Development of a Multidimensional App Quality Assessment Tool for Health-Related Apps (AQUA)

Teresa O'Rourke[1*], Rüdiger C. Pryss[2], Winfried Schlee[3] & Thomas Probst[1]

[1] Department of Psychotherapy and Biopsychosocial Health, Danube University Krems, Austria

[2] Institute for Databases and Information Systems, Faculty of Engineering, Computer Sciences and Psychology, University of Ulm, Germany

[3] Department of Psychiatry and Psychotherapy, Faculty of Medicine, University of Regensburg Germany

Abstract

Background: A multitude of health-related mobile applications are available to the public in app stores. Many of these apps were not developed by health professionals and are not scientifically valid. To facilitate a safe handling and use of such apps, it is important to assess their quality in a standardized way. Some instruments for app quality assessment already exist, although they have some limitations, which we want to improve upon with a new multi-dimensional assessment tool.

Objectives: The objective of this paper is to explain the development of a new multidimensional criteria-based tool for the quality assessment of health-related apps (AQUA).

Method: Based on existing app-quality assessment tools and guidelines for evaluating health-related app-quality, questionnaire items were constructed to assess the quality of mHealth apps from the perspective of both experts and users. Before the finalization of the questionnaire that would form the basis of AQUA, we conducted a pretest of the original German items with six participants, who gave qualitative feedback on the items while filling them out as they completed the surveys.

Results: An expert and a user version of AQUA were developed in English and German. The expert version consists of 31 items in seven dimensions: Usability; User Engagement; Content; Visual Design; Therapeutic Quality; Security; and Information. The user version consists of 31 items in the following dimensions: Usability; User Engagement; Content; Visual Design; Therapeutic Quality; Impact; and Information.

Conclusion: AQUA is a brief multidimensional app-quality assessment tool that can be used by experts and app-users to quickly determine the quality of health-related and mental health-related apps.

Keywords: mHealth, health-related apps, digital health, quality assessment, evaluation

1 Introduction

Mobile app-based mental health interventions are becoming increasingly popular in the field of mobile health (mHealth) as they have the potential to reach more people in need of mental health care than traditional mental health services (Kazdin, 2017). They are low-threshold interventions that can support people in the self-management of their health if adequately designed (Baldwin, Singh, Sittig, & Giardina, 2017). Different mHealth apps exist for a multitude of conditions, including diabetes (Barreda-Pérez, De la Torre, & López-Coronado, 2013), depression (Shen et al., 2015), and smoking-cessation (Abroms, Padmanabhan, Thaweethai, & Phillips, 2011). The impact of such interventions depends on different factors, such as users' motivation and ability to engage with them (Kohl, Crutzen, & de Vries, 2013). To

improve user engagement in health-related mobile apps, the concept of therapeutic persuasiveness, meaning a persuasive design and an app's therapeutic potential (Webb, Joseph, Yardley, & Michie, 2010), or a human factor (e.g. a chatbot or avatar), that supports the therapeutic alliance between user and intervention in the app (Ly, Ly, & Anderson, 2017), can be incorporated. A myriad of health-related and mental health-related apps are currently available in different app stores such as the Google Play Store[1] and the Apple Store[2]. However, many of these apps are often hastily developed without an empirical basis (BinDhim, Hawkey & Trevena, 2015). Consequently, quality assessment for these apps is becoming increasingly crucial to ensure their

---

[1]   https://play.google.com/store

[2]   https://www.apple.com/at/ios/app-store/

safe and effective use. BinDhim, Hawkey, & Trevena (2015), for example, suggested the use of evidence-based guidelines to systematically assess app content and to avoid consumer ratings to judge the quality of health-related apps. A recent meta-analysis by Nouri et al. (2018) of the existing literature on mHealth app quality reviewed 23 studies concerning mHealth app quality assessment tools. Based on this review they suggested seven main categories of app quality assessment criteria: Design; Information/Content; Usability; Functionality; Ethical Issues; Security and Privacy; and User-perceived value. Based on an earlier systematic review of headache apps, Hundert et al. (2014) established a set of seven quality criteria an ideal headache app should contain. Among other criteria, it should "be created with clinical and/or scientific headache expertise, have undergone testing…, measure clinically relevant headache variables" and "be usable". To measure the usability of an app, they suggested a list of ten usability heuristics, which includes criteria such as aesthetic design, and user anonymity and/or transparency regarding saved data (Hundert et al., 2014, p. 3). Mcmillan et al. (2016) identified the following nine dimensions relevant for app quality assessment based on the National Institute for Health and Care Excellence behavior change guidance (2014): Purpose; Planning and development; Usability; Initial assessment and tailoring; Behavior change technique; Maintenance & relapse prevention; Evaluation; Documentation; and Data protection.

Boudreaux et al. (2014, p. 363) proposed the following seven strategies for evaluating the quality of mHealth apps:

"(1) Review the scientific literature,

(2) Search app clearinghouse websites,

(3) Search app stores,

(4) Review app descriptions, user ratings, and reviews,

(5) Conduct a social media query within professional and, if available, patient networks,

(6) Pilot the apps, and

(7) Elicit feedback from patients."

