). Toward the development of a motivational model of pain self-management. *The Journal of Pain* 4, 477–492. doi: 10.1016/S1526-5900(03)00779-X

Jibb, L. A., Stevens, B. J., Nathan, P. C., Seto, E., Cafazzo, J. A., Johnston, D. L., et al. (2018). Perceptions of Adolescents With Cancer Related to a Pain Management App and Its Evaluation: Qualitative Study Nested Within a Multicenter Pilot Feasibility Study. *JMIR mHealth and uHealth* 6, e80. doi: 10.2196/mhealth.9319

Johnson, D., Deterding, S., Kuhn, K.-A., Staneva, A., Stoyanov, S., and Hides, L. (2016). Gamification for health and wellbeing: A systematic review of the literature. *Internet Interventions* 6, 89–106. doi: 10.1016/j.invent.2016.10.002

Jones, M. C., and Johnston, D. W. (2000). Reducing distress in first level and student nurses: a review of the applied stress management literature. *Journal of Advanced Nursing* 32, 66–74. doi: 10.1046/j.1365-2648.2000.01421.x

Kamel Boulos, M. N., Gammon, S., Dixon, M. C., MacRury, S. M., Fergusson, M. J., Miranda Rodrigues, F., et al. (2015). Digital games for type 1 and type 2 diabetes: underpinning theory with three illustrative examples. *JMIR serious games* 3, e3. doi: 10.2196/games.3930

Kanfer, F. H. (1970). "Self-regulation: Research, issues and speculations," in *Behavior modification in clinical psychology 1970*, eds. C. Neuringer, and J. L. Michael (Appleton-Century-Crofts), 178–220

Kanfer, F. H., and Gaelick-Buys, L. (1991). "Self-Management methods," in *Helping people change: A textbook of methods 1991*, eds. F. H. Kanfer, and A. P. Goldstein (Allyn band Bacon), 305–360

Kanner, A. D., Coyne, J. C., Schaefer, C., and Lazarus, R. S. (1981). Comparison of two modes of stress measurement: Daily hassles and uplifts versus major life events. *Journal of behavioral medicine* 4, 1–39. doi: 10.1007/BF00844845

Kao, C.-K., and Liebovitz, D. M. (2017). Consumer Mobile Health Apps: Current State, Barriers, and Future Directions. *PM & R: the journal of injury, function, and rehabilitation* 9, 106–115. doi: 10.1016/j.pmrj.2017.02.018

Katule, N., Rivett, U., and Densmore, M. (2016). "A Family Health App," in *Proceedings of the 7th Annual Symposium on Computing for Development 2016* (New York, New York, USA: ACM Press), 1–10

Kaushik, R. M., Kaushik, R., Mahajan, S. K., and Rajesh, V. (2006). Effects of mental relaxation and slow breathing in essential hypertension. *Complementary therapies in medicine* 14, 120–126. doi: 10.1016/j.ctim.2005.11.007

King, A. C., Hekler, E. B., Grieco, L. A., Winter, S. J., Sheats, J. L., Buman, M. P., et al. (2013). Harnessing different motivational frames via mobile phones to promote daily physical activity and reduce sedentary behavior in aging adults. *PloS one* 8, e62613. doi: 10.1371/journal.pone.0062613

Kirschstein, T. (2008). Wie entsteht das EEG? *Das Neurophysiologie-Labor* 30, 29–37. doi: 10.1016/j.neulab.2008.05.003

Klasnja, P., and Pratt, W. (2012). Healthcare in the pocket: mapping the space of mobile-phone health interventions. *Journal of biomedical informatics* 45, 184–198. doi: 10.1016/j.jbi.2011.08.017

Klatt, M. D., Buckworth, J., and Malarkey, W. B. (2009). Effects of low-dose mindfulness-based stress reduction (MBSR-ld) on working adults. *Health education & behavior: the official publication of the Society for Public Health Education* 36, 601–614. doi: 10.1177/1090198108317627

Knight, J. F., Baber, C., Schwirtz, A., and Bristow, H. W. (2002). "The Comfort Assessment of Wearable Computers," in *Proceedings of the 6th international symposium on wearable computers 2002* (IEEE), 65–72

Kocielnik, R., and Sidorova, N. (2015). Personalized Stress Management: Enabling Stress Monitoring with LifelogExplorer. *Künstliche Intelligenz* 29, 115–122. doi: 10.1007/s13218-015-0348-1

Koivisto, J., and Hamari, J. (2019). The rise of motivational information systems: A review of gamification research. *International Journal of Information Management* 45, 191–210. doi: 10.1016/j.ijinfomgt.2018.10.013

Koldijk, S., Kraaij, W., and Neerincx, M. A. (2016). Deriving Requirements for Pervasive Well-Being Technology From Work Stress and Intervention Theory: Framework and Case Study. *JMIR mHealth and uHealth* 4, e79. doi: 10.2196/mhealth.5341

Koskinen, E., and Salminen, J. (2007). "A customizable mobile tool for supporting health behavior interventions," in *Proceedings of the 29th Annual International Conference of the IEEE EMBS Cité Internationale 2007* (IEEE), 5908–5911

Kotz, D. (2011). "A threat taxonomy for mHealth privacy," in *Proceedings of the 3rd International Conference on Communication Systems and Networks 2011* (Piscataway, New Jersey, USA: IEEE), 1–6

Krebs, P., and Duncan, D. T. (2015). Health App Use Among US Mobile Phone Owners: A National Survey. *JMIR mHealth and uHealth* 3, e101. doi: 10.2196/mhealth.4924

Krigolson, O. E., Williams, C. C., Norton, A., Hassall, C. D., and Colino, F. L. (2017). Choosing MUSE: Validation of a Low-Cost, Portable EEG System for ERP Research. *Frontiers in neuroscience* 11, 109. doi: 10.3389/fnins.2017.00109

Kropotov, J. D. (2009). *Quantitative EEG, event-related potentials and neurotherapy*. Elsevier/Academic

Kuo, M., and Chuang, T. (2016). How gamification motivates visits and engagement for online academic dissemination – An empirical study. *Computers in Human Behavior* 55, 16–27. doi: 10.1016/j.chb.2015.08.025

Labbé, E., Schmidt, N., Babin, J., and Pharr, M. (2007). Coping with stress: the effectiveness of different types of music. *Applied psychophysiology and biofeedback* 32, 163–168. doi: 10.1007/s10484-007-9043-9

Lalloo, C., Jibb, L. A., Rivera, J., Agarwal, A., and Stinson, J. N. (2015). "There's a Pain App for That": Review of Patient-targeted Smartphone Applications for Pain Management. *The Clinical journal of pain* 31, 557–563. doi: 10.1097/AJP.0000000000000171

Larsen, J. E., Cuttone, A., and Jørgensen, S. L. (2013). "QS Spiral: Visualizing periodic quantified self data," in *CHI. Proceedings of the Workshop on Personal Informatics in the Wild: Hacking Habits for Health & Happiness 2013* (New York, New York: ACM Press)

Laugwitz, B., Held, T., and Schrepp, M. (2008). "Construction and Evaluation of a User Experience Questionnaire," in *HCI and usability for education and work: 4th symposium of the workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society 2008*, ed. A. Holzinger (Berlin: Springer)

Lazarus, R. S. (1985). The Psychology of Stress and Coping. *Issues in Mental Health Nursing* 7, 399–418. doi: 10.3109/01612848509009463

Leape, C., Fong, A., and Ratwani, R. M. (2016). Heuristic Usability Evaluation of Wearable Mental State Monitoring Sensors for Healthcare Environments. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 60, 583–587. doi: 10.1177/1541931213601134

Lee, J., Bedra, M., and Finkelstein, J. (2014). A critical review of consumer health devices for stress self-management. *Studies in health technology and informatics* 202, 221–224. doi: 10.3233/978-1-61499-423-7-221

Lee, J.-A., Choi, M., Lee, S. A., and Jiang, N. (2018). Effective behavioral intervention strategies using mobile health applications for chronic disease management: a systematic review. *BMC medical informatics and decision making* 18, 12. doi: 10.1186/s12911-018-0591-0

Lee, K. H. (2009). Evaluation of Attention and Relaxation Levels of Archers in Shooting Process using Brain Wave Signal Analysis Algorithms. *감성과학* 12, 341–350

Lee, R. A., and Jung, M. E. (2018). Evaluation of an mHealth App (DeStressify) on University Students' Mental Health: Pilot Trial. *JMIR mental health* 5, e2. doi: 10.2196/mental.8324

Lee, S., Shin, Y., Woo, S., Kim, K., and Lee, H.-N. (2013). "Review of Wireless Brain-Computer Interface Systems," in *Brain-computer interface system: Recent progress and future prospects 2013*, ed. R. Fazel-Rezai (Rijeka, Croatia: InTech)

Lenarduzzi, V., and Taibi, D. (2016). "MVP Explained: A Systematic Mapping Study on the Definitions of Minimal Viable Product," in *SEAA: 42nd Euromicro Conference on Software Engineering and Advanced Applications 2016* (Piscataway, New Jersey, USA: IEEE), 112–119

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

Lewis, T. L., and Wyatt, J. C. (2014). mHealth and mobile medical Apps: a framework to assess risk and promote safer use. *Journal of medical Internet research* 16, e210. doi: 10.2196/jmir.3133

Li, I., Dey, A. K., and Forlizzi, J. (2011). "Understanding my data, myself," in *Proceedings of the 13th international conference on Ubiquitous computing 2011*, eds. J. Landay, Y. Shi, D. J. Patterson, Y. Rogers, and X. Xie (New York, New York, USA: ACM Press), 405

Li, I., Medynskiy, Y., Froehlich, J., and Larsen, J. (2012). "Personal informatics in practice," in *Proceedings of the ACM annual conference extended abstracts on Human Factors in Computing Systems Extended Abstracts 2012*, eds. J. A. Konstan, E. H. Chi, and K. Höök (New York, New York, USA: ACM Press), 2799

Liao, L. D., Wang, I. J., Chen, S. F., Chang, J. Y., and Lin, C. T. (2011). Design, fabrication and experimental validation of a novel dry-contact sensor for measuring electroencephalography signals without skin preparation. *Sensors* 11, 5819–5834. doi: 10.3390/s110605819

Libenson, M. H. (2010). *Practical approach to electroencephalography: Chapter 7: Filters in the Electroencephalogram*. Philadelphia, Pennsylvania: Elsevier/Saunders

Lopez-Calderon, J., and Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in human neuroscience* 8, 213. doi: 10.3389/fnhum.2014.00213

Lopez-Gordo, M., Morillo, D., and Valle, F. (2014). Dry EEG Electrodes. *Sensors* 14, 12847–12870. doi: 10.3390/s140712847

Lu, L. (1991). Daily hassles and mental health: A longitudinal study. *British Journal of Psychology* 82, 441–447. doi: 10.1111/j.2044-8295.1991.tb02411.x

Luxton, D. D., McCann, R. A., Bush, N. E., Mishkind, M. C., and Reger, G. M. (2011). mHealth for mental health: Integrating smartphone technology in behavioral healthcare. *Professional Psychology: Research and Practice* 42, 505–512. doi: 10.1037/a0024485

Lyles, A. A., Amresh, A., Huberty, J., Todd, M., and Lee, R. E. (2017). A Mobile, Avatar-Based App for Improving Body Perceptions Among Adolescents: A Pilot Test. *JMIR serious games* 5, e4. doi: 10.2196/games.6354

Lynch, M. E., Campbell, F., Clark, A. J., Dunbar, M. J., Goldstein, D., Peng, P., et al. (2008). A systematic review of the effect of waiting for treatment for chronic pain. *Pain* 136, 97–116. doi: 10.1016/j.pain.2007.06.018

