Evaluating the Requirements of Digital Stress Management Systems: A Modified Delphi Study

Kim Janine Blankenhagel(✉), Miriam Linker, and Rüdiger Zarnekow

Technical University Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany
k.blankenhagel@tu-berlin.de

Abstract. Stress is a major health problem in this century, and it is associated with adverse health consequences. Its prevention and management are a great challenge, and only a minority of the affected persons receive treatment. New digital technologies offer opportunities to provide effective psychological interventions to address the negative consequences of occupational stress. However, the knowledge of the importance of different functions and features of digital stress management systems remains largely unexplored. This work closes that research gap by conducting a Delphi study, in which 20 experts prioritized requirements in three rounds. The purpose of the present study is to enable developers of digital stress management systems (DSMS) to profitably select and use functions and characteristics of those systems, taking into account the available resources. Thus, the aim is to find DSMS that better counteract excessive stress. Finally, 82% of all requirements meet the consensus threshold.

Keywords: Digital stress management systems · Burnout prevention · eHealth

1 Introduction

Burnout is a major public health issue due to the generally elevated levels of stress and its growing complexity in the professional context [1]. Stress can be caused by work-related factors, such as excessive workload, or through individual sources, for example, work-family conflicts [2]. Long-term stress is not only associated with an increased risk of burnout but also with other mental disorders and somatic problems such as cardiovascular disease [3]. Undoubtedly, stress and the associated health consequences lead to high direct and indirect costs incurred by both employers and the society due to healthcare costs, lower productivity, and absenteeism [2]. Nevertheless, the rate of participation in conventional preventive services is low, and face-to-face stress reduction interventions require excessive human resource allocations and suffer time conflict issues [4]. At the same time, the technological capability to mine, interpret, and respond to a large amount of data for promoting the welfare of human subjects is growing [5]. Therefore, research continues to focus on the use of digital technologies. They are promising in terms of acceptability, effectiveness, and economic sustainability to support stress management and reduce the negative consequences of work-related stress [4]. In addition to adequate physical exercise, the ability to cope with stress is one of the most important factors in

G. M. P. O'Hare et al. (Eds.): MobiHealth 2019, LNICST 320, pp. 229–248, 2020.
https://doi.org/10.1007/978-3-030-49289-2_18

preventing burnout [6]. The term preventive digital stress management systems (DSMS), as defined in this article, encompasses coping strategies for the preservation of mental, physical, and social health. The aim is to reduce stress caused by work-related pressure and contribute to finding a healthy balance. DSMS are mainly provided in the digital form (derived from [7]). Currently, there are only a few DSMS on the market (e.g., applications on smartphones). Moreover, there are also certain initial studies dealing with the effectiveness and moderating factors of DSMS [8–10].

However, in general, the market is commercially driven, and the importance of the different functions and features of DSMS are poorly represented in the existing research. In order to close this gap, we conducted a Delphi study with 20 experts who prioritized the requirements for DSMS in three rounds. The obtained list of requirements was developed by the authors in 2018 using 15 semi-structured qualitative interviews. This list is the foundation for the present study. The purpose is to first confirm the requirements and then to enable developers of DSMS to profitably select and use the functions and characteristics, keeping in mind the available resources. Thus, the aim is to develop DSMS that better counteract excessive stress.

The Delphi study design is a suitable method for the present prioritization because it structures the necessary communication processes well and can produce a well-founded consensus due to its mixture of questionnaires and controlled opinion feedback.

2 Background

DSMS usually aim to reduce the symptoms of work-related stress, thereby increasing the wellbeing of users [11]. They provide new ways to feel, think, and act in stressful situations to reduce stressors, improve reactions to stressors, or mitigate the physiological or psychological effects of stress [2]. They focus on teaching different techniques to cope with stress [12] and differ in terms of various aspects, such as delivery, intervention content, length, or scope [10]. Often, DSMS are delivered in sessions over several weeks, whereby there are short systems (such as 2-week interventions); in contrast, some are also delivered over months or years [10]. In general, the interventions and exercises focus on an individual level and comprise meditation, mindfulness, breathing, and relaxation techniques, biofeedback, time management, and other cognitive behavioral elements [2, 13]. There are two types of systems, which differ in guidance. Guided interventions have some kind of human support, such as e-mail reminders or counselor support, whereas unguided ones have no support or only technical support [2].