Although such a comprehensive evaluation process can help health care providers in selecting reliable mHealth apps, it might require too much effort to be efficiently used in routine practice. Standardized app quality assessment tools can therefore help facilitate this process.

One widely used instrument for specifically assessing the quality of apps is the Mobile App Rating Scale (MARS; Stoyanov et al., 2015). MARS is a 23-item multidimensional app quality assessment tool with the following five dimensions: *Engagement; Functionality; Aesthetics; Information quality;* and *Subjective quality*. The user version (uMARS) consists of 20 items in the same dimensions (Stoyanov, Hides, Kavanagh, & Wilson, 2016). Another tool for measuring the quality of apps is *Enlight*, a multidimensional set of scales that was developed for the quality assessment of internet-based electronic health (eHealth) programs (Baumel, Faber, Mathur, Kane, & Muench, 2017), but can also be employed for the assessment of app-based (mHealth) interventions for (mental) health related behavior. To our knowledge, it

is the first instrument to incorporate the concepts of therapeutic persuasiveness and therapeutic alliance into the quality assessment of web-based and app-based interventions. Both of these assessment tools are valuable in the process of mHealth app quality assessment but have some limitations. While MARS is easy to use, it does not include all relevant criteria that have to be considered when assessing mHealth app quality and shows psychometric inconsistencies such as only moderate inter-rater reliability described on some items (Stoyanov et al., 2015). *Enlight* consists of a quality assessment section and an additional checklist section, which increases the amount of information that can be obtained for the assessment, although the differing answer formats complicate the assessment process. Going one step further, assessing an app's quality during the development phase – before its release to the public – would facilitate the process of selecting reliable mHealth apps for both health care providers and end users by making developers responsible for ensuring the quality of an app. In this context, our aim was to create an app quality assessment tool to assess an app's quality during the development process, that can be used by both experts involved in the development process and possible users. This would ensure all quality dimensions are optimized before an app is released to the public.

Table 1. Similarities and Differences between MARS and Enlight

MARS	Enlight
5 Dimensions of quality criteria (MARS & uMARS)	7 Dimensions of quality criteria
Engagement	User engagement
Aesthetics	Visual design
Information quality	Content
Functionality	Usability
Subjective app quality	General subjective evaluation
	Therapeutic persuasiveness
	Therapeutic alliance
	Checklist section
	Credibility
	Privacy explanation
	Basic security
	Evidence-based program ranking

We therefore wanted to create a multidimensional mHealth app quality assessment tool that combines the strengths of the above-mentioned instruments by not only assessing further relevant quality criteria, but also by establishing an easy and efficient evaluation process. In contrast to the existing tools, we want to create an assessment instrument specifically tailored to people involved in app development, to obtain valid app ratings. As Grundy, Wang, & Bero (2016) suggest involving end users in the study design when developing health-related apps, as well as using multi-dimensional quality assessment tools, another aim was to also create a user version of AQUA to enable a quality assessment process of health-related apps from the user's point

14

of view. Such a tool would allow end users to be involved in the development process, which would give app developers the opportunity to improve certain aspects that are important to end users before releasing the app to the public.

2 Method

The AQUA questionnaire was developed in the context of the EU Project CHRODIS+. CHRODIS+ is an initiative of 21 EU-countries to improve the treatment of chronic diseases with a three-year time frame (2017-2020) As part of this project, health-related apps for patients with chronic diseases, such as Tinnitus, are being developed and tested respectively by developers and app users in a two-step process. The principal aim of the development of AQUA is to facilitate this process by making it possible for both experts and end users to assess app-quality and demonstrate which dimensions of an app need to be improved. The primary approach to the development of the AQUA app was to combine the existing set of criteria included in MARS (Stoyanov et al., 2015) and *Enlight* (Baumel et al., 2017) and format it in a coherent way to enable a more efficient app-quality assessment process. We constructed a pool of items based on the dimensions included in these instruments. Additionally, we conducted an exploratory literature review to identify further app-quality criteria not mentioned in *Enlight* or MARS. English and German language articles from the years 2010 to 2019 were retrieved from PsycInfo and Scopus. The search terms used were "mobile" OR "mhealth" AND "app*" PAIRED WITH "quality" AND "assess*" OR "criteria" OR "evaluat*". Relevant criteria for the development of AQUA were gathered using a novel guideline for the certification of internet-based self-management interventions proposed by Klein et al. (2018). They introduced a set of eight criteria, with 17 sub-criteria, that internet-based interventions including mobile app-based interventions must fulfill in order to be certified. The central criteria are therapeutic quality requirements, patient and data security, and evidence of efficacy. Klein et al. suggest differentiating between so-called knockout (K.O.)-criteria, that must be fulfilled by an intervention for it to be certified, and descriptive quality criteria, which should be fulfilled but do not influence whether an intervention receives certification. Named K.O.-criteria include: transparency regarding the aim of the intervention; an evidence base; and a detailed and comprehensible description of the intervention including disclosure of eventual costs. Descriptive criteria include transparency regarding which user data is saved, where it is saved, and for what purposes. Ease of use and intuitive navigation are also included in the descriptive criteria (Klein et al., 2018).