Macea, D. D., Gajos, K., Daglia Calil, Y. A., and Fregni, F. (2010). The efficacy of Web-based cognitive behavioral interventions for chronic pain: a systematic review and meta-analysis. *The journal of pain: official journal of the American Pain Society* 11, 917–929. doi: 10.1016/j.jpain.2010.06.005

Maddox, M. M., Lopez, A., Mandava, S. H., Boonjindasup, A., Viriyasiripong, S., Silberstein, J. L., et al. (2015). Electroencephalographic Monitoring of Brain Wave Activity During Laparoscopic Surgical Simulation to Measure Surgeon Concentration and Stress: Can the Student Become the Master? *Journal of endourology* 29, 1329–1333. doi: 10.1089/end.2015.0239

Malarkey, W. B., and Mills, P. J. (2007). Endocrinology: the active partner in PNI research. *Brain, behavior, and immunity* 21, 161–168. doi: 10.1016/j.bbi.2006.10.008

Marcengo, A., and Rapp, A. (2014). "Visualization of Human Behavior Data," in *Innovative Approaches of Data Visualization and Visual Analytics 2014*, eds. D. Taniar, M. L. Huang, and W. Huang (IGI Global), 236–265

Mason, L. J. (2001). *Guide to stress reduction*. Berkeley, California: Celestial Arts

Masterson Creber, R. M., Maurer, M. S., Reading, M., Hiraldo, G., Hickey, K. T., and Iribarren, S. (2016). Review and Analysis of Existing Mobile Phone Apps to Support Heart Failure Symptom Monitoring and Self-Care Management Using the Mobile Application Rating Scale (MARS). *JMIR mHealth and uHealth* 4, e74. doi: 10.2196/mhealth.5882

Matthews, R., Turner, P. J., McDonald, N. J., Ermolaev, K., Manus, T., Shelby, R. A., et al. (2008). "Real time workload classification from an ambulatory wireless EEG system using hybrid EEG electrodes," in *30th Annual International IEEE EMBS Conference 2008* (IEEE), 5871–5875

McDowell, K., Chin-Teng, L., Oie, K. S., Tzyy-Ping, J., Gordon, S., Whitaker, K. W., et al. (2013). Real-World Neuroimaging Technologies. *IEEE Access* 1, 131–149. doi: 10.1109/ACCESS.2013.2260791

McFarland, D. J., and Wolpaw, J. R. (2011). Brain-Computer Interfaces for Communication and Control. *Communications of the ACM* 54, 60–66. doi: 10.1145/1941487.1941506

McKeown, S., Krause, C., Shergill, M., Siu, A., and Sweet, D. (2016). Gamification as a strategy to engage and motivate clinicians to improve care. *Healthcare management forum* 29, 67–73. doi: 10.1177/0840470415626528

Meingast, M., Roosta, T., and Sastry, S. (2006). "Security and privacy issues with health care information technology," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2006* (IEEE), 5453–5458

Melnik, A., Legkov, P., Izdebski, K., Kärcher, S. M., Hairston, W. D., Ferris, D. P., et al. (2017). Systems, Subjects, Sessions: To What Extent Do These Factors Influence EEG Data? *Frontiers in human neuroscience* 11, 18200. doi: 10.3389/fnhum.2017.00150

Mihajlovic, V., Grundlehner, B., Vullers, R., and Penders, J. (2015). Wearable, wireless EEG solutions in daily life applications: what are we missing? *IEEE journal of biomedical and health informatics* 19, 6–21. doi: 10.1109/JBHI.2014.2328317

Miller, G. (2012). The Smartphone Psychology Manifesto. *Perspectives on psychological science: a journal of the Association for Psychological Science* 7, 221–237. doi: 10.1177/1745691612441215

Molander, P., Dong, H.-J., Äng, B., Enthoven, P., and Gerdle, B. (2018). The role of pain in chronic pain patients' perception of health-related quality of life: a cross-sectional SQRP study of 40,000 patients. *Scandinavian journal of pain* 18, 417–429. doi: 10.1515/sjpain-2018-0003

Mornhinweg, G. C. (1992). Effects of music preference and selection on stress reduction. *Journal of holistic nursing: official journal of the American Holistic Nurses' Association* 10, 101–109. doi: 10.1177/089801019201000202

Morris, M. E., Kathawala, Q., Leen, T. K., Gorenstein, E. E., Guilak, F., Labhard, M., et al. (2010). Mobile Therapy: Case Study Evaluations of a Cell Phone Application for Emotional Self-Awareness. *Journal of medical Internet research* 12, e10. doi: 10.2196/jmir.1371

Mosconi, P., Radrezza, S., Lettieri, E., and Santoro, E. (2019). Use of Health Apps and Wearable Devices: Survey Among Italian Associations for Patient Advocacy. *JMIR mHealth and uHealth* 7, e10242. doi: 10.2196/10242

MPG (2019). *Bundesministerium der Justiz und für Verbraucherschutz*, https://www.gesetze-im-internet.de/mpg/MPG.pdf

Müller, K.-R., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., and Blankertz, B. (2008). Machine learning for real-time single-trial EEG-analysis: from brain-computer interfacing to mental state monitoring. *Journal of neuroscience methods* 167, 82–90. doi: 10.1016/j.jneumeth.2007.09.022

Murphy, L. R. (1996). Stress management in work settings: a critical review of the health effects. *American journal of health promotion* 11, 112–135. doi: 10.4278/0890-1171-11.2.112

Murray, E., Burns, J., See, T. S., Lai, R., and Nazareth, I. (2005). Interactive Health Communication Applications for people with chronic disease. *The Cochrane database of systematic reviews*, CD004274. doi: 10.1002/14651858.CD004274.pub4

Naparstek, B. (1995). *Staying Well With Guided Imagery: How to Harness the Power of Your Imagination for Healing Health*. Warner Books

Nawrot, P., Jordan, S., Eastwood, J., Rotstein, J., Hugenholtz, A., and Feeley, M. (2003). Effects of caffeine on human health. *Food additives and contaminants* 20, 1–30. doi: 10.1080/0265203021000007840

NeuroSky. *Body and Mind. Quantified*, http://neurosky.com/

Nielsen, J. (1994). *Usability engineering*. San Francisco, California: Morgan Kaufmann Publishers

Nowak, K. L., and Rauh, C. (2005). The Influence of the Avatar on Online Perceptions of Anthropomorphism, Androgyny, Credibility, Homophily, and Attraction. *Journal of Computer-Mediated Communication* 11, 153–178. doi: 10.1111/j.1083-6101.2006.tb00308.x

Oinas-Kukkonen, H., and Harjumaa, M. (2009). Persuasive Systems Design: Key Issues, Process Model, and System Features. *Communications of the Association for Information Systems* 24. doi: 10.17705/1CAIS.02428

Olff, M. (2015). Mobile mental health: a challenging research agenda. *European journal of psychotraumatology* 6, 27882. doi: 10.3402/ejpt.v6.27882

O'Reilly, G. A., and Spruijt-Metz, D. (2013). Current mHealth technologies for physical activity assessment and promotion. *American journal of preventive medicine* 45, 501–507. doi: 10.1016/j.amepre.2013.05.012

Oyebode, O., Gordon-Dseagu, V., Walker, A., and Mindell, J. S. (2014). Fruit and vegetable consumption and all-cause, cancer and CVD mortality: analysis of Health Survey for England data. *Journal of epidemiology and community health* 68, 856–862. doi: 10.1136/jech-2013-203500

Pagano, R. R. (2010). *Understanding statistics in the behavioral sciences*. Belmont, Australia: Wadsworth Cengage Learning

Parks, P., Cruz, R., and Ahn, S. J. (2014). Don't Hurt My Avatar: The Use and Potential of Digital Self-Representation in Risk Communication. *International Journal of Robots, Education and Art* 4, 10–20

Payne, H. E., Wilkinson, J., West, J. H., and Bernhardt, J. M. (2016). A content analysis of precede-proceed constructs in stress management mobile apps. *mHealth* 2, 5. doi: 10.3978/j.issn.2306-9740.2016.02.02

Peng, P., Choiniere, M., Dion, D., Intrater, H., Lefort, S., Lynch, M., et al. (2007a). Challenges in accessing multidisciplinary pain treatment facilities in Canada. *Canadian journal of anaesthesia* 54, 977–984. doi: 10.1007/BF03016631

Peng, P., Stinson, J. N., Choiniere, M., Dion, D., Intrater, H., Lefort, S., et al. (2007b). Dedicated multidisciplinary pain management centres for children in Canada: the current status. *Canadian journal of anaesthesia* 54, 985–991. doi: 10.1007/BF03016632

Pereira, P., Duarte, E., Rebelo, F., and Noriega, P. (2014). "A Review of Gamification for Health-Related Contexts," in *Design, User Experience, and Usability. User Experience Design for Diverse Interaction Platforms* *and Environments 2014*, eds. D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, A. Kobsa, F. Mattern, et al. (Cham: Springer), 742–753

Pérez, C., Margarit, C., Sánchez-Magro, I., de Antonio, A., and Villoria, J. (2017). Chronic Pain Features Relate to Quality of Life More than Physiopathology: A Cross-Sectional Evaluation in Pain Clinics. *Pain practice: the official journal of World Institute of Pain* 17, 866–878. doi: 10.1111/papr.12533

Pilcher, J. J., Ginter, D. R., and Sadowsky, B. (1997). Sleep quality versus sleep quantity: Relationships between sleep and measures of health, well-being and sleepiness in college students. *Journal of psychosomatic research* 42, 583–596. doi: 10.1016/s0022-3999(97)00004-4

Portelli, P., and Eldred, C. (2016). A quality review of smartphone applications for the management of pain. *British journal of pain* 10, 135–140. doi: 10.1177/2049463716638700

Proudfoot, J., Parker, G., Hadzi Pablovic, D., Manicavasagar, V., Adler, E., and Whitton, A. (2010). Community Attitudes to the Appropriation of Mobile Phones for Monitoring and Managing Depression, Anxiety, and Stress. *Journal of medical Internet research* 12, e64. doi: 10.2196/jmir.1475

Przegalinska, A., Ciechanowski, L., Magnuski, M., and Gloor, P. (2018). “Muse Headband: Measuring Tool or a Collaborative Gadget?” in *Collaborative Innovation Networks 2018*, eds. F. Grippa, J. Leitão, J. Gluesing, K. Riopelle, and P. Gloor (Cham: Springer), 93–101

Rabbi, M., Pfammatter, A., Zhang, M., Spring, B., and Choudhury, T. (2015). Automated personalized feedback for physical activity and dietary behavior change with mobile phones: a randomized controlled trial on adults. *JMIR mHealth and uHealth* 3, e42. doi: 10.2196/mhealth.4160

Rasche, P., Wille, M., Bröhl, C., Theis, S., Schäfer, K., Knobe, M., et al. (2018). Prevalence of Health App Use Among Older Adults in Germany: National Survey. *JMIR mHealth and uHealth* 6, e26. doi: 10.2196/mhealth.8619

Ray, W., and Cole, H. (1985). EEG alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes. *Science* 228, 750–752. doi: 10.1126/science.3992243

Rebolledo-Mendez, G., Dunwell, I., Martínez-Mirón, E. A., Vargas-Cerdán, M. D., de Freitas, S., Liarokapis, F., et al. (2009). “Assessing NeuroSky’s Usability to Detect Attention Levels in an Assessment Exercise,” in *HCI international: 13th international conference 2009*, ed. J. A. Jacko (Berlin: Springer), 149–158

Rieiro, H., Diaz-Piedra, C., Morales, J. M., Catena, A., Romero, S., Roca-Gonzalez, J., et al. (2019). Validation of Electroencephalographic Recordings Obtained with a Consumer-Grade, Single Dry Electrode, Low-Cost Device: A Comparative Study. *Sensors* 19. doi: 10.3390/s19122808