The effectiveness of psychological and psychotherapeutic health interventions provided via the internet is frequently studied, and their results are promising [10]. Whether these findings can be transferred to DSMS and internet interventions can also be effective in prevention in the field of stress remains little studied and no conclusive evidence of effectiveness is available. A few analyses have been performed to examine the effectiveness of DSMS in comparison to a group on waiting list, which show a significant reduction of stress [3, 11, 14, 15]. Moreover, comparisons with a no-treatment group [16] or to an attention group [17] indicate the effectiveness of DSMS too. Elena Heber and Wasantha Jayawardene explored related research, and both of the meta-analyses demonstrate positive effects of DSMS [4, 10]. As many relevant parameters, such as

type and length of the interventions, usage of guidance, outcomes, measurements, or settings, vary in the studies, the results are not quite comparable and generalizability is difficult to assess [2, 18]. Furthermore, the effects of individual interventions of DSMS are mostly unexplored, sample sizes of the trials are often small, and the measurement of stress as a success control is often carried out solely through self-assessments (e.g., Perceived Stress Scale) rather than through objective methods, such as biomarkers (e.g. cortisol levels). It is striking that DSMS are often tested on groups, constituting people with relevant symptoms in a stressful period of life (e.g., healthcare professionals who have higher burnout, depression, and suicide rates) [13–15, 18–20]. Thus, from the perspective of a universal preventive stress setting, no conclusions can be drawn regarding people in their everyday life with lower stress levels. In addition, studies analyze outcomes such as stress, perceived stress, depression, etc. after a short time of using DSMS [10, 12, 19, 20]. Change in stress management, attitudes, and behavior are fundamental basic psychological processes, which may take a long period to manifest. Therefore, study periods may not be enough to cover long-term effects, and the first research results should be interpreted with care.

In summary, there is insufficient knowledge available of what makes DSMS successful in reducing stress and preventing burnout. Thus, we aim to bridge the research gap by prioritizing the requirements of DSMS.

3 Methods

This analysis is based on an explorative study conducted in 2018, replacing the first phase of a typical Delphi study. As it provides an underlying list of requirements, the study is crucial and described in more detail below. Afterward, an in-depth description of the methodical approach of this study is presented.

3.1 Preliminary Study

The requirements for DSMS were derived from a qualitative interview study and its analysis [21]. As the body of knowledge lacks assessment, interpretation patterns, and action orientations as well as the identification of individual perspectives, semi-structured, guideline-based interviews were chosen as a suitable methodology to identify the requirements. In order to better understand the specificity of the DSMS application, the interviews were conducted from four different perspectives (health insurance companies, care providers, the private sector, and users). The interview partners were selected according to Flickl following case selection using a qualitative sampling plan, in order to include a targeted selection of particularly meaningful cases [22]. The employed interview guideline was based on the principles established by Döring and Bortz and was deductively derived from the literature [23]. The interviews were analyzed as per the qualitative content analysis presented by Mayring, utilizing a software for qualitative analysis (MaxQDA) [24]. The requirements identified from the interviews were aggregated and coded, rule-based, on a developed coding guide, which defined the characteristics of the individual categories. The interview material was cross-validated, and

the Cohen's Kappa coefficient was determined to be 0.85 [21]. Table 1 collates all the requirements.