The criteria extracted as a result of the literature review were categorized into dimensions and added to the already existing dimensions extracted from MARS and *Enlight*. Following this categorization, items were formulated in a way to assess the extent to which health-related apps adhere to these criteria from both an expert and a user point of view. After the items were

created, we conducted a pretest of the user questionnaire similar to cognitive pretesting with a convenience sample of possible future users, which consisted of six participants (3 women and 3 men, aged 19 to 58 years old). We chose participants who had already installed some mHealth apps on their phones, and expressed an interest in being involved in the process of mHealth app development by assessing an app's quality before release. We pretested the user version of the questionnaire to ensure that the items were as clear as possible to future participants not involved in the development of apps. After verbal consent was obtained from participants, they were asked to each test a health-related app already installed on their phone, before assessing the quality of this app with the questionnaire. The tested apps were Headspace[3], Runtastic[4], Moodpath[5], and TinnitusTipps[6]. Only the German version of the questionnaire was pretested, as all the participants were German-speakers. In cognitive pretesting, the participants are usually asked to think aloud while filling out a questionnaire (Hilton, 2017). In our case, we asked the participants to write any emerging thoughts as well as positive or negative feedback next to the items in the questionnaire while filling it out. The items were then optimized based on the resulting feedback.

3 Results

These methods resulted in the following eight basic dimensions of app-quality: Usability; User Engagement; Content; Visual Design; Therapeutic Quality; Security; Information; and Impact. We developed both an expert version and a user version of AQUA to assess mHealth app quality from different perspectives. All items use a 5-point Likert scale from 5 – strongly agree to 1 – strongly disagree, except for the item *email confirmation* which uses a yes/no format. By calculating the mean scores for each dimension, the quality of an app with regard to each of the dimensions can be determined and compared. A total score can also be calculated by calculating the overall mean. This is less precise, but it allows for the comparison of different health-related apps. Both versions of AQUA exist in English and German languages.

4 Expert Version

The expert version of AQUA is designed for persons who are actively involved in the development of apps and are thus more aware of the aspects to be considered in the design of an app than users or even (mental) health professionals. It consists of 31 items in the following seven dimensions: Usability (4); User Engage-

---

[3] https://www.headspace.com

[4] https://www.runtastic.com/de/

[5] https://mymoodpath.com/en/

[6] https://tinnitustipps.lenoxug.de/

ment (5); Content (5); Visual Design (4); Therapeutic Quality (5); Security (3); and Information (5). The dimension Usability is composed of the items: *Performance*; *Navigation*; *Learnability*; and *Ease of Use.* User Engagement is composed of: *Personalization*; *Interest*; *Entertainment*; *Interaction*; and *Gamification.* Content comprises the items: *Evidence-Base*; *Information*; *Completeness & Conciseness*; *Goal & Purpose;* and *Sequence.* Visual design contains: *Aesthetics*; *Graphics*; *Format;* and *Size.* Therapeutic Quality comprises the items: *Call to Action*; *Therapeutic Principle*; *Expectations*; *Adaptive Content;* and *Therapeutic Alliance.* Security includes: *Privacy Policy*; *Subjective Security;* and *Credibility.* Finally, the dimension Information consists of the items: *Developer*; *Frequency of Use;* *Costs*; *Crisis Support;* and *Email.*

5   User Version

The user version consists of 31 items in the following 7 dimensions: Usability (4); User Engagement (5); Content (4); Visual Design (4); Therapeutic Quality (5); Impact (4); and Information (5). The Impact dimension was added to the user version of the questionnaire to assess the subjective effectiveness of a health-related app from the user's point of view. Most other items are consistent with the expert version. The item *Evidence-Base* was not included in the user version because it is not a subjective quality criterion but an objective one and should be assessed by health-care professionals. The items included in the Impact dimension are: *Achievement*; *Symptoms*; *Health Behavior;* and *Recommendation*.

6   Discussion

The principal aim of the development of AQUA was to facilitate the assessment process of newly developed mHealth apps by making it possible for app-quality to be assessed by both experts and users, thereby enabling identification of the specific dimensions of an app that need to be improved. In contrast to MARS, which does not explicitly specify the difference between users and experts, we specify experts as persons involved in the development of mHealth apps. The resulting dimensions in the expert and user versions of AQUA resemble in large part the seven main categories of app quality criteria proposed by Nouri et al. (2018).