Rivera, J., McPherson, A., Hamilton, J., Birken, C., Coons, M., Iyer, S., et al. (2016). Mobile Apps for Weight Management: A Scoping Review. *JMIR mHealth and uHealth* 4, e87. doi: 10.2196/mhealth.5115

Roberts, B. W., Walton, K. E., and Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: a meta-analysis of longitudinal studies. *Psychological bulletin* 132, 1–25. doi: 10.1037/0033-2909.132.1.1

Rohit, F., Kulathumani, V., Kavi, R., Elwarfalli, I., Kecojevic, V., and Nimbarte, A. (2017). Real-time drowsiness detection using wearable, lightweight brain sensing headbands. *IET Intelligent Transport Systems* 11, 255–263. doi: 10.1049/iet-its.2016.0183

Rosser, B. A., and Eccleston, C. (2011). Smartphone applications for pain management. *Journal of telemedicine and telecare* 17, 308–312. doi: 10.1258/jtt.2011.101102

Russ, T. C., Stamatakis, E., Hamer, M., Starr, J. M., Kivimäki, M., and Batty, G. D. (2012). Association between psychological distress and mortality: individual participant pooled analysis of 10 prospective cohort studies. *BMJ* 345, e4933. doi: 10.1136/bmj.e4933

Ruzic, L., and Sanford, J. A. (2018). Needs Assessment—mHealth Applications for People Aging with Multiple Sclerosis. *Journal of Healthcare Informatics Research* 2, 71–98. doi: 10.1007/s41666-018-0023-z

Sailer, M., Hense, J. U., Mayr, S. K., and Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior* 69, 371–380. doi: 10.1016/j.chb.2016.12.033

Sardi, L., Idri, A., and Fernández-Alemán, J. L. (2017). A systematic review of gamification in e-Health. *Journal of biomedical informatics* 71, 31–48. doi: 10.1016/j.jbi.2017.05.011

Sauro, J. (2011). *A practical guide to the system usability scale: Background, benchmarks & best practices*. Denver, Colorado: Measuring Usability LCC

Sax, M., Helberger, N., and Bol, N. (2018). Health as a Means Towards Profitable Ends: mHealth Apps, User Autonomy, and Unfair Commercial Practices. *Journal of Consumer Policy* 41, 103–134. doi: 10.1007/s10603-018-9374-3

Scheufele, P. M. (1999). Effects of Progressive Relaxation and Classical Music on Measurements of Attention, Relaxation, and Stress Responses. *Journal of behavioral medicine*

Schrepp, M. (2015). User Experience Questionnaire Handbook. *All you need to know to apply the UEQ successfully in your project*. doi: 10.13140/RG.2.1.2815.0245

Shaw, J. C. (2003). *The brain's alpha rhythms and the mind: A review of classical and modern studies of the alpha rhythm component of the electroencephalogram with commentaries on associated neuroscience and neuropsy-chology*. Elsevier

Shephard, R. J. (2001). Absolute versus relative intensity of physical activity in a dose-response context. *Medicine and Science in Sports and Exercise* 33, 400-418. doi: 10.1097/00005768-200106001-00008

Smith, A. (2013). *Smartphone Ownership — 2013 Update*, https://www.pewinternet.org/2013/06/05/smartphone-ownership-2013/

Solbrig, L., Jones, R., Kavanagh, D., May, J., Parkin, T., and Andrade, J. (2017). People trying to lose weight dislike calorie counting apps and want motivational support to help them achieve their goals. *Internet Interventions* 7, 23–31. doi: 10.1016/j.invent.2016.12.003

Spil, T., Sunyaev, A., Thiebes, S., and van Baalen, R. (2017). "The Adoption of Wearables for a Healthy Lifestyle: Can Gamification Help?" in *Proceedings of the 50th Hawaii International Conference on System Sciences 2017* (Honolulu: HICSS)

Stansfeld, S., and Candy, B. (2006). Psychosocial work environment and mental health - a meta-analytic review. *Scandinavian journal of work, environment & health* 32, 443–462. doi: 10.5271/sjweh.1050

Statista (2019a). *Anzahl der Smartphone-Nutzer in Deutschland in den Jahren 2009 bis 2018 (in Millionen)*, https://de.statista.com/statistik/daten/studie/198959/umfrage/anzahl-der-smartphonenutzer-in-deutschland-seit-2010//

Statista (2019b). *Market share comparison of Android and iOS in smartphone sales in Germany from March 2012 to March 2019*. Accessed July 18, 2019, https://www.statista.com/statistics/461900/android-vs-ios-market-share-in-smartphone-sales-germany/

Statista (2019c). *Number of mHealth app downloads worldwide from 2013 to 2017 (in billions)*, https://www.statista.com/statistics/625034/mobile-health-app-downloads/

Stead, W. W., Haynes, R. B., Fuller, S., Friedman, C. P., Travis, L. E., Beck, J. R., et al. (1994). Designing medical informatics research and library--resource projects to increase what is learned. *Journal of the American Medical Informatics Association* 1, 28–33. doi: 10.1136/jamia.1994.95236134

Stinson, J., McGrath, P., Hodnett, E., Feldman, B., Duffy, C., Huber, A., et al. (2010). Usability testing of an online self-management program for adolescents with juvenile idiopathic arthritis. *Journal of medical Internet research* 12, e30. doi: 10.2196/jmir.1349

Stinson, J., White, M., Isaac, L., Campbell, F., Brown, S., Ruskin, D., et al. (2013). Understanding the information and service needs of young adults with chronic pain: perspectives of young adults and their providers. *The Clinical journal of pain* 29, 600–612. doi: 10.1097/AJP.0b013e31826dce65

Stinson, J. N., Jibb, L. A., Nguyen, C., Nathan, P. C., Maloney, A. M., Dupuis, L. L., et al. (2013). Development and testing of a multidimensional iPhone pain assessment application for adolescents with cancer. *Journal of medical Internet research* 15, e51. doi: 10.2196/jmir.2350

Stinson, J. N., Lalloo, C., Harris, L., Isaac, L., Campbell, F., Brown, S., et al. (2014). iCanCope with Pain™: User-centred design of a web- and mobile-based self-management program for youth with chronic pain based on identified health care needs. *Pain research & management* 19, 257–265

Stone, A. A., Shiffman, S., Schwartz, J. E., Broderick, J. E., and Hufford, M. R. (2002). Patient non-compliance with paper diaries. *BMJ* 324, 1193–1194. doi: 10.1136/bmj.324.7347.1193

Stoyanov, S. R., Hides, L., Kavanagh, D. J., and Wilson, H. (2016). Development and Validation of the User Version of the Mobile Application Rating Scale (uMARS). *JMIR mHealth and uHealth* 4, e72. doi: 10.2196/mhealth.5849

Stoyanov, S. R., Hides, L., Kavanagh, D. J., Zelenko, O., Tjondronegoro, D., and Mani, M. (2015). Mobile app rating scale: a new tool for assessing the quality of health mobile apps. *JMIR mHealth and uHealth* 3, e27. doi: 10.2196/mhealth.3422

Sullivan, T. J., Deiss, S. R., and Cauwenberghs, G. (2007). "A Low-Noise, Non-Contact EEG/ECG Sensor," in *IEEE Biomedical Circuits and Systems Conference 2007* (IEEE), 154–157

Taylor, S., Ferguson, C., Peng, F., Schoeneich, M., and Picard, R. W. (2019). Use of In-Game Rewards to Motivate Daily Self-Report Compliance: Randomized Controlled Trial. *Journal of medical Internet research* 21, e11683. doi: 10.2196/11683

Thiel, S.-K., Reisinger, M., and Röderer, K. (2016). ""I'm too old for this!": Influence of Age on Perception of Gamified Public Participation," in *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia 2016*, ed. F. Alt (New York, New York, USA: ACM Press), 343–346

Thiele, C., Laireiter, A.-R., and Baumann, U. (2002). Diaries in clinical psychology and psychotherapy: a selective review. *Clinical Psychology & Psychotherapy* 9, 1–37. doi: 10.1002/cpp.302

Thompson, D. V., Hamilton, R. W., and Rust, R. T. (2005). Feature Fatigue: When Product Capabilities Become Too Much of a Good Thing. *Journal of Marketing Research* 42, 431–442. doi: 10.1509/jmkr.2005.42.4.431

Thorpe, A. S., and Roper, S. (2017). The Ethics of Gamification in a Marketing Context. *Journal of Business Ethics* 23, 123. doi: 10.1007/s10551-017-3501-y

Tudor-Locke, C., and Bassett, D. R. (2004). How many steps/day are enough? Preliminary pedometer indices for public health. *Sports medicine* 34, 1–8. doi: 10.2165/00007256-200434010-00001

van Boxtel, G. J. M., Denissen, A. J. M., Jäger, M., Vernon, D., Dekker, M. K. J., Mihajlović, V., et al. (2012). A novel self-guided approach to alpha activity training. *International journal of psychophysiology: official journal of the International Organization of Psychophysiology* 83, 282–294. doi: 10.1016/j.ijpsycho.2011.11.004

van Velsen, L., Evers, M., Bara, C.-D., Op den Akker, H., Boerema, S., and Hermens, H. (2018). Understanding the Acceptance of an eHealth Technology in the Early Stages of Development: An End-User Walkthrough Approach and Two Case Studies. *JMIR formative research* 2, e10474. doi: 10.2196/10474

van Vugt, H. C., Konijn, E. A., Hoorn, J. F., and Veldhuis, J. (2009). When too heavy is just fine: Creating trustworthy e-health advisors. *International Journal of Human-Computer Studies* 67, 571–583. doi: 10.1016/j.ijhcs.2009.02.005

Vandelanotte, C., Spathonis, K. M., Eakin, E. G., and Owen, N. (2007). Website-Delivered Physical Activity Interventions. *American journal of preventive medicine* 33, 54–64. doi: 10.1016/j.amepre.2007.02.041

Varvogli, L., and Darviri, C. (2011). Stress Management Techniques: evidence-based procedures that reduce stress and promote health. *Health Science Journal* 5, 74–89

Venkatesh, V., and Bala, H. (2008). Technology Acceptance Model 3 and a Research Agenda on Interventions. *Decision Sciences* 39, 273–315. doi: 10.1111/j.1540-5915.2008.00192.x

Ventä, L., Isomursu, M., Ahtinen, A., and Ramiah, S. (2008). ""My Phone is a Part of My Soul" - How People Bond with Their Mobile Phones," in *Proceedings of the 2nd International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies 2008* (IEEE), 311–317

Vincent, W. J. (2008). *Statistics in kinesiology*. Human Kinetics

Virzi, R. A. (1992). Refining the Test Phase of Usability Evaluation: How Many Subjects Is Enough? *Human Factors: The Journal of the Human Factors and Ergonomics Society* 34, 457–468. doi: 10.1177/001872089203400407

Wallace, L. S., and Dhingra, L. K. (2014). A systematic review of smartphone applications for chronic pain available for download in the United States. *Journal of opioid management* 10, 63–68. doi: 10.5055/jom.2014.0193

Wang, J., Yao, N. A., Liu, Y., Geng, Z., Wang, Y., Shen, N., et al. (2017). Development of a Smartphone Application to Monitor Pediatric Patient-Reported Outcomes. *Computers, informatics, nursing* 35, 590–598. doi: 10.1097/CIN.0000000000000357

Wantland, D. J., Portillo, C. J., Holzemer, W. L., Slaughter, R., and McGhee, E. M. (2004). The effectiveness of Web-based vs. non-Web-based interventions: a meta-analysis of behavioral change outcomes. *Journal of medical Internet research* 6, e40. doi: 10.2196/jmir.6.4.e40

Webber, K. H., Tate, D. F., Ward, D. S., and Bowling, J. M. (2010). Motivation and Its Relationship to Adherence to Self-monitoring and Weight Loss in a 16-week Internet Behavioral Weight Loss Intervention. *Journal of Nutrition Education and Behavior* 42, 161–167. doi: 10.1016/j.jneb.2009.03.001