Table 1. Underlying Requirements

ID	Requirement
1	Consecutive, adaptive, small-step goals; behavioral goals derived from health goals
2	Reminder functions to goals, open exercises or other mature interactions; individually configurable
3	Short interaction times (about 5-15 min per day); quick check-in, clear structure and menu navigation, overviews
4	Simple and intuitive usability, fun to use
5	Hidden content, only a few technical terms
6	Highly personal and clear everyday relevance
7	Everyday suitability and high wearing comfort
8	Clear presentation of cause and effect relationships
9	High individualization including personal on-boarding, individual configuration options and tailoring to the user
10	Support to identify stressors and derive appropriate measures
11	Measurement of stress level
12	Customization to specific user situations
13	High autonomy; proactive construction
14	Feedback functions in acute stress situations and effectiveness of implemented measures and exercises; reports on medium- to long-term trends in stress levels
15	Diary or documentation function
16	Continuous measurements of heart rate variability and respiratory rate
17	Analysis of vocal pitch, muscular tension in the neck area and skin resistance
18	At least one stress-measuring functionality is available and transformed into an understandable main metric; detailed drill down capabilities for stress levels
19	Detection of deviations from the normal pattern
20	Adaption of stress monitoring over time
21	Interventions include exercises in psychological self-education
22	Interventions include exercises for self-reflection
23	Interventions include exercises to build up anti-stress resources
24	Interventions are scientifically driven
25	Possibility of integration of a doctor, psychotherapist, or medical professional
26	Adequate overview function for a supervising physician
27	Profound data concept regarding data security and data protection

(*continued*)

Table 1. (*continued*)

ID	Requirement
28	Encapsulated and encrypted storage and processing of users data
29	Anonymous and aggregated data transmission
30	Transparency and information regarding data usage
31	Raw data material available in a machine-readable and structured format; possibility of deletion
32	Integration of artificial intelligence for individualization
33	High mobility; offline use
34	Compatibility and integration with other applications

Finally, a total of 34 requirements in the domains of Human Centricity, Medicine, and Technology were determined, which are mainly non-functional. Therefore, it can be assumed that aspects of user-friendliness and the method of implementation are of great importance for the success of DSMS. As equivalent DSMS that meet all the requested requirements can be highly complex and very expensive, this study focused on validation and prioritization.

3.2 Follow-up Study

We used the established method of a Delphi study; it is based on the concept of pooled intelligence intended to enhance individual judgments and capture the collective opinion of experts [25]. Due to the incomplete state of knowledge about prioritization of the requirements of DSMS, we considered the Delphi study a suitable research tool to augment unanimity in opinions. Delphi studies that use ranking have been widely used in information systems research to develop group consensus regarding the relative importance of issues, particularly in health sciences [26–28]. The literature does not present a consistent definition of the method, but it is possible to define particular features characterizing its nature. It is an instrument for the improved recording of group opinions, the basic concept being the use of expert knowledge to solve problems in several iterations. The process is characterized by the addition of anonymous feedback regarding the general opinion of all experts after each round [29]. The averages produced in the decision-making processes of expert groups prove to be better than the averages of individual expert responses [30]. Furthermore, compared to other methods of forming opinions in groups (e.g., focus groups), the Delphi method neutralizes the influence of dominant opinion leaders (e.g., by reason of their authority, personality, reputation, etc.) due to its anonymity [31]. The involvement of experts and the use of a formalized questionnaire are further characteristics of Delphi studies [29]. As we have already derived the requirements of DSMS in the context of a previous qualitative study, the open-ended initial phase of a traditional Delphi process has been omitted in the present study, and this is, therefore, to be considered a modified Delphi study.

Figure 1 demonstrates the methodical approach, which is presented in detail below.