Calculating a total score for all items increases the comparability between different apps but contradicts the purpose of multidimensionality. For this reason, AQUA was formatted in a way that allows for the determination of a score for each dimension by calculating the mean item score for each dimension. This procedure preserves the multidimensionality and makes it easier for developers to identify the strengths and weaknesses of an app, while the calculation of a total score permits global comparison of different health-related apps. The same level of measurement scales is used throughout the questionnaire to al-

low for factor and item analyses in the future. Data security is a crucial aspect for health- and mental health-related apps. Ideally, the assessment of security and data protection quality in the field of mHealth should be undertaken in cooperation with experts in that field and persons involved in the process of developing such apps. This is why we developed an expert version of AQUA explicitly for people actively involved in the development of mHealth apps.

7   Limitations

A study reviewing data security and privacy policies of health apps focused on depression found that most of them lack transparency regarding data security (O'Loughlin, Neary, Adkins, & Schueller, 2018). Assessing the quality of data security is therefore a crucial part of improving the use of health-related apps, and the development of methods to identify possible risks to personal privacy and data security is important (Grundy et al., 2016). The dimension of security and data protection in AQUA takes this aspect into consideration, yet in its current state does not cover enough criteria of the data protection aspect of mHealth. We therefore plan to improve AQUA in future collaborations with data protection specialists. This was an exploratory pilot study with the aim of creating a preliminary mHealth app quality assessment tool which can be used in the development of mHealth apps. Another limitation is that we did not conduct an exhaustive literature review as this was not the aim of this study, and thus did not cover all of the relevant literature on the topic of app quality assessment. Furthermore, the current sample size did not allow for reliable testing of the psychometric qualities of AQUA, which is why we plan to conduct this testing in a future study with a larger sample size.

Pretesting was only undertaken with the German user version of the questionnaire. The expert version and the English versions should be tested with the same procedure in the future. Furthermore, AQUA does not establish K.O.-criteria and all of the items in AQUA have the same weight. This approach was taken to ensure an efficient assessment process and for better comparability between the dimensions. To highlight the importance of data security and evidence-based content, it could be relevant to more clearly distinguish between the user experience, data security, and the evidence base and to reflect this in the further development of specific expert questionnaires for (mental) health professionals and app developers. AQUA does not assess any ethical questions. As this is also an important aspect to consider in the field of mHealth (Nouri et al., 2018), this should be incorporated in the further development of this tool.

8   Conclusions and Future Prospects

The AQUA is a brief multidimensional app-quality assessment tool that enables a quick and efficient assessment of health-re-

lated apps by both experts and users in both the English and German language. We plan to test the psychometric qualities of this instrument in a study with a stress-monitoring app and a larger sample. We furthermore plan to obtain approval from an ethics commission for that study. As the development of any scientifically valid mHealth app should also include clinicians, we plan to create an expert version of AQUA specifically targeted at (mental) health professionals. In addition to the questionnaires, a semi-structured Interview guide has been developed to gather more detailed feedback from app-users and is available in both English and German.

Future studies in this area should especially focus on the dimensions of privacy and data security. While the dimensions of user engagement, usability, and functionality are crucial for users, the dimensions of data security, as well as content and therapeutic quality, are important aspects for experts rating the quality of an app. Future studies could also investigate how the quality ratings of experts and users differ from one another and what aspects of specific dimensions are important to these two groups. For example, the dimension of user engagement is highly relevant for users, as continuous user engagement with mHealth apps provides a challenge and various factors can influence user engagement (Mclean, 2018). Incentive management is a promising approach to ensuring ongoing user engagement with mHealth apps (Agrawal et al., 2018). Gamification and serious gaming is another approach that should be considered in this regard when assessing the quality of mHealth apps, although most gamification applications seem to yield only short-term effects on user engagement (Sardi, Idri, & Fernández-Alemán, 2017). Therefore, more research is needed on how to improve user engagement with serious gaming and how to identify valuable gamification aspects to assess app quality.

9 References

Abroms, L., Padmanabhan, N., Thaweethai, L., & Phillips, T. (2011). Iphone apps for smoking cessation: A content analysis. *American Journal of Preventive Medicine, 40,* 279–285. doi: 10.1016/j.amepre.2010.10.032

Agrawal, K., Mehdi, M., Reichert, M., Hauck, F., Schlee, W., Probst, T., & Pryss, R. (2018). Towards incentive management mechanisms in the context of crowdsensing technologies based on TrackYourTinnitus insights. *Procedia Computer Science, 134*, 145–152. doi: 10.1016/j.procs.2018.07.155

Baldwin, J., Singh, H., Sittig, D., & Giardina, T. (2017). Patient portals and health apps: Pitfalls, promises, and what one might learn from the other. Healthcare, 5, 81–85. doi: 10.1016/j.hjdsi.2016.08.004. doi: 10.1089/tmj.2012.0110

Barreda-Pérez, M, De la Torre, I, López-Coronado, M. (2013). Development and evaluation of a web-based tool to estimate type 2 diabetes risk: Diab_Alert. *Telemed J E Health, 19,* 81–87. doi: 10.1089/tmj.2012.0110