Weber, K. (2015). "MEESTAR: Ein Modell zur ethischen Evaluierung sozio-technischer Arrangements in der Pflege- und Gesundheitsversorgung," in *Technisierung des Alltags – Beitrag für ein gutes Leben?* eds. K. Weber, D. Frommeld, A. Manzeschke, and H. Fangerau (Franz Steiner Verlag)

Weigensberg, M. J., Lane, C. J., Winners, O., Wright, T., Nguyen-Rodriguez, S., Goran, M. I., et al. (2009). Acute effects of stress-reduction Interactive Guided Imagery (SM) on salivary cortisol in overweight Latino adolescents. *Journal of alternative and complementary medicine* 15, 297–303. doi: 10.1089/acm. 2008.0156

Wharton, C. M., Johnston, C. S., Cunningham, B. K., and Sterner, D. (2014). Dietary self-monitoring, but not dietary quality, improves with use of smartphone app technology in an 8-week weight loss trial. *Journal of Nutrition Education and Behavior* 46, 440–444. doi: 10.1016/j.jneb.2014.04.291

Wiegner, L., Hange, D., Björkelund, C., and Ahlborg, G. (2015). Prevalence of perceived stress and associations to symptoms of exhaustion, depression and anxiety in a working age population seeking primary care - an observational study. *BMC Family Practice* 16, 78. doi: 10.1186/s12875-015-0252-7

Wilcox, R. R. (2013). *Introduction to Robust Estimation and Hypothesis Testing*. San Diego, California, USA: Elsevier Science & Technology Books

Wilkowska, W., and Ziefle, M. (2012). Privacy and data security in E-health: requirements from the user's perspective. *Health informatics journal* 18, 191–201. doi: 10.1177/1460458212442933

Williams, K. N., and Kemper, S. (2010). Interventions to reduce cognitive decline in aging. *Journal of psychosocial nursing and mental health services* 48, 42–51. doi: 10.3928/02793695-20100331-03

Wilson, G. F., and Russell, C. A. (2003). Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. *Human factors* 45, 635–643. doi: 10.1518/hfes.45.4.635.27088

World Health Organization (2011). *mHealth: New horizons for health through mobile technologies*, http://www.who.int/goe/publications/goe_mhealth_web.pdf

World Medical Association (2013). Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 310, 2191–2194. doi: 10.1001/jama.2013.281053

Zagel, C., and Bodendorf, F. (2014). "Gamification: Auswirkungen auf Usability, Datenqualität und Motivation," in *Tagungsband: Mensch und Computer 2014*, eds. M. Koch, A. Butz, and J. Schlichter (Oldenbourg Wissenschaftsverlag), 15–24

Zander, T. O., and Kothe, C. (2011). Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general. *Journal of neural engineering* 8, 25005. doi: 10. 1088/1741-2560/8/2/025005

Zichermann, G., and Cunningham, C. (2011). *Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps*. O'Reilley

Appendix

Appendix 1: Main effects and interactions for five-factor ANOVA with factors (channel, frequency band, session, task, and device).

Factors	**df1**	**df2**	***F***	***p***	**Partial η^2**
Channel	1.51	27.18	15.45	≤ .001	.46
Frequency Band	1.00	18.00	5.74	.03	.24
Session	1.00	18.00	0.86	.37	.05
Task	1.60	28.82	18.33	≤ .001	.51
Device	1.00	18.00	22.81	≤ .001	.56
Channel * Frequency Band	1.21	21.76	36.48	≤ .001	.67
Channel * Session	2.63	47.37	1.23	.31	.06
Frequency Band * Session	1.00	18.00	0.52	.48	.03
Channel * Frequency Band * Session	1.70	30.58	0.14	.84	.01
Channel * Task	2.25	40.57	29.37	≤ .001	.62
Frequency Band * Task	1.71	30.74	11.96	≤ .001	.40
Channel * Frequency Band * Task	2.05	36.93	3.05	.06	.15
Session * Task	2.51	45.24	0.66	.56	.04
Channel * Session * Task	4.49	80.87	1.48	.21	.08
Frequency Band * Session * Task	2.48	44.66	0.51	.65	.03
Channel * Frequency Band * Session * Task	3.57	64.17	0.55	.68	.03
Channel * Device	2.88	51.83	11.09	≤ .001	.38
Frequency Band * Device	1.00	18.00	9.28	.007	.34
Channel * Frequency Band * Device	1.57	28.21	0.31	.69	.02
Session * Device	1.00	18.00	0.03	.88	.00
Channel * Session * Device	2.37	42.61	0.76	.49	.04
Frequency Band * Session * Device	1.00	18.00	1.81	.20	.09
Channel * Frequency Band * Session * Device	1.88	33.78	0.43	.64	.02
Task * Device	2.32	41.75	45.16	≤ .001	.72
Channel * Task * Device	3.89	69.98	11.03	≤ .001	.38
Frequency Band * Task * Device	1.65	29.76	33.16	≤ .001	.65
Channel * Frequency Band * Task * Device	3.88	69.87	11.74	≤ .001	.40
Session * Task * Device	2.57	46.24	0.47	.68	.03
Channel * Session * Task * Device	4.36	78.51	0.79	.55	.04
Frequency Band * Session * Task * Device	1.97	35.39	0.21	.81	.01
Channel * Frequency Band * Session * Task * Device	5.30	95.33	0.93	.47	.05

Appendix 2: t-test results for the interaction channel * frequency band.

t-test pairs	***t***	**df**	***p***	**Cohen's *d***	**Sign. after Bonferroni adjustment**
Fp1_alpha - Fp2_alpha	-1.16	18	.26	0.27	
Fp1_alpha - A1_alpha	-6.80	18	≤ .001	1.56	*
Fp1_alpha - A2_alpha	-6.07	18	≤ .001	1.39	*

t-test pairs	*t*	df	*p*	Cohen's *d*	Sign. after Bonferroni adjustment
Fp2_alpha - A1_alpha	-6.23	18	≤ .001	1.43	*
Fp2_alpha - A2_alpha	-5.80	18	≤ .001	1.33	*
A1_alpha - A2_alpha	-1.07	18	.30	0.25	
Fp1_beta - Fp2_beta	-0.82	18	.43	0.19	
Fp1_beta - A1_beta	1.27	18	.22	0.29	
Fp1_beta - A2_beta	2.43	18	.03	0.56	
Fp2_beta - A1_beta	1.95	18	.07	0.45	
Fp2_beta - A2_beta	3.28	18	.004	0.75	*
A1_beta - A2_beta	3.85	18	.001	0.88	*

Appendix 3: t-test results for the interaction channel * task.

t-test pairs	*t*	df	*p*	Cohen's *d*	Sign. after Bonferroni adjustment
Fp1_eyes open - Fp1_eyes closed	0.52	18	.61	0.12	
Fp1_eyes open - Fp1_brainstorming	1.67	18	.11	0.38	
Fp1_eyes open - Fp1_breathing	1.09	18	.29	0.25	
Fp1_eyes closed - Fp1_brainstorming	2.44	18	.03	0.56	
Fp1_eyes closed - Fp1_breathing	1.04	18	.31	0.24	
Fp1_brainstorming - Fp1_breathing	-0.97	18	.35	0.22	
Fp2_eyes open - Fp2_eyes closed	-0.61	18	.55	0.14	
Fp2_eyes open - Fp2_brainstorming	0.343	18	.74	0.08	
Fp2_eyes open - Fp2_breathing	-0.12	18	.91	0.03	
Fp2_eyes closed - Fp2_brainstorming	2.00	18	.06	0.46	
Fp2_eyes closed - Fp2_breathing	0.86	18	.40	0.20	
A2_eyes open - A2_eyes closed	-6.96	18	≤ .001	1.60	*
A2_eyes open - A2_brainstorming	-4.59	18	≤ .001	1.05	*
A2_eyes open - A2_breathing	-5.54	18	≤ .001	1.27	*
A2_eyes closed - A2_brainstorming	6.94	18	≤ .001	1.59	*
A2_eyes closed - A2_breathing	1.46	18	.16	0.34	
A2_breathing - A2_brainstorming	4.47	18	≤ .001	1.03	*
A1_eyes open - A1_eyes closed	-6.75	18	≤ .001	1.55	*
A1_eyes open - A1_brainstorming	-4.36	18	≤ .001	1.00	*
A1_eyes open - A1_breathing	-6.24	18	≤ .001	1.43	*
A1_eyes closed - A1_brainstorming	5.07	18	≤ .001	1.16	*
A1_eyes closed - A1_breathing	-0.94	18	.36	0.22	
A1_brainstorming - A1_breathing	-5.72	18	≤ .001	1.31	*

Appendix 4: t-test results for the interaction frequency band * task.

t-test pairs	*t*	df	*p*	Cohen's *d*	Sign. after Bonferroni adjustment
eyes open_alpha - eyes closed_alpha	-6.75	18	≤ .001	1.55	*
eyes open_alpha - breathing_alpha	-5.95	18	≤ .001	1.36	*

t-test pairs	***t***	**df**	***p***	**Cohen's *d***	**Sign. after Bonferroni adjustment**
eyes open_alpha - brainstorming_alpha	-4.58	18	≤ .001	1.05	*
eyes closed_alpha - breathing_alpha	0.97	18	.34	0.22	
eyes closed_alpha - brainstorming_alpha	5.12	18	≤ .001	1.17	*
breathing_alpha - brainstorming_alpha	4.61	18	≤ .001	1.06	*
eyes open_beta - eyes closed_beta	-1.39	18	.18	0.32	
eyes open_beta - breathing_beta	-1.07	18	.30	0.25	
eyes open_beta - brainstorming_beta	0.15	18	.88	0.03	
eyes closed_beta - breathing_beta	0.67	18	.51	0.15	
eyes closed_beta - brainstorming_beta	3.76	18	.001	0.86	*
breathing_beta - brainstorming_beta	2.26	18	.04	0.52	

Appendix 5: t-test results for the interaction channel * device.

t-test pairs	***t***	**df**	***p***	**Cohen's *d***	**Sign. after Bonferroni adjustment**
wearable_Fp1 - gel-based_Fp1	1.51	18	.15	0.35	
wearable_Fp2 - gel-based_Fp2	1.06	18	.30	0.24	
wearable_A1 – gel-based_A1	8.41	18	≤ .001	1.93	*
wearable_A2 – gel-based_A2	5.69	18	≤ .001	1.31	*

Appendix 6: t-test results for the interaction frequency band * device.

t-test pairs	***t***	**df**	***p***	**Cohen's *d***	**Sign. after Bonferroni adjustment**
wearable_alpha – gel-based_alpha	3.55	18	.002	0.81	*
wearable_beta – gel-based_beta	4.53	18	≤ .001	1.04	*

Appendix 7: t-test results for the interaction task * device.

t-test pairs	***t***	**df**	***p***	**Cohen's *d***	**Sign. after Bonferroni adjustment**
wearable_eyes open – gel-based_eyes open	-3.59	18	.002	0.82	*
wearable_eyes closed – gel-based_eyes closed	6.49	18	≤ .001	1.49	*
wearable_breathing – gel-based_breathing	6.88	18	≤ .001	1.58	*
wearable_brainstorming – gel-based_brainstorming	5.55	18	≤ .001	1.27	*

Appendix 8: t-test results for the interaction channel * task * device.