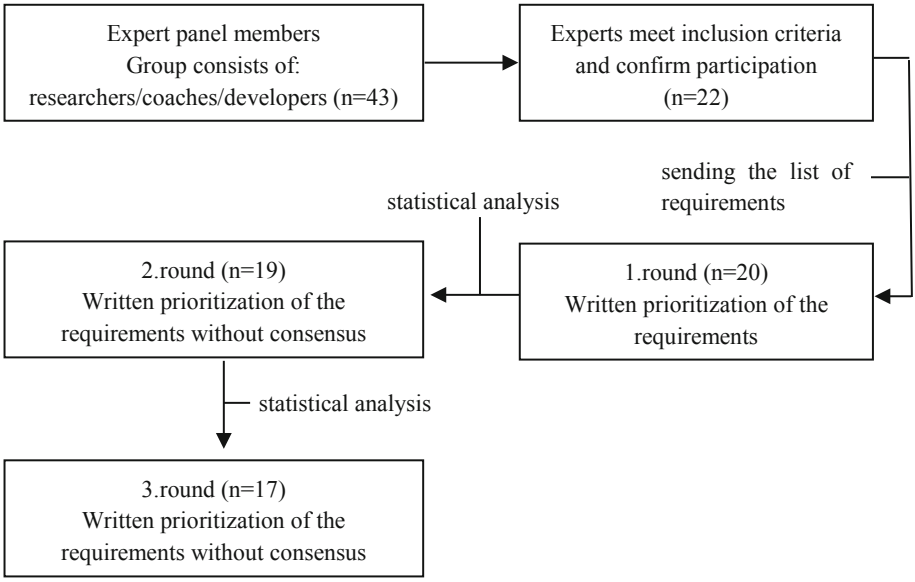

Fig. 1. Methodical approach

Expert Panel Selection. Initially, we recruited experts who are highly familiarized with the combination of stress management and digital services. In order to cover different perspectives, we asked experts to participate in the study who are active researchers, coaches/consultants, and developers. Scientists were selected according to their research activities and publications in the field of digital stress management, coaches and consultants according to their orientation and popularity, and developers according to their proximity with the topic. In total, we contacted 43 experts by telephone and e-mail as well as provided an information sheet regarding the research team, the content, the objective, and the proposed duration of the study. In addition, we asked the selected experts to self-assess their expert status in terms of stress management and digital health services by classifying them into four categories (low, average, high, very high) as compared to other experts in their field. As the level of knowledge of the expert panel is a critical success factor in Delphi studies [29], we only included experts in the study if they rated their expert knowledge as high or very high.

There is disagreement in the literature regarding the optimal scope of a Delphi expert group. According to Woudenberg, a panel size of three people is considered too small [32], and Brooks prefers a maximum of 25 experts [33]. Even though larger panels generally reduce any possible biases, in an experiment on the necessary panel size, Duffield showed that the results of two different groups ($n = 16$ versus $n = 34$) are in agreement by 92% [34]. On this basis, we aimed for a group of 15–25 experts. A total of 22 experts pledged their participation at the end of the recruitment phase.

Round 1. We conducted the first round of the Delphi study in October 2018. In order to reduce costs, time, and effort and avoid geographical boundaries, we conducted the

study digitally. The experts received the list of 34 requirements via e-mail and were asked to prioritize each requirement. In addition, open comment fields were provided. The scale consisted of the four classic categories listed below:

"Must": The requirement is absolutely essential.
"Should": The requirement should be met but is not absolutely essential.
"Could": The requirement could be met if a more valuable requirement did not interfere.
"Won't": The requirement should not be met

Out of the participants, 20 experts returned the prioritized list on time, and their replies were included in the evaluation.

The objective of the study was to build consensus among the expert group on the importance of individual requirements for effective and successful DSMS. Therefore, we defined the following two essential criteria for reaching consensus, and both had to be met:

1. 70% of all experts select the same category.
2. The interquartile range (IQR) is less than or equal to one.

With the two criteria, we ensured that consensus can only be reached if a clear majority of the experts had the same opinion (Criterion 1) and if the answers were little scattered (Criterion 2: 75% of all answers fluctuate between two categories at most). After the first round, we evaluated all the questionnaires, calculated the mean, mode, relative frequency, and interquartile range. Requirements that fulfilled both criteria were removed from the process as a consensus had already been reached. At the end of the first round, 15 of the 34 requirements were endorsed by the experts, and the remaining 19 did not meet the threshold.

Round 2. Subsequently, we sent a revised list of requirements to the same experts in a second round. In this round, only those requirements were sent that did not fulfill the previous criteria for consensus. In addition, we visually presented the results of the first round in general and also informed each expert of his own answer. The total distribution was represented by circles of different colors and sizes. They comprised the three categories "more than 40%…," "20–40%…", or "less than 20% of all experts chose the respective category." Then, we again asked the experts for another set of prioritizations, considering the general opinion, and received a total of 19 completed questionnaires at the end of the deadline. Afterward, we evaluated them in the same way as the first round. Overall, consensus was reached on 11 out of the 19 requirements.