Baumel, A., Faber, K., Mathur, N., Kane, J., & Muench, F. (2017). Enlight: A comprehensive quality and therapeutic potential evalua-

tion tool for mobile and web-based ehealth interventions. *Journal of medical Internet research, 19*(3), e82. doi:10.2196/jmir.7270

Boudreaux, E. D., Waring, M. E., Hayes, R. B., et al. (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

BinDhim, N. F., Hawkey, A. & Trevena, L. (2015). A systematic review of quality assessment methods for smartphone health apps. *Telemedicine and E-Health, 21*, 97–104. Doi: 10.1089/tmj.2014.0088

Grundy, Q. H., Wang, Z. & Bero, L. A. (2016). Challenges in assessing mobile health app quality: A systematic review of prevalent and innovative methods. *American Journal of Preventetive Medicine, 51 (6),* 1051–1059. doi: 10.1016/j.amepre.2016.07.009

Hilton, C. (2017). The importance of pretesting questionnaires: A field research example of cognitive pretesting the Exercise referral Quality of Life Scale (ER-QLS). *International Journal of Social Research Methodology, 20*, 21–34. doi: 10.1080/13645579.2015.1091640

Hundert, A. S., Huguet, A., McGrath, P. J., Stinson J. N. & Wheaton, M. (2014). Commercially available mobile phone headache diary apps: A systematic review. *Journal of Medical Internet Research Mhealth Uhealth, 2*(3):e36. doi: 10.2196/mhealth.3452

Kazdin, A. E. (2017). Adressing the treatment gap: a key challenge for extending evidence-based psychosocial interventions. *Behaviour Research and Therapy, 88,* 7–18. doi: 10.1016/j.brat.2016.06.004

Klein, J. P., Knaevelsrud, C., Bohus, M. et al. (2018). Internetbasierte Selbstmanagementinterventionen. *Nervenarzt, 89* (11), 1277–1286. doi: https://doi.org/10.1007/s00115-018-0591-4

Kohl, L. F., Crutzen, R. & de Vries, N. K. (2013). Online prevention aimed at lifestyle   behaviors: a systematic review of reviews. *Journal of Medical Internet Research, 15*(7):e146. doi: 10.2196/jmir.2665.

Ly, K. H., Ly, A. M. & Anderson, G. A. (2017). A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. *Internet Interventions, 10*, 39–46. doi: 10.1016/j.invent.2017.10.002

Mclean, G. (2018). Examining the determinants and outcomes of mobile app engagement – A longitudinal perspective. *Computers in Human Behavior, 84*, 392–403. doi: 10.1016/j.chb.2018.03.015

Mcmillan, B., Hickey, E., Patel, M., & Mitchell, C. (2016). Quality assessment of a sample of mobile app-based health behavior change interventions using a tool based on the National Institute of Health and Care Excellence behavior change guidance. *Patient Education and Counseling, 99*(3), 429–435. doi: 10.1016/j.pec.2015.10.023

National Institute for Health and Care Excellence (2014). Behavior change: Individual approaches, *NICE*, London. https://www.nice.org.uk/guidance/ph49

Nouri, R., R Niakan Kalhori, S., Ghazisaeedi, M., Marchand, G., & Yasini, M. (2018). Criteria for assessing the quality of mHealth apps: A systematic review. *Journal of the American Medical Informatics Association, 25*, 1089–1098. doi: 10.1093/jamia/ocy050

O'Loughlin, K., Neary, M., Adkins, E. C., & Schueller, S. M. (2018). Reviewing the data security and privacy policies of mobile apps for depression. *Internet interventions*, *15*, 110–115. doi: 10.1016/j.invent.2018.12.001

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

Shen, N., Levitan, M., Johnson, A., Bender, J., Hamilton-Page, M., Jadad, A., & Wiljer, D. (2015). Finding a Depression App: A Review

and Content Analysis of the Depression App Marketplace. *Jmir Mhealth And Uhealth, 3,* E16.

Stoyanov, S. R, Hides, L., Kavanagh, D. J., Zelenko, O., Tjondronegoro, D. & Mani, M. (2015). Mobile app rating scale: A new tool for assessing the quality of health mobile apps. *Journal of Medical Internet Research Mhealth Uhealth,* 3(1): e27. doi: 10.2196/mhealth.3422

Stoyanov, S. R., Hides, L., Kavanagh, D. J. & Wilson, H. (2016). Development and validation of the user version of the mobile application rating scale (uMARS). *Journal of Medical Internet Research Mhealth and Uhealth, 4*(2), e72. doi: 10.2196/mhealth.5849

Webb, T., Joseph, J., Yardley, L. & Michie, S. (2010). Using internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. *Journal of Medical Internet Research, 12*(1):e4. Doi: 10.2196/jmir.1376

*Corresponding Author

Teresa O'Rourke, BSc.
Danube University Krems
Department for Psychotherapy and Biopsychosocial Health
Dr.-Karl-Dorrek-Straße 30
3500 Krems an der Donau
Tel: +43 2732 893-2531
Mail: teresa.orourke@donau-uni.ac.at

Declaration of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Conflicts of Interest

The authors have no conflict of interest to report.