t-test pairs	***t***	**df**	***p***	**Cohen's *d***	**Sign. after Bonferroni adjustment**
wearable_eyes open_Fp1 – gel-based_eyes open_Fp1	-5.09	18	≤ .001	1.17	*
wearable_eyes open_Fp2 – gel-based_eyes open_Fp2	-5.73	18	≤ .001	1.31	*
wearable_eyes open_TP9 – gel-based_eyes open_TP9	3.42	18	.003	0.78	*

t-test pairs	*t*	df	*p*	Cohen's *d*	Sign. after Bonferroni adjustment
wearable_eyes open_TP10 – gel-based_eyes open_TP10	3.86	18	≤ .001	0.89	*
wearable_eyes closed_Fp1 – gel-based_eyes closed_Fp1	4.06	18	.001	0.93	*
wearable_eyes closed_Fp2 – gel-based_eyes closed_Fp2	4.04	18	.001	0.93	*
wearable_eyes closed_TP9 – gel-based_eyes closed_TP9	3.02	18	.007	0.69	*
wearable_eyes closed_TP10 – gel-based_eyes closed_TP10	5.16	18	≤ .001	1.18	*
wearable_breathing_Fp1 – gel-based_breathing_Fp1	4.93	18	≤ .001	1.13	*
wearable_breathing_Fp2 – gel-based_breathing_Fp2	4.62	18	≤ .001	1.06	*
wearable_breathing_TP9 – gel-based_breathing_TP9	7.99	18	≤ .001	1.83	*
wearable_breathing_TP10 – gel-based_breathing_TP10	3.00	18	.008	0.69	*
wearable_brainstorming_Fp1 – gel-based_brainstorming_Fp1	3.58	18	.002	0.82	*
wearable_brainstorming_Fp2 – gel-based_brainstorming_Fp2	2.77	18	.01	0.64	*
wearable_brainstorming_TP9 – gel-based_brainstorming_TP9	5.86	18	≤ .001	1.34	*
wearable_brainstorming_TP10 – gel-based_brainstorming_TP10	3.55	18	.002	0.81	*
wearable_eyes open_Fp1 – wearable_eyes open_Fp2	0.41	18	.69	0.09	
wearable_eyes open_Fp1 – wearable_eyes open_TP9	-1.27	18	.22	0.29	
wearable_eyes open_Fp1 - wearable_eyes open_TP10	-0.58	18	.57	0.13	
wearable_eyes open_Fp2 – wearable_eyes open_TP9	-1.54	18	.14	0.35	
wearable_eyes open_Fp2 – wearable_eyes open_TP10	-0.85	18	.40	0.20	
wearable_eyes closed_Fp1 – wearable_eyes closed_Fp2	-0.74	18	.47	0.17	
wearable_eyes closed_Fp1 – wearable_eyes closed_TP9	-4.68	18	≤ .001	1.07	*
wearable_eyes closed_Fp1 – wearable_eyes closed_TP10	-5.74	18	≤ .001	1.32	*
wearable_eyes closed_Fp2 – wearable_eyes closed_TP9	-3.92	18	≤ .001	0.90	*
wearable_eyes closed_Fp2 – wearable_eyes closed_TP10	-4.93	18	≤ .001	1.13	*

t-test pairs	*t*	df	*p*	Cohen's *d*	Sign. after Bonferroni adjustment
wearable_eyes closed_TP9 – wearable_eyes closed_TP10	-0.82	18	.42	0.19	
wearable_breathing_Fp1 – wearable_breathing_Fp2	-0.54	18	.60	0.12	
wearable_breathing_Fp1 – wearable_breathing_TP9	-6.64	18	≤ .001	1.52	*
wearable_breathing_Fp1 – wearable_breathing_TP10	-4.78	18	≤ .001	1.10	*
wearable_breathing_Fp2 – wearable_breathing_TP9	-5.36	18	≤ .001	1.23	*
wearable_breathing_Fp2 – wearable_breathing_TP10	-4.48	18	≤ .001	1.03	*
wearable_breathing_TP9 – wearable_breathing_TP10	2.13	18	.05	0.49	
wearable_brainstorming_Fp1 – wearable_brainstorming_Fp2	-0.79	18	.44	0.18	
wearable_brainstorming_Fp1 – wearable_brainstorming_TP9	-4.43	18	≤.000	1.02	*
wearable_brainstorming_Fp1 – wearable_brainstorming_TP10	-3.88	18	.001	0.89	*
wearable_brainstorming_Fp2 – wearable_brainstorming_TP9	-2.61	18	.02	0.60	
wearable_brainstorming_Fp2 – wearable_brainstorming_TP10	-2.53	18	.02	0.58	
wearable_brainstorming_TP9 – wearable_brainstorming_TP10	-0.08	18	.94	0.02	
gel-based_eyes open_Fp1 – gel-based_eyes open_Fp2	0.59	18	.56	0.14	
gel-based_eyes open_Fp1 – gel-based_eyes open_TP9	4.16	18	.001	0.95	*
gel-based_eyes open_Fp1 – gel-based_eyes open_TP10	4.53	18	≤ .001	1.04	*
gel-based_eyes open_Fp2 – gel-based_eyes open_TP9	3.81	18	.001	0.87	*
gel-based_eyes open_Fp2 – gel-based_eyes open_TP10	4.32	18	≤ .001	0.99	*
gel-based_eyes open_TP9 – gel-based_eyes open_TP10	3.85	18	.001	0.88	*
gel-based_eyes closed_Fp1 – gel-based_eyes closed_Fp2	-1.70	18	.11	0.39	
gel-based_eyes closed_Fp1 – gel-based_eyes closed_TP9	-7.09	18	≤ .001	1.63	*
gel-based_eyes closed_Fp1 – gel-based_eyes closed_TP10	-6.24	18	≤ .001	1.43	*
gel-based_eyes closed_Fp2 – gel-based_eyes closed_TP9	-4.96	18	≤ .001	1.14	*

t-test pairs	t	df	p	Cohen's d	Sign. after Bonferroni adjustment
gel-based_eyes closed_Fp2 – gel-based_eyes closed_TP10	-4.61	18	≤ .001	1.06	*
gel-based_eyes closed_TP9 – gel-based_eyes closed_TP10	-0.12	18	.90	0.03	
gel-based_breathing_Fp1 – gel-based_breathing_Fp2	-1.94	18	.07	0.44	
gel-based_breathing_Fp1 – gel-based_breathing_TP9	-7.28	18	≤ .001	1.67	*
gel-based_breathing_Fp1 – gel-based_breathing_TP10	-6.12	18	≤ .001	1.40	*
gel-based_breathing_Fp2 – gel-based_breathing_TP9	-4.85	18	≤ .001	1.11	*
gel-based_breathing_Fp2 – gel-based_breathing_TP10	-4.35	18	≤ .001	1.00	*
gel-based_breathing_TP9 – gel-based_breathing_TP10	0.06	18	.95	0.01	
gel-based_brainstorming_Fp1 – gel-based_brainstorming_Fp2	-1.72	18	.10	0.39	
gel-based_brainstorming_Fp1 – gel-based_brainstorming_TP9	-4.45	18	≤ .001	1.02	*
gel-based_brainstorming_Fp1 – gel-based_brainstorming_TP10	-3.94	18	.001	0.90	*
gel-based_brainstorming_Fp2 – gel-based_brainstorming_TP9	-2.52	18	.02	0.58	
gel-based_brainstorming_Fp2 – gel-based_brainstorming_TP10	-2.37	18	.03	0.54	
gel-based_brainstorming_TP9 – gel-based_brainstorming_TP10	-0.33	18	.74	0.08	

Appendix 9: t-test results for the interaction frequency band * task * device.

t-test pairs	t	df	p	Cohen's d	Sign. after Bonferroni adjustment
wearable_eyes open_alpha – gel-based_eyes open_alpha	-1.60	18	.13	0.37	
wearable_eyes closed_alpha – gel-based_eyes closed_alpha	-2.87	18	.01	0.66	
wearable_breathing_alpha – gel-based_breathing_alpha	-3.47	18	.003	0.80	*
wearable_brainstorming_alpha – gel-based_brainstorming_alpha	-4.10	18	.001	0.94	*
wearable_eyes open_alpha – wearable_eyes closed_alpha	-3.09	18	.006	0.71	*
wearable_eyes open_alpha – wearable_breathing_alpha	-2.85	18	.01	0.65	

t-test pairs	*t*	df	*p*	Cohen's *d*	Sign. after Bonferroni adjustment
wearable_eyes open_alpha – wearable_brainstorming_alpha	-2.62	18	.02	0.60	
wearable_eyes closed_alpha – wearable_breathing_alpha	1.11	18	.28	0.25	
wearable_eyes closed_alpha – wearable_brainstorming_alpha	0.89	18	.39	0.20	
wearable_breathing_alpha – wearable_brainstorming_alpha	0.44	18	.67	0.10	
gel-based_eyes open_alpha – gel-based_eyes closed_alpha	-3.70	18	.002	0.85	*
gel-based_eyes open_alpha – gel-based_breathing_alpha	-3.56	18	.002	0.82	*
gel-based_eyes open_alpha – gel-based_brainstorming_alpha	-3.20	18	.005	0.73	*
gel-based_eyes closed_alpha – gel-based_breathing_alpha	0.27	18	.79	0.06	
gel-based_eyes closed_alpha – gel-based_brainstorming_alpha	1.37	18	.19	0.32	
gel-based_breathing_alpha – gel-based_brainstorming_alpha	1.01	18	.33	0.23	
wearable_eyes open_beta – gel-based_eyes open_beta	-2.03	18	.06	0.47	
wearable_eyes closed_beta – gel-based_eyes closed_beta	0.99	18	.33	0.23	
wearable_breathing_beta – gel-based_breathing_beta	0.62	18	.55	0.14	
wearable_brainstorming_beta – gel-based_brainstorming_beta	2.23	18	.04	0.51	
wearable_eyes open_beta – wearable_eyes closed_beta	-0.11	18	.91	0.03	
wearable_eyes open_beta – wearable_breathing_beta	0.56	18	.58	0.13	
wearable_eyes open_beta – wearable_brainstorming_beta	-0.86	18	.40	0.20	
wearable_eyes closed_beta – wearable_breathing_beta	0.74	18	.47	0.17	
wearable_eyes closed_beta – wearable_brainstorming_beta	-0.84	18	.41	0.19	
wearable_breathing_beta – wearable_brainstorming_beta	-4.41	18	≤ .001	1.01	*
gel-based_eyes open_beta – gel-based_eyes closed_beta	3.17	18	.005	0.73	*
gel-based_eyes open_beta – gel-based_breathing_beta	2.89	18	.001	0.66	*
gel-based_eyes open_beta – gel-based_brainstorming_beta	2.75	18	.01	0.63	

t-test pairs	*t*	df	*p*	Cohen's *d*	Sign. after Bonferroni adjustment
gel-based_eyes open_beta – gel-based_brainstorming_beta	-0.05	18	.96	0.01	
gel-based_eyes closed_beta – gel-based_breathing_beta	-1.30	18	.21	0.30	
gel-based_eyes closed_beta – gel-based_brainstorming_beta	-1.49	18	.16	0.34	
gel-based_breathing_beta – gel-based_brainstorming_beta	-1.60	18	.13	0.37	

Appendix 10: Main effects and interactions revealed by the ANOVAs for the four channels.