Round 3. The third round included a repeated revision of the requirement list by excluding the requirements that had gained consensus previously. We again asked the experts to prioritize the remaining requirements and invited them to explain their choice if it differed from the general group perspective. From the participants, 17 experts participated in the third round, and consensus was reached on 2 of the remaining 8 requirements. Since the difference between the second and third rounds was negligible, we did not expect any further movements in a potential fourth round; thus, we completed the study.

4 Results

The participants in round 1 of the modified Delphi study self-identified as one of the following: coach (n = 9, 45%), developer of DSMS (n = 6, 30%), or researcher (n = 5, 25%). The proportions of the genders were almost at par (11 males, 9 females). Lastly, 70% of all participants reported very high expert status in both stress management and digital health services.

Table 2 anonymously lists the characteristics of all the participating experts from round 1.

Table 2. Expert characteristics

ID	Background	Gender	Expert level – stress management	Expert level – digital health services
1	Developer		High	Very high
2	Researcher	Female	Very high	Very high
3	Developer	Male	High	Very high
4	Developer	Male	High	Very high
5	Coach	Male	Very high	Very high
6	Coach	Male	Very high	High
7	Coach	Female	Very high	High
8	Coach	Female	High	High
9	Researcher	Male	Very high	Very high
10	Researcher	Male	Very high	Very high
11	Coach	Female	Very high	Very high
12	Developer	Female	Very high	Very high
13	Researcher	Female	Very high	Very high
14	Coach	Male	Very high	High
15	Coach	Male	Very high	Very high
16	Developer	Female	Very high	Very high
17	Coach	Female	High	High
18	Researcher	Male	High	High
19	Coach	Female	Very high	Very high
20	Developer	Male	Very high	Very high

4.1 Results of Round 1

The prioritization from the first round resulted in 15 requirements with consensus. Overall, it is especially striking that the requirements with meta-level and conceptual characteristics reached consensus (e.g., 1, 7, 12, 26). In contrast, concrete requirements about

implementation resulted in less consensus (e.g., 2, 16, 21–23). This indicates that experts agree on the importance of basic design issues, but that the importance of several functions leads to different opinions and still offers potential for discussion. In the following three sections we present the results of all 15 requirements with consensus in the first round.

Requirements with the Prioritization "Must" (n = 9). In total, 9 out of 15 requirements were classified as essential for successful DSMS. This includes the idea that DSMS should incorporate multi-purpose, adaptive, small-scale goals in order to create positive incentives for the user and motivate them (Requirement 1). Without exception, all coaches selected the category "Must." Their work requires an excellent understanding of the areas of goal setting and motivation. Therefore, their clear prioritization robustly confirms that setting appropriate goals also plays an important role in the context of digital systems and should definitely be implemented. Across all participating groups, 86% chose the category "Must". Thus, this requirement meets the consensus threshold (IQR = 0).

Furthermore, the experts agree that DSMS processing times should be very short (Requirement 3). The fast pace of (work) life today explains the increasing importance of time requirements. People who suffer from stress are willing to use a DSMS only if they can operate it quickly. This understanding is particularly pronounced among developers (all of them marked the category "Must"). With regard to sustainable utilization, simple and intuitive usability as well as fun to use (Requirement 4) are also essential requirements according to expert ratings. The consideration of a clear, everyday reference and high everyday suitability of DSMS are also highly important requirements (Requirements 6 and 7), and it should be ensured that any intervention is scientifically proven (Requirement 24). There is agreement among the participating experts that DSMS must have a strong data protection concept (Requirement 27) and must inform the user transparently and comprehensively about the use of the data (Requirement 30). Both requirements found 100% consensus among the researchers as well as among the developers. The provision of the raw data material in a machine-readable and structured format as well as the possibility of deletion are also very high priorities for the experts (Requirement 31).