Funding

Supplementary Material

Table 2. English Expert Version

	strongly agree	agree	undecided	disagree	strongly disagree
Usability					
Performance: The app functions without problems (bugs, crashing, etc.).	5	4	3	2	1
Navigation: It is easy to navigate through the app.	5	4	3	2	1
Learnability: It is easy to learn how to use the app.	5	4	3	2	1
Ease of use: It is easy to use the app.	5	4	3	2	1
User Engagement					
Personalization: Settings (e.g. volume or brightness) can be customized.	5	4	3	2	1
Interest: The app is interesting.	5	4	3	2	1
Entertainment: It is fun to use the app.	5	4	3	2	1
Interaction: The app has interactive functions (e.g. user input, sharing, reminders, etc.)	5	4	3	2	1
Gamification: The app uses game-typical elements such as rewards, point systems, badges, etc.	5	4	3	2	1
Content					
Evidence base: The app offers evidence-based measures to accomplish its intended purpose.	5	4	3	2	1
Information: The information is depicted in a coherent way (complexity, grammar, etc.).	5	4	3	2	1
Completeness & Conciseness The content is complete, but not redundant or irrelevant.	5	4	3	2	1
Goal & Purpose: The goal and purpose of the app is clear.	5	4	3	2	1
Sequence: The sequence of elements (features, images, exercises, etc.) is reasonable.	5	4	3	2	1
Visual Design					
Aesthetic: The app's visual design is appealing (colors, images, fonts, etc.).	5	4	3	2	1
Graphics: The quality of the graphics is good.	5	4	3	2	1
Format: The app is well structured.	5	4	3	2	1
Size: The size of the elements is appropriate and can be customized if necessary.	5	4	3	2	1
Therapeutic Quality					
Call to action: The app offers actions that can be successfully completed with little effort.	5	4	3	2	1
Therapeutic Principle: The underlying therapeutic principle is clear.	5	4	3	2	1
Expectations: It is clear what is expected of the user.	5	4	3	2	1
Adaptive Content: The app content adapts to the user's progress.	5	4	3	2	1
Therapeutic Alliance The app establishes a therapeutic alliance with the user (e.g. by chatting) through a human factor (doctor, therapist, avatar).	5	4	3	2	1
Security					
Privacy Policy Declaration: It is explained to the user whether and for what purpose user data is stored and how it is protected.	5	4	3	2	1
Data Protection: The app undertakes relevant data protection measures.	5	4	3	2	1
Credibility: The app was developed by a credible source (existing website, contact information, etc.).	5	4	3	2	1
Information					
Developer: Who developed, provides and distributes the app is displayed transparently.	5	4	3	2	1
Frequency of Use: The app's ideal frequency of use is displayed in a transparent way.	5	4	3	2	1
Costs: All costs of the app are transparently disclosed.	5	4	3	2	1
Crisis Support: The app provides advice on how to handle psychological crises. Contact information for emergency support services or helplines is displayed.	5	4	3	2	1
Email Confirmation: Is an email confirmation required to use the app?	yes	no			