ANOVA	Factors	df1	df2	*F*	*p*	Partial η^2
Fp1	Frequency Band	1.00	18.00	57.95	≤ .001	.76
	Session	1.00	18.00	3.16	.09	.15
	Task	1.65	29.78	1.50	.24	.08
	Device	1.00	18.00	2.29	.15	.11
	Frequency Band * Session	1.00	18.00	0.70	.41	.04
	Frequency Band * Task	1.54	27.71	5.78	.01	.24
	Session * Task	1.93	34.75	0.44	.64	.02
	Frequency Band * Session * Task	1.79	32.17	0.84	.43	.04
	Frequency Band * Device	1.00	18.00	1.67	.21	.09
	Session * Device	1.00	18.00	0.35	.56	.02
	Frequency Band * Session * Device	1.00	18.00	0.90	.36	.05
	Task * Device	2.02	36.44	38.87	≤ .001	.68
	Frequency Band * Task * Device	1.64	29.55	24.42	≤ .001	.58
	Session * Task * Device	2.19	39.38	0.07	.95	.00
	Frequency Band * Session * Task * Device	2.13	38.39	1.13	.34	.06
Fp2	Frequency Band	1.00	18.00	60.39	≤ .001	.77
	Session	1.00	18.00	1.31	.27	.07
	Task	1.86	33.48	0.48	.61	.03
	Device	1.00	18.00	1.13	.30	.06
	Frequency Band * Session	1.00	18.00	0.17	.68	.01
	Frequency Band * Task	1.57	28.33	2.21	.14	.11
	Session * Task	2.65	47.78	0.34	.77	.02
	Frequency Band * Session * Task	2.23	40.13	0.34	.74	.02
	Frequency Band * Device	1.00	18.00	1.39	.25	.07
	Session * Device	1.00	18.00	0.08	.78	.00
	Frequency Band * Session * Device	1.00	18.00	0.01	.93	.00
	Task * Device	2.40	43.25	39.38	≤ .001	.69
	Frequency Band * Task * Device	1.63	29.40	24.75	≤ .001	.58
	Session * Task * Device	2.66	47.91	1.36	.27	.07
	Frequency Band * Session * Task * Device	2.31	41.65	0.66	.54	.04
A2	Frequency Band	1.00	18.00	6.80	.02	.27

ANOVA	Factors	df1	df2	*F*	*p*	Partial η^2
	Session	1.00	18.00	0.02	.90	.00
	Task	1.54	27.78	32.15	≤ .001	.64
	Device	1.00	18.00	32.36	≤ .001	.64
	Frequency Band * Session	1.00	18.00	0.29	.60	.02
	Frequency Band * Task	1.49	26.77	11.69	.001	.39
	Session * Task	2.44	43.91	1.03	.38	.05
	Frequency Band * Session * Task	2.41	43.39	0.31	.78	.02
	Frequency Band * Device	1.00	18.00	9.06	.008	.33
	Session * Device	1.00	18.00	0.35	.56	.02
	Frequency Band * Session * Device	1.00	18.00	1.59	.22	.08
	Task * Device	2.50	45.07	1.50	.23	.08
	Frequency Band * Task * Device	2.55	45.93	5.68	.003	.24
	Session * Task * Device	1.96	35.36	0.52	.60	.03
	Frequency Band * Session * Task * Device	2.07	37.34	0.37	.70	.02
	Frequency Band	1.00	18.00	3.35	.08	.16
	Session	1.00	18.00	0.00	.96	.00
	Task	1.54	27.67	34.17	≤ .001	.65
	Device	1.00	18.00	70.68	≤ .001	.80
	Frequency Band * Session	1.00	18.00	0.00	.98	.00
	Frequency Band * Task	1.54	27.63	9.76	.001	.35
	Session * Task	2.70	48.52	1.39	.26	.07
A1	Frequency Band * Session * Task	2.70	48.52	0.65	.57	.03
	Frequency Band * Device	1.00	18.00	4.18	.05	.19
	Session * Device	1.00	18.00	0.76	.39	.04
	Frequency Band * Session * Device	1.00	18.00	0.16	.70	.01
	Task * Device	2.59	46.55	4.39	.01	.20
	Frequency Band * Task * Device	2.86	51.42	3.01	.04	.14
	Session * Task * Device	2.38	42,78	0.58	.60	.03
	Frequency Band * Session * Task * Device	2.11	38,04	0.31	.75	.02

Appendix 11: t-test results regarding the ANOVA for Fp1 for the interaction device* condition * frequency band.

t-test pairs	*t*	df	*p*	Cohen's *d*	Sign. after Bonferroni adjustment
wearable_eyes open_alpha – wearable_eyes closed_alpha	-4.24	18	≤ .001	0.97	*
wearable_eyes open_alpha – wearable_breathing_alpha	-3.73	18	.002	0.85	*
wearable_eyes open_alpha – wearable_brainstorming_alpha	-3.05	18	.007	0.70	*
wearable_eyes closed_alpha – wearable_breathing_alpha	0.79	18	.44	0.18	
wearable_eyes closed_alpha – wearable_brainstorming_alpha	2.97	18	.008	0.68	*

t-test pairs	*t*	df	*p*	Cohen's *d*	Sign. after Bonferroni adjustment
wearable_breathing_alpha – wearable_brainstorming_alpha	2.98	18	.008	0.68	*
gel-based_eyes open_alpha – gel-based_eyes closed_alpha	-1.33	18	.20	0.30	
gel-based_eyes open_alpha – gel-based_breathing_alpha	0.05	18	.96	0.01	
gel-based_eyes open_alpha – gel-based_brainstorming_alpha	0.82	18	.43	0.19	
gel-based_eyes closed_alpha – gel-based_breathing_alpha	1.69	18	.11	0.39	
gel-based_eyes closed_alpha – gel-based_brainstorming_alpha	3.14	18	.006	0.72	*
gel-based_breathing_alpha – gel-based_brainstorming_alpha	1.72	18	.10	0.39	
wearable_eyes open_beta – wearable_eyes closed_beta	-2.76	18	.01	0.63	
wearable_eyes open_beta – wearable_breathing_beta	-2.60	18	.02	0.60	
wearable_eyes open_beta – wearable_brainstorming_beta	-2.47	18	.02	0.57	
wearable_eyes closed_beta – wearable_breathing_beta	0.17	18	.87	0.04	
wearable_eyes closed_beta – wearable_brainstorming_beta	0.95	18	.36	0.22	
wearable_breathing_beta – wearable_brainstorming_beta	0.73	18	.47	0.17	
gel-based_eyes open_beta – gel-based_eyes closed_beta	4.03	18	.001	0.92	*
gel-based_eyes open_beta – gel-based_breathing_beta	4.55	18	≤ .001	1.04	*
gel-based_eyes open_beta – gel-based_brainstorming_beta	4.16	18	.001	0.95	*
gel-based_eyes closed_beta – gel-based_breathing_beta	1.17	18	.26	0.27	
gel-based_eyes closed_beta – gel-based_brainstorming_beta	0.98	18	.34	0.23	
gel-based_breathing_beta – gel-based_brainstorming_beta	-0.55	18	.59	0.13	
wearable_eyes open_alpha – gel-based_eyes open_alpha	-2.36	18	.03	0.54	
wearable_eyes closed_alpha – gel-based_eyes closed_alpha	1.41	18	.18	0.32	
wearable_breathing_alpha – gel-based_breathing_alpha	2.12	18	.05	0.49	
wearable_brainstorming_alpha – gel-based_brainstorming_alpha	1.57	18	.13	0.36	

t-test pairs	*t*	df	*p*	Cohen's *d*	Sign. after Bonferroni adjustment
wearable_eyes open_beta – gel-based_eyes open_beta	-4.47	18	≤ .001	1.02	*
wearable_eyes closed_beta – gel-based_eyes closed_beta	4.50	18	≤ .001	1.03	*
wearable_breathing_beta – gel-based_breathing_beta	5.86	18	≤ .001	1.34	*
wearable_brainstorming_beta – gel-based_brainstorming_beta	3.51	18	.002	0.81	*

Appendix 12: t-test results regarding the ANOVA for Fp2 for the interaction device* condition * frequency band.

t-test pairs	*t*	df	*p*	Cohen's *d*	Sign. after Bonferroni adjustment
wearable_eyes open_alpha – wearable_eyes closed_alpha	-5.62	18	≤ .001	1.29	*
wearable_eyes open_alpha – wearable_breathing_alpha	-4.98	18	≤ .001	1.14	*
wearable_eyes open_alpha – wearable_brainstorming_alpha	-3.33	18	.004	0.76	*
wearable_eyes closed_alpha – wearable_breathing_alpha	0.79	18	.44	0.18	
wearable_eyes closed_alpha – wearable_brainstorming_alpha	2.14	18	.05	0.49	
wearable_breathing_alpha – wearable_brainstorming_alpha	1.86	18	.08	0.43	
gel-based_eyes open_alpha – gel-based_eyes closed_alpha	-2.24	18	.04	0.51	
gel-based_eyes open_alpha – gel-based_breathing_alpha	-0.93	18	.37	0.21	
gel-based_eyes open_alpha – gel-based_brainstorming_alpha	-0.41	18	.69	0.09	
gel-based_eyes closed_alpha – gel-based_breathing_alpha	1.95	18	.07	0.45	
gel-based_eyes closed_alpha – gel-based_brainstorming_alpha	2.25	18	.04	0.52	
gel-based_breathing_alpha – gel-based_brainstorming_alpha	0.93	18	.37	0.21	
wearable_eyes open_beta – wearable_eyes closed_beta	-4.30	18	≤ .001	0.99	*
wearable_eyes open_beta – wearable_breathing_beta	-3.60	18	.002	0.83	*
wearable_eyes open_beta – wearable_brainstorming_beta	-2.25	18	.04	0.52	
wearable_eyes closed_beta – wearable_breathing_beta	0.29	18	.78	0.07	

t-test pairs	*t*	df	*p*	Cohen's *d*	Sign. after Bonferroni adjustment
wearable_eyes closed_beta – wearable_brainstorming_beta	0.79	18	.44	0.18	
wearable_breathing_beta – wearable_brainstorming_beta	0.40	18	.70	0.09	
gel-based_eyes open_beta – gel-based_eyes closed_beta	3.50	18	.003	0.80	*
gel-based_eyes open_beta – gel-based_breathing_beta	3.72	18	.002	0.85	*
gel-based_eyes open_beta – gel-based_brainstorming_beta	3.60	18	.002	0.83	*
gel-based_eyes closed_beta – gel-based_breathing_beta	0.77	18	.45	0.18	
gel-based_eyes closed_beta – gel-based_brainstorming_beta	0.75	18	.46	0.17	
gel-based_breathing_beta – gel-based_brainstorming_beta	-0.04	18	.97	0.01	
wearable_eyes open_alpha – gel-based_eyes open_alpha	-2.77	18	.01	0.64	*
wearable_eyes closed_alpha – gel-based_eyes closed_alpha	0.51	18	.62	0.12	
wearable_breathing_alpha – gel-based_breathing_alpha	1.52	18	.15	0.35	
wearable_brainstorming_alpha – gel-based_brainstorming_alpha	0.45	18	.66	0.10	
wearable_eyes open_beta – gel-based_eyes open_beta	-4.88	18	≤ .001	1.12	*
wearable_eyes closed_beta – gel-based_eyes closed_beta	5.03	18	≤ .001	1.15	*
wearable_breathing_beta – gel-based_breathing_beta	4.81	18	≤ .001	1.10	*
wearable_brainstorming_beta – gel-based_brainstorming_beta	2.93	18	.009	0.67	*

Appendix 13: t-test results regarding the ANOVA for A1 for the interaction device* condition * frequency band.