Requirements with the Prioritization "Should" (n = 3). The experts consider a high degree of individualization (Requirement 9) of the DSMS to be preferable but not absolutely necessary. Tailoring to individual users' needs can be implemented, for example, by employing a software that gets to know the user, identifies his stressors, and supports him with personalized instructions and exercises. The answers of the coaches and researchers are all located in the two highest evaluation categories. Moreover, with a total of 74% agreement between all of the experts on "Should," Requirement 9 narrowly achieved the consensus criteria. With an interquartile range of 0, one check mark at "Could" (developer) and none at "Won't," we assume that the "Should" prioritization tends to be of higher importance. In addition to tailoring the system to a person, an individual adaption to the user's environment is also considered desirable by the experts (Requirement 12). It should be possible to disable or individually configure the interaction with the DSMS adapted to the circumstances of personal everyday life.

Among the researchers, there is absolute agreement on the prioritization of "Should," but the opinion among the developers is incoherent. This distribution can be traced to an increased degree of difficulty in implementation, thereby leading to disagreement or rather uncertainty among the developers. Furthermore, high autonomy as well as a proactive structure of the DSMS (Requirement 13) are seen as a "Should" requirement for appropriate application by the experts.

Requirements with the prioritization "Could" (n = 3). The experts agree that the measurement of stress levels on the basis of voice analysis, muscular tension in the neck area and skin resistance (Requirement 17) has a low priority and should only be done if there are available resources left after the implementation of more important requirements. Expert 02 comments that

> *"reliable and suitable stress detection [with the mentioned technique] is not yet satisfactorily possible."*

Due to the abstract nature of the symptom stress, its measurability is generally difficult and the experts believe that voice analyses, measurement of skin resistance as well as muscular tensions should be given low priority. No one selected the category Must, but two experts selected the category Won't. Similarly, the experts rate the involvement of medical specialists and the provision of overviews for physicians (Requirements 25 and 26) as low on importance. Because DSMS are designed to reduce excessive stress and prevent burnout, they are naturally preventive and take effect before the onset of a disease. Therefore, the focus is on self-management and the integration of physicians does not seem necessary.

Table 3 contains all of the requirements with a consensus after round 1, and also specifies the mean, mode and interquartile ranges.

Table 3. Requirements with consensus in round 1

Requirement	Mean	Mode	IQR
Must:			
6	3.9	4	0
30	3.89	4	0
27	3.85	4	0
4	3.8	4	0
1	3.75	4	0
3	3.7	4	0.5
24	3.65	4	1
7	3.6	4	0.5
31	3.63	4	1
Should:			

(*continued*)

Table 3. (*continued*)

Requirement	Mean	Mode	IQR
9	3.16	3	0
13	3	3	0
12	2.9	3	0
Could:			
26	2.35	2	0.5
25	2.15	2	0
17	2	2	0
Must ≙ 4, Should ≙ 3, Could ≙ 2, Won't ≙ 1			

4.2 Results of Round 2

The prioritization in the second round resulted in 11 requirements gaining consensus. Overall, the experts agree on the topics of stress measurement and the design of exercises (e.g., 11, 18, 21).

Requirements with the Prioritization "Must" (n = 8). The experts evaluate individually configurable reminder functions for goals, open exercises, or other interactions with the system (Requirement 2) as crucial features. In addition, a clear presentation of cause and effect relationships (Requirement 8) is highly important due to the motivating effect. For example, after a completed exercise, the DSMS visualizes the reduced stress levels and emphasizes the casual link. The developers particularly consider this presentation to be very important (100% chose the category "Must"), whereas the researchers and coaches vary between "Must" and "Should," with more focus on "Must".