Table 3. German Expert Version

	trifft zu	trifft eher zu	teils – teils	trifft eher nicht zu	trifft nicht zu
Usability					
Leistung: Die App funktioniert ohne Probleme (Programmfehler, Abstürzen, etc.).	5	4	3	2	1
Navigation: Es ist einfach sich durch die App zu navigieren.	5	4	3	2	1
Erlernbarkeit: Die Benutzung der App ist einfach zu erlernen.	5	4	3	2	1
Benutzerfreundlichkeit: Die App ist einfach zu benutzen.	5	4	3	2	1
User Engagement					
Personalisierung: Einstellungen wie z.B. Lautstärke oder Helligkeit können personalisiert werden.	5	4	3	2	1
Interesse: Die App ist interessant.	5	4	3	2	1
Unterhaltung: Es macht Spaß, die App zu benutzen.	5	4	3	2	1
Interaktion: Die App hat interaktive Funktionen (Nutzerinput, Inhalte teilen, Erinnerungen, etc.).	5	4	3	2	1
Gamification: Die App verwendet spieltypische Elemente, wie Belohnungen, Punkte, Abzeichen, etc.	5	4	3	2	1
Inhalt					
Evidenzbasis: Es werden evidenzbasierte Maßnahmen bereitgestellt, um den Zweck der App zu erreichen.	5	4	3	2	1
Informationsdarstellung: Die Informationen werden in einer klaren und verständlichen Sprache dargestellt (bzgl. Komplexität, Grammatik, etc.)	5	4	3	2	1
Vollständigkeit & Prägnanz: Der Inhalt ist vollständig, aber nicht ausschweifend oder irrelevant.	5	4	3	2	1
Ziel & Zweck: Ziel und Zweck der App sind klar.	5	4	3	2	1
Reihenfolge: Die Reihenfolge der Elemente (Funktionen, Bilder, Übungen, etc.) ist sinnvoll.	5	4	3	2	1
Visuelle Gestaltung					
Ästhetik: Das visuelle Design der App ist ansprechend (Farben, Bilder, Schriftarten, etc.).	5	4	3	2	1
Grafik: Die Qualität der Grafik ist gut.	5	4	3	2	1
Format: Die App ist gut strukturiert.	5	4	3	2	1
Größe: Die Größe der Elemente ist passend.	5	4	3	2	1
Therapeutische Qualität					
Handlungsaufforderung: Die App beinhaltet Handlungen, die mit wenig Aufwand erfolgreich durchgeführt werden können.	5	4	3	2	1
Therapeutisches Prinzip: Das zugrundeliegende therapeutische Prinzip ist klar.	5	4	3	2	1
Erwartungen: Die Erwartungen der App an den Nutzer sind klar.	5	4	3	2	1
Adaptiver Inhalt: Die App passt sich dem Fortschritt des Nutzers an.	5	4	3	2	1
Therapeutisches Bündnis: Die App stellt durch einen menschlichen Faktor (Arzt, Therapeut, etc.) ein therapeutisches Bündnis mit dem Nutzer her (z.B. durch Chatten).	5	4	3	2	1
Sicherheit					
Datenschutzerklärung: Dem Nutzer wird erklärt ob und wofür Nutzerdaten gespeichert und wie sie geschützt werden.	5	4	3	2	1
Datenschutz: Die App unternimmt entsprechende Datenschutzmaßnahmen.	5	4	3	2	1
Vertrauenswürdigkeit: Die App wurde von einer vertrauenswürdigen Quelle entwickelt (z. B. vorhandene Kontaktinformationen, Website, etc.).	5	4	3	2	1
Information					
Entwickler: Es liegt ein Impressum vor, in dem transparent beschrieben wird, wer die App entwickelt hat, anbietet und vertreibt.	5	4	3	2	1
Verwendungshäufigkeit: Es wird transparent darüber informiert, in welcher Häufigkeit und Frequenz die App genutzt werden sollte.	5	4	3	2	1
Kosten: Die vollständigen Kosten der App werden transparent dargestellt.	5	4	3	2	1
Krisenunterstützung: Die App gibt Empfehlungen zum Umgang mit psychischen Krisen. Es werden Hilfsangebote bzw. Kontaktinformationen für Notfälle angegeben.	5	4	3	2	1
Emailbestätigung: Wird eine Emailbestätigung benötigt, um die App benutzen zu können?	ja	nein			

Table 4. English User Version

	strongly agree	agree	undecided	disagree	strongly disagree
Usability					
Performance: The app functions without problems (bugs, crashing, etc.).	5	4	3	2	1
Navigation: It is easy to navigate through the app.	5	4	3	2	1
Learnability: It is easy to learn how to use the app.	5	4	3	2	1
Ease of use: It is easy to use the app.	5	4	3	2	1
User Engagement					
Personalization: Settings (e.g. volume or brightness) can be customized.	5	4	3	2	1
Interest: The app is interesting.	5	4	3	2	1
Entertainment: The app is fun to use.	5	4	3	2	1
Interaction: The app has interactive functions (e.g. user input, sharing, reminders, etc.).	5	4	3	2	1
Gamification: The app uses game-typical elements such as rewards, point systems, badges, etc.	5	4	3	2	1
Content					
Quality of information: The information is depicted in a coherent way (complexity, grammar, etc.).	5	4	3	2	1
Quantity of information: The content is complete, but not redundant or irrelevant.	5	4	3	2	1
Purpose and Goals: The goal and purpose of the app is clear.	5	4	3	2	1
Sequence: The sequence of elements (features, images, exercises, etc.) is reasonable.	5	4	3	2	1
Visual Design					
Aesthetic: The app's visual design is appealing (colors, images, fonts, etc.).	5	4	3	2	1
Graphics: The quality of the graphics is good.	5	4	3	2	1
Format: The app is well structured.	5	4	3	2	1
Size: The size of the elements is appropriate and can be adjusted if necessary.	5	4	3	2	1
Therapeutic Quality					
Call to action: The app offers actions that can be successfully completed with little effort.	5	4	3	2	1
Therapeutic Principle: The underlying therapeutic principle is clear.	5	4	3	2	1
Expectations: It is clear what is expected of me.	5	4	3	2	1
Adaptive Content: The app content adapts to my progress.	5	4	3	2	1
Therapeutic Alliance There is a human related to the app (doctor, therapist, etc.) I feel connected to (e.g. through chatting).	5	4	3	2	1
Impact					
Achievement: I have achieved the app's intended goal.	5	4	3	2	1
Symptoms: My symptoms have improved using this app.	5	4	3	2	1
Health behavior: My health behavior (mindfulness, habits, etc.) has improved using this app.	5	4	3	2	1
Recommendation: I would recommend this app.	5	4	3	2	1
Information					
Developer: Who developed, provides and distributes the app is disclosed transparently.	5	4	3	2	1
Frequency of Use: The app's ideal frequency of use is displayed in a transparent way.	5	4	3	2	1
Costs: All costs of the app are transparently disclosed.	5	4	3	2	1
Crisis Support: The app provides advice on how to handle psychological crises. Contact information for emergency support services or helplines is displayed.	5	4	3	2	1
Email Confirmation: Is an email confirmation required to use the app?	yes	no			