t-test pairs	*t*	df	*p*	Cohen's *d*	Sign. after Bonferroni adjustment
wearable_eyes open_alpha – wearable_eyes closed_alpha	-6.42	18	≤ .001	1.47	*
wearable_eyes open_alpha – wearable_breathing_alpha	-6.08	18	≤ .001	1.39	*
wearable_eyes open_alpha – wearable_brainstorming_alpha	-4.75	18	≤ .001	1.09	*
wearable_eyes closed_alpha – wearable_breathing_alpha	-1.55	18	.14	0.36	

t-test pairs	***t***	**df**	***p***	**Cohen's *d***	**Sign. after Bonferroni adjustment**
wearable_eyes closed_alpha – wearable_brainstorming_alpha	3.88	18	.001	0.89	*
wearable_breathing_alpha – wearable_brainstorming_alpha	4.96	18	≤ .001	1.14	*
gel-based_eyes open_alpha – gel-based_eyes closed_alpha	-6.02	18	≤ .001	1.38	*
gel-based_eyes open_alpha – gel-based_breathing_alpha	-6.10	18	≤ .001	1.40	*
gel-based_eyes open_alpha – gel-based_brainstorming_alpha	-4.47	18	≤ .001	1.02	*
gel-based_eyes closed_alpha – gel-based_breathing_alpha	1.06	18	.30	0.24	
gel-based_eyes closed_alpha – gel-based_brainstorming_alpha	5.50	18	≤ .001	1.26	*
gel-based_breathing_alpha – gel-based_brainstorming_alpha	4.68	18	≤ .001	1.07	*
wearable_eyes open_beta – wearable_eyes closed_beta	-3.63	18	.002	0.83	*
wearable_eyes open_beta – wearable_breathing_beta	-4.46	18	≤ .001	1.02	*
wearable_eyes open_beta – wearable_brainstorming_beta	-2.71	18	.01	0.62	
wearable_eyes closed_beta – wearable_breathing_beta	-1.59	18	.13	0.37	
wearable_eyes closed_beta – wearable_brainstorming_beta	2.05	18	.06	0.47	
wearable_breathing_beta – wearable_brainstorming_beta	3.87	18	.001	0.89	*
gel-based_eyes open_beta – gel-based_eyes closed_beta	-2.62	18	.02	0.60	
gel-based_eyes open_beta – gel-based_breathing_beta	-2.22	18	.04	0.51	
gel-based_eyes open_beta – gel-based_brainstorming_beta	-0.95	18	.36	0.22	
gel-based_eyes closed_beta – gel-based_breathing_beta	0.78	18	.45	0.18	
gel-based_eyes closed_beta – gel-based_brainstorming_beta	4.40	18	≤ .001	1.01	*
gel-based_breathing_beta – gel-based_brainstorming_beta	3.40	18	.003	0.78	*
wearable_eyes open_alpha – gel-based_eyes open_alpha	4.37	18	≤ .001	1.00	*
wearable_eyes closed_alpha – gel-based_eyes closed_alpha	1.36	18	.19	0.31	
wearable_breathing_alpha – gel-based_breathing_alpha	4.47	18	≤ .001	1.03	*

t-test pairs	t	df	p	Cohen's d	Sign. after Bonferroni adjustment
wearable_brainstorming_alpha – gel-based_brainstorming_alpha	2.52	18	.02	0.58	
wearable_eyes open_beta – gel-based_eyes open_beta	2.13	18	.05	0.49	
wearable_eyes closed_beta – gel-based_eyes closed_beta	4.25	18	≤ .001	0.98	*
wearable_breathing_beta – gel-based_breathing_beta	8.54	18	≤ .001	1.96	*
wearable_brainstorming_beta – gel-based_brainstorming_beta	7.89	18	≤ .001	1.81	*

Appendix 14: t-test results regarding the ANOVA for A2 for the interaction device* condition * frequency band.

t-test pairs	t	df	p	Cohen's d	Sign. after Bonferroni adjustment
wearable_eyes open_alpha – wearable_eyes closed_alpha	-6.75	18	≤ .001	1.55	*
wearable_eyes open_alpha – wearable_breathing_alpha	-5.14	18	≤ .001	1.18	*
wearable_eyes open_alpha – wearable_brainstorming_alpha	-5.01	18	≤ .001	1.15	*
wearable_eyes closed_alpha – wearable_breathing_alpha	1.07	18	.30	0.25	
wearable_eyes closed_alpha – wearable_brainstorming_alpha	5.09	18	≤ .001	1.17	*
wearable_breathing_alpha – wearable_brainstorming_alpha	3.50	18	.003	0.80	*
gel-based_eyes open_alpha – gel-based_eyes closed_alpha	-5.69	18	≤ .001	1.31	*
gel-based_eyes open_alpha – gel-based_breathing_alpha	-5.55	18	≤ .001	1.27	*
gel-based_eyes open_alpha – gel-based_brainstorming_alpha	-4.43	18	≤ .001	1.02	*
gel-based_eyes closed_alpha – gel-based_breathing_alpha	0.94	18	.36	0.22	
gel-based_eyes closed_alpha – gel-based_brainstorming_alpha	4.85	18	≤ .001	1.11	*
gel-based_breathing_alpha – gel-based_brainstorming_alpha	3.93	18	.001	0.90	*
wearable_eyes open_beta – wearable_eyes closed_beta	-5.57	18	≤ .001	1.28	*
wearable_eyes open_beta – wearable_breathing_beta	-3.94	18	.001	0.90	*
wearable_eyes open_beta – wearable_brainstorming_beta	-3.32	18	.004	0.76	*

t-test pairs	*t*	df	*p*	Cohen's *d*	Sign. after Bonferroni adjustment
wearable_eyes closed_beta – wearable_breathing_beta	1.28	18	.22	0.29	
wearable_eyes closed_beta – wearable_brainstorming_beta	5.05	18	≤ .001	1.16	*
wearable_breathing_beta – wearable_brainstorming_beta	2.44	18	.03	0.56	
gel-based_eyes open_beta – gel-based_eyes closed_beta	-2.64	18	.02	0.60	
gel-based_eyes open_beta – gel-based_breathing_beta	-2.10	18	.05	0.48	
gel-based_eyes open_beta – gel-based_brainstorming_beta	-0.72	18	.48	0.17	
gel-based_eyes closed_beta – gel-based_breathing_beta	1.17	18	.26	0.27	
gel-based_eyes closed_beta – gel-based_brainstorming_beta	5.60	18	≤.000	1.28	*
gel-based_breathing_beta – gel-based_brainstorming_beta	3.62	18	.002	0.83	*
wearable_eyes open_alpha – gel-based_eyes open_alpha	4.37	18	≤ .001	1.00	*
wearable_eyes closed_alpha – gel-based_eyes closed_alpha	2.17	18	.04	0.50	
wearable_breathing_alpha – gel-based_breathing_alpha	1.28	18	.22	0.29	
wearable_brainstorming_alpha – gel-based_brainstorming_alpha	1.17	18	.26	0.27	
wearable_eyes open_beta – gel-based_eyes open_beta	2.81	18	.01	0.64	*
wearable_eyes closed_beta – gel-based_eyes closed_beta	7.69	18	≤ .001	1.76	*
wearable_breathing_beta – gel-based_breathing_beta	4.81	18	≤ .001	1.10	*
wearable_brainstorming_beta – gel-based_brainstorming_beta	6.09	18	≤ .001	1.40	*

Appendix 15: Statements and their respective rating scales that were filled out after each exercise for assessment.

Questions for All Exercises					
Question	**Rating**				
	1 (strongly disagree)	2 (somewhat disagree)	3 (neutral)	4 (somewhat agree)	5 (strongly agree)
The exercise instructions were easy to follow	☐	☐	☐	☐	☐
It was easy for me to perform the exercise	☐	☐	☐	☐	☐
I enjoyed the exercise	☐	☐	☐	☐	☐
The audio recording is of good general quality	☐	☐	☐	☐	☐
The music is pleasant	☐	☐	☐	☐	☐
Additional Questions for Exercises with Audio Instructions (PMR, Mindfulness Meditation, Breathing)					
Question	**Rating**				
	1 (strongly disagree)	2 (somewhat disagree)	3 (neutral)	4 (somewhat agree)	5 (strongly agree)
The speakers voice is pleasant	☐	☐	☐	☐	☐
The speakers voice is understandable at all times	☐	☐	☐	☐	☐
The speed of speech is too fast	☐	☐	☐	☐	☐
The speed of speech is too slow	☐	☐	☐	☐	☐

Appendix 16: Statements and their respective rating scales that were filled out at the end of the study for assessment.

Questions for All Exercises					
Question	**Rating**				
	1 (strongly disagree)	2 (somewhat disagree)	3 (neutral)	4 (somewhat agree)	5 (strongly agree)
I enjoyed the exercise “meditation”	☐	☐	☐	☐	☐
I enjoyed the exercise “guided imagery”	☐	☐	☐	☐	☐
I enjoyed the exercise “breathing”	☐	☐	☐	☐	☐
I enjoyed the exercise “progressive muscle relaxation”	☐	☐	☐	☐	☐

Questions for All Exercises					
Question	**Rating**				
	1 (strongly disagree)	2 (somewhat disagree)	3 (neutral)	4 (somewhat agree)	5 (strongly agree)
I enjoyed the exercise "music"	☐	☐	☐	☐	☐
It was easy to perform the exercise "meditation"	☐	☐	☐	☐	☐
It was easy to perform the exercise "guided imagery"	☐	☐	☐	☐	☐
It was easy to perform the exercise "breathing"	☐	☐	☐	☐	☐
It was easy to perform the exercise "progressive muscle relaxation"	☐	☐	☐	☐	☐
It was easy to perform the exercise "music"	☐	☐	☐	☐	☐

Appendix 17: List of gamification techniques according to Hoffmann et al. (2017) and how they are realized within Stress-Mentor.

Gamification Technique **(Hoffmann et al 2017)**	**Included**	**Realization within Stress-Mentor**
Economic		
Marketplace and economies	✓	Shop, currency
Digital rewards	✓	Badges, points, positive appearance of avatar in accordance to behaviors that meet the health recommendations
Real world prizes	-	-
Social		
Avatar	✓	Bird-like cartoon animal (Elwetritsch) that reflects the user's behavior
Agent	✓	Wise owl that guides the user through the app and provides information
Competition	-	-
Teams	-	-
Parallel communication systems	-	-
Social pressure	-	-
Performance-oriented		
Feedback	✓	Badges, color coding of diary entries, diary overview diagram, progress bars
Levels	✓	Avatar development, increasingly difficult tasks with progression in app
Secondary game objectives	✓	Make avatar look as healthy as possible, learn as many different stress management techniques as possible, try to achieve all badges
Ranks of achievement	-	-

Gamification Technique (Hoffmann et al 2017)	Included	Realization within Stress-Mentor
Leaderboards	-	-
Time pressure	-	-
Embedding-focused		
Narrative context	✓	Responsibility for rearing the avatar, avatar grows with user's progress
3-D environments	-	-

Appendix 18: List of behavior change techniques according to Abraham and Michie (2008) and how they are realized within Stress-Mentor.