Table 5. German User Version

	trifft zu	trifft eher zu	teils – teils	trifft eher nicht zu	trifft nicht zu
Usability					
Leistung: Die App funktioniert ohne Probleme (Programmfehler, Abstürzen, etc.).	5	4	3	2	1
Navigation: Es ist einfach sich durch die App zu navigieren.	5	4	3	2	1
Erlernbarkeit: Die Benutzung der App ist einfach zu erlernen.	5	4	3	2	1
Benutzerfreundlichkeit: Die App ist einfach zu benutzen.	5	4	3	2	1
User Engagement					
Personalisierung: Einstellungen wie z.B. Lautstärke oder Helligkeit können personalisiert werden.	5	4	3	2	1
Interesse: Die App ist interessant.	5	4	3	2	1
Unterhaltung: Es macht Spaß, die App zu benutzen.	5	4	3	2	1
Interaktion: Die App hat interaktive Funktionen (Nutzerinput, Inhalte teilen, Erinnerungen, etc.).	5	4	3	2	1
Gamification: Die App verwendet spieltypische Elemente, wie Belohnungen, Punkte, Abzeichen, etc.	5	4	3	2	1
Inhalt					
Informationsdarstellung: Die Informationen werden in einer klaren und verständlichen Sprache dargestellt (bzgl. Komplexität, Grammatik, etc.)	5	4	3	2	1
Vollständigkeit & Prägnanz: Der Inhalt ist vollständig, aber nicht ausschweifend oder irrelevant.	5	4	3	2	1
Ziel & Zweck: Ziel und Zweck der App sind mir klar.	5	4	3	2	1
Reihenfolge: Die Reihenfolge der Elemente (Funktionen, Bilder, Übungen, etc.) ist sinnvoll.	5	4	3	2	1
Visuelle Gestaltung					
Ästhetik: Das visuelle Design (Farben, Bilder, Schriftarten, etc.) der App ist ansprechend.	5	4	3	2	1
Grafik: Die Qualität der Grafik ist gut.	5	4	3	2	1
Format: Die App ist sinnvoll strukturiert.	5	4	3	2	1
Größe: Die Größe der Elemente ist passend.	5	4	3	2	1
Therapeutische Qualität					
Handlungsaufforderung: Die App beinhaltet Handlungen, die mit wenig Aufwand erfolgreich durchgeführt werden können.	5	4	3	2	1
Therapeutisches Prinzip: Das zugrunde liegende therapeutische Prinzip ist mir klar.	5	4	3	2	1
Erwartungen: Mir ist klar, welche Erwartungen die App an mich hat.	5	4	3	2	1
Adaptiver Inhalt: Die App passt sich an meinen Fortschritt an.	5	4	3	2	1
Therapeutisches Bündnis: Es gibt einen Menschen (Arzt, Therapeut, etc.) oder einen Avatar, mit dem ich mich verbunden fühle (z.B. durch Chatten).	5	4	3	2	1
Wirkung					
Zielerreichung: Ich habe das von der App angestrebte Ziel erreicht.	5	4	3	2	1
Symptome: Meine Symptome/mein Gesundheitszustand haben/hat sich durch die Nutzung der App verbessert.	5	4	3	2	1
Gesundheitsverhalten: Mein Gesundheitsverhalten (Achtsamkeit, Gewohnheiten, etc.) hat sich durch die Nutzung der App verbessert.	5	4	3	2	1
Empfehlung: Ich würde die App weiterempfehlen.	5	4	3	2	1

Continued Table 5. German User Version

	trifft zu	trifft eher zu	teils – teils	trifft eher nicht zu	trifft nicht zu
Information					
Entwickler: Es liegt ein Impressum vor, in dem transparent beschrieben wird, wer die App entwickelt hat, anbietet und vertreibt.	5	4	3	2	1
Verwendungshäufigkeit: Es wird transparent darüber informiert, in welcher Häufigkeit und Frequenz die App genutzt werden sollte.	5	4	3	2	1
Kosten: Die vollständigen Kosten der App werden transparent dargestellt.	5	4	3	2	1
Krisenunterstützung: Die App gibt Empfehlungen zum Umgang mit psychischen Krisen. Es werden Hilfsangebote bzw. Kontaktinformationen für Notfälle angegeben.	5	4	3	2	1
Emailbestätigung: Wird eine Emailbestätigung benötigt, um die App benutzen zu können?	ja	nein			