Behavior Change Technique (Abraham and Michie 2008)	Included	Realization within Stress-Mentor
1. Provide information about behavior health link	✓	Health information for every diary category, e.g. how much caffeine should be consummated at the most; agent provides tips on stress and stress management
2. Provide information on consequences	✓	Information about stress is linked to nutrition, regular exercise, emotions and the appraisal of events
3. Provide information about others' approval	-	-
4. Prompt intention formation	✓	Every day the user can choose 1 out of 3 stress management tasks which he or she wants to accomplish today
5. Prompt barrier identification	-	-
6. Provide general encouragement	✓	Praise at the end of the tasks; agent provides motivational quotes
7. Set graded tasks	✓	Task difficulty increases within the different categories, e.g. relaxation exercises start with detailed audio instructions, followed by shortened audio versions and expert versions with text instructions. By completing the easy tasks more difficult tasks will be available in future sessions
8. Provide instruction	✓	Multiple types of instruction are used: text, audio files, multiple choice quizzes, and photos
9. Model or demonstrate the behavior	✓	Avatar reflects user's behavior through its appearance
10. Prompt specific goal setting	✓	Task for setting identifying priorities and goals; task of the day and task of the week feature
11. Prompt review of behavioral goals	-	-
12. Prompt self-monitoring of behavior	✓	Every day the user fills out a diary to monitor stress related behaviors
13. Provide feedback on performance	✓	The user's knowledge about stress management, irrational ideas and health is tested in quiz tasks with direct appraisal; traffic-light coloring of the diary entries provides feedback on whether health recommendations are met, development of the avatar reflects user's progress

Behavior Change Technique (Abraham and Michie 2008)	**Included**	**Realization within Stress-Mentor**
14. Provide contingent rewards	✓	Badges reward for behavior that meets health recommendations and contingent usage of the app; user receives coins for diary entries, tasks and stress checklist
15. Teach to use prompts or cues	✓	Embedded in the tasks for time management and relaxation, e.g. short breathing exercises can be performed each time before answering the phone
16. Agree on behavioral contract	✓	User must agree to a behavioral contract with the agent in order to take over responsibility of rearing the avatar
17. Prompt practice	✓	Reminders to complete the tasks of the day and tasks of the week
18. Use follow-up prompts	✓	Four weeks after finishing the intervention the avatar sends a postcard reminding the user of the learned exercises
19. Provide opportunities for social comparison	✓	The photobook provides the opportunity to share pictures of the avatar with others
20. Plan social support or social change	✓	Separate task category, e.g. the user is asked to integrate help from friends and family in his daily life
21. Prompt identification as a role model	-	-
22. Prompt self-talk	✓	Used within some relaxation exercise
23. Relapse prevention	-	-
24. Stress management	✓	Task categories comprise relaxation methods, time management, revealing and refuting irrational ideas, assertiveness training, planning social support, general knowledge about stress management, euthymic methods, and physical tasks for muscle relaxation and stress relief
25. Motivational interviewing	-	-
26. Time management	✓	Separate task category which teaches effective planning and to set priorities

Appendix 19: Means (*M*), standard deviations (*SD*), and statistical values of Mann-Whitney-U tests for uMARS ratings of the control group (CG) and the experimental group (EG) of Stress-Mentor by item. The new alpha level is set at .01 according to the Bonferroni correction method.

uMARS Sub-Category	***M* CG**	***M* EG**	***SD* CG**	***SD* EG**	***U***	***p***	***r***
Engagement	3.90	4.20	0.61	0.57	291.50	.07	0.03
Functionality	4.54	4.74	0.47	0.29	293.50	.06	0.03
Aesthetics	4.35	4.62	0.46	0.41	251.50	.01	0.04
Information	4.37	4.42	0.39	0.48	367.50	.53	0.01

Appendix 20: Means (*M*), standard deviations (*SD*), and statistical values of Mann-Whitney-U tests for uMARS ratings of the control group (CG) and the experimental group (EG) of Stress-Mentor by item. The new alpha level is set at .025 according to the Bonferroni correction method.

Category	Item	*M* CG	*M* EG	*SD* CG	*SD* EG	*U*	*p*	*r*
Engagement	Entertainment	3.93	4.10	0.94	0.90	364.50	.49	0.09
	Interest	4.18	4.38	0.86	0.68	361.50	.45	0.10
	Customization	3.68	4.00	1.19	1.04	347.00	.32	0.13
	Interactivity	3.18	3.79	1.06	1.01	264.00	.02	0.31
	Target Group	4.57	4.48	0.63	0.79	396.00	.91	0.03
Functionality	Performance	4.75	4.79	0.52	0.56	379.00	.65	0.09
	Ease of Use	4.46	4.62	0.64	0.62	347.50	.33	0.14
	Navigation	4.36	4.48	0.56	0.57	357.00	.43	0.12
	Gestural Design	4.61	4.79	0.74	0.49	358.50	.29	0.14
Aesthetics	Layout	4.61	4.66	0.57	0.72	366.00	.47	0.11
	Graphics	4.5	4.72	0.64	0.59	322.00	.13	0.22
	Visual Appeal	3.93	4.34	0.66	0.67	275.00	.03	0.30
Information	Quality of Information	4.32	4.45	0.61	0.63	358.50	.45	0.11
	Quantity of Information	4.54	4.48	0.58	0.57	385.00	.75	0.05
	Visual Information	4.07	4.38	1.09	0.98	324.50	.15	0.19
	Credibility of Source	4.50	4.38	0.58	1.02	405.50	.99	0.00
Subjective Quality	Recommendation to Others	4.00	4.38	0.94	0.94	308.50	.09	0.22
	Amount of Uses	4.36	4.55	1.16	0.99	361.50	.38	0.12
	Willingness to Pay	2.54	3.00	1.40	1.56	335.00	.25	0.15
	Overall Star Rating	3.89	4.21	0.63	0.90	293.00	.06	0.26
Perceived Impact	Awareness	3.64	3.83	1.25	1.20	367.50	.54	0.09
	Knowledge	3.36	3.90	1.10	1.01	285.50	.04	0.27
	Attitudes	3.04	3.45	1.35	1.45	334.50	.25	0.15
	Intention to Change	3.32	3.93	1.19	1.31	269.00	.02	0.30
	Helps Seeking	3.36	3.28	1.28	1.36	393.00	.84	0.03
	Behavior Change	4.11	3.93	0.88	1.03	373.50	.59	0.07

Appendix 21: List of additional questions for the semi-structured expert interview.

	Question	Answer options
1.	Do you think the app would appeal to patients?	N/A I cannot assess this 1 definitely not 2 rather not 3 probably 4 very likely 5 of course
2.	Did you like the gamification concept?	N/A I cannot assess this 1 not at all 2 a little 3 so/so 4 mostly 5 completely
3.	Did the app meet your expectations?	N/A I cannot assess this 1 not at all 2 a little 3 so/so 4 mostly 5 completely
4.	How useful is the diary?	N/A I cannot assess this 1 not at all 2 not useful 3 so/so 4 mostly 5 very
5.	How useful is the symptoms checklist?	N/A I cannot assess this 1 not at all 2 not useful 3 so/so 4 mostly 5 very
6.	How useful is the concept of the daily exercises?	N/A I cannot assess this 1 not at all 2 not useful 3 so/so 4 mostly 5 very
7.	How useful is it to apply the app in addition to therapy?	N/A I cannot assess this 1 not at all 2 not useful 3 so/so 4 mostly 5 very
8.	What do you think could keep patients from using the app?	

Appendix 22: Means (*M*) and standard deviations (*SD*) of the experts' ratings for each category of the MARS questionnaire.

MARS Category	*M*	*SD*
General app quality	4.51	0.54
Subjective Quality	4.51	0.31
App specific	4.27	0.76
Engagement	4.32	0.62
Functionality	4.73	0.50
Aesthetics	4.45	0.56
Information	4.53	0.49

Appendix 23: Listed are the questions at least one expert felt unable to assess and the number of experts that were unable to assess each of these questions.

Question	N
Customization	1
Navigation	1
Goals	1
Information quantity	1
Attitudes	1
Intention to change	1
Awareness	2
Interactivity	3
Gestural design	5
Behavior change	6

Acknowledgements

Writing a PhD thesis is a long, hard, and for the most part, also a very enjoyable endeavor. It continuously presented me with new challenges and opportunities that helped me grow as a scientist, as well as a person. On my road I was supported by many whom I would like to thank for their support at this point. First and foremost, I thank my supervisor Dr. Corinna A. Faust-Christmann for her guidance and support. She provided advice and critical comments, shared her knowledge, and maybe most importantly, gave me the freedom to learn while retaining faith in my competence throughout. Moreover, I would like to thank Dr. Gabriele Bleser and Prof. Dr. Thomas Lachmann for their guidance and unconditional support throughout this process. Without them this thesis would not have been possible. Thanks are also due to my colleagues for continuously backing me up. This dissertation would not have been possible without the contribution of Dr. Gregor Zolynski, an excellent android developer. Thanks to Petra Ludowicy for sharing in my ups and downs, for feeling with me and providing emotional support, as well as advice when I needed it. Furthermore, I want to thank my other colleagues that include Dr. Bertram Taetz, Markus Miezal, Michael Lorenz, Wolfgang Teufl, Felix Laufer, and Alexander Weidmann for making my experience in the wearHEALTH research group exciting and fun. Thanks are also due to the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF, reference number: 16SV7115) for providing the funding for this project. Moreover, thanks are due to all study participants. Without them, I would not have been able to conduct my studies and make my scientific discoveries. Last but not least, I would like to express my deepest gratitude to my family and friends. Especially to my fiancé, Peter Eisenhauer, who was my rock when I needed someone to lean on and who kept me on track when I would have lost course. Thanks to my sister and mother for always having an open ear and being there whenever I need them. This dissertation would not have been possible without their warm love, continued patience, and endless support.

Notes

Parts of Chapter 2 were previously published in a modified version in peer-review articles. Chapter 2.1 was modified with the permission from Springer Nature Customer Service Centre GmbH: Springer Nature, Proceedings of the 20th Congress of the International Ergonomics Association, "Towards more interactive stress-related self-monitoring tools to improve quality of life" by Corinna A. Christmann, Gregor Zolynski, Alexandra Hoffmann, and Gabriele Bleser, 2018. An altered version of Chapter 2.2 was previously published at the conference Human Computer Interaction International. The publication was modified by permission from Springer Nature Customer Service Centre GmbH: Springer, Digital Human Modeling. Applications in Health, Safety, Ergonomics, and Risk Management: Health and Safety, "Effective Visualization of Long Term Health Data to Support Behavior Change" by Corinna A. Christmann, Gregor Zolynski, Alexandra Hoffmann, and Gabriele Bleser, 2017. Also, Chapter 2.4 was modified by permission from Springer Nature Customer Service Centre GmbH: Springer, HCI International 2018 – Posters' Extended Abstracts, "Stress-Mentor: Linking Gamification and Behavior Change Theory in a Stress Management Application" by Corinna A. Christmann, Alexandra Hoffmann, Gregor Zolynski, and Gabriele Bleser, 2018.

Chapter 3 was slightly adapted by permission from Springer Nature Customer Service Centre GmbH: Springer, Design, User Experience, and Usability. Application Domains Association, "Gamification of a Stress Management App: Results of a User Study" by Alexandra Hoffmann, Corinna A. Faust-Christmann, Gregor Zolynski, and Gabriele Bleser, 2019.

Chapter 4 is based on an altered form of the article "Does gamification affect the usage behavior of stress management apps: A longitudinal user study with "Stress-Mentor" by Alexandra Hoffmann, Corinna A. Faust-Christmann, Gregor Zolynski, and Gabriele Bleser, which is currently under review at the JMIR Serious Games.

An altered version of the article "Towards gamified pain management apps: A MARS-based quality assessment of "Pain-Mentor's" first prototype through an expert study" by Alexandra Hoffmann, Corinna A. Faust-Christmann, Gregor Zolynski, and Gabriele Bleser, which is published at JMIR Formative Research, is found in Chapter 5.

Declaration

I hereby confirm that this thesis titled “The iterative development and evaluation of the gamified stress management app “Stress-Mentor”” is the result of my own work. All sources and materials that were used are listed and specified in the thesis. The information derived from the literature has been duly acknowledged in the text and a list of references provided. I confirm that this thesis has not yet been submitted as part of another examination process.

Kaiserslautern, March 11, 2020

Alexandra Hoffmann

Curriculum Vitae

Persönliche Angaben

Name:	Alexandra Hoffmann
Anschrift	Gottlieb-Daimler Straße 48, 67663 Kaiserslautern
E-Mail:	hoffmann@cs.uni-kl.de

Ausbildung

12/2015 – heute	Wissenschaftliche Mitarbeiterin an der Technischen Universität Kaiserslautern, Fachbereich Informatik, AG wearHEALTH, Dr. Gabriele Bleser
10/2013 – 12/2015	Master-Studium Kognitionswissenschaften an der Technischen Universität Kaiserslautern Thema der Masterarbeit: „Stress App Analysis in Regards to Gamification, Behavior Change Techniques and Stress Management Techniques“
10/2009 – 09/2013	Bachelor-Studium Biowissenschaften an der Technischen Universität Kaiserslautern Thema der Bachelorarbeit: „Vergleichbare Fluoreszenz- in *situ*- Hybridisierung an Tropf- und Tupfpräperaten oligodendroglialer Tumore “

Kaiserslautern, 11.03.2020