Furthermore, it is very important that DSMS provide functions that support the identification of the user's stressors and, if necessary, derive appropriate measures to reduce stress (Requirement 10). In the same vein, the measurement of stress (Requirement 11) constitutes a basic function because it is suitable for drawing conclusions about possible stressors. This logic matches expert ratings, because Requirement 11 reaches consensus with the prioritization "Must." Both coaches and developers evaluate the two related requirements (10 and 11) similarly. Only the opinion of the researchers is slightly more inconsistent. In addition, agreeing on the great importance of measuring stress levels, the experts also consider the translation of the stress levels measured into an understandable metric and its detailed breakdown (for example, subdividing stress into low-, medium-, and high-stress phases or specifying the duration of the stress phase) to be essential (Requirement 18). With a relative frequency of 79% and an interquartile range of 1, this requirement meets the consensus threshold. Only the coaches express a few different opinions, and Expert 17 comments:

"The measurement and presentation of stress levels could be also counterproductive, because the affected person is confronted directly with having a very high level of stress."

In conclusion, the experts consider stress measurement as a meaningful and profitable functionality, although it cannot be ruled out that in individual cases, the existing stress load of the user may be further aggravated by its measurement and presentation.

Both the psycho-educative and educational interventions (Requirement 21) and reflective exercises (Requirement 22) are must-have requirements. The mean values indicate that the provision of reflective interventions is of a little more importance than psychoeducational ones (mean of 3.79 versus 3.68). The field of self-reflection includes exercises for better body perception (body scan), increased awareness and sensitization, as well as self-observation exercises. The present prioritization emphasizes that in addition to several new tracking functions, the stimulation of self-reflection also is vital in the area of digital systems. Expert 08 states:

"People should also get a feeling for their body and intuition again and learn this through self-reflection and questioning."

After the second round, the technical requirement for anonymous and aggregated data transmission (Requirement 29) also achieves the prioritization "Must". As 86% of the experts with a very high level knowledge of digital health technologies chose the "Must" category (among those with high expert status, it is only 40%) and because Requirement 29 necessitates deep technical understanding, the distribution indicates very high importance. If data transmission is not necessary, Expert 01 alternatively recommends that applications can work on a smartphone in a completely encapsulated way, including data processing, so that personalized systems can be implemented.

Requirements with the Prioritization "Should" (n = 1). The requirement of adapting the stress-measuring monitoring element to development over time and period of use (Requirement 20) reaches the prioritization of "Should." Therefore, it is desirable to offer frequent learning opportunities at the beginning of usage to quantify and evaluate one's own stress levels and identify the stressors. Afterward, medium- to long-term trends of stress development gain spotlight rather than the cause-effect relationships and individual situations.

Requirements with the Prioritization "Could" (n = 2). The detection of deviations from normal behavior (for example, the user only stays in the office and hardly ever leaves it) has a lower priority among the experts (Requirement 19). This suggests that such deviations can either not be sufficiently causally attributed to stress or that they provide little profitable information. Moreover, the possibility of integrating data and results from other applications (e.g., pedometer) into the DSMS (Requirement 34) is granted lower priority. Without exception, all answers are located in the two categories "Should" and "Could," whereby the experts with very high expert status in terms of digital health services tend more toward "Could" compared to those with high expert status (75% versus 66%).

Table 4 contains all the requirements with consensus after round 2 as well as specifies the mean, mode, and interquartile ranges.

Table 4. Requirements with consensus in round 2

Requirement	Mean	Mode	IQR
Must:			
22	3.79	4	1
2	3.68	4	1
18	3.68	4	1
21	3.68	4	1
8	3.67	4	1
10	3.63	4	1
11	3.58	4	1
29	3.58	4	1
Should:			
20	2.79	3	0
Could:			
34	2.26	2	1
19	2.21	2	0
Must $\triangleq$ 4, Should $\triangleq$ 3, Could $\triangleq$ 2, Won't $\triangleq$ 1			

4.3 Results of Round 3

In general, the third round showed only a few changes in the overall opinion and yielded two more requirements with consensus.

Requirements with the Prioritization "Must" (n = 2). The participating experts agree that in addition to providing psycho-educative and reflective exercises, those for building up anti-stress resources (Requirement 23) are also of utmost importance for improving the balance of stress periods and recreation (examples are resource analysis or time management). DSMS offering a high degree of mobility (Requirement 33) is also considered extremely important. Thus, they should be accessible on a smartphone or other devices that can be carried every day and used any time. Furthermore, a higher flexibility can be achieved if the application is available offline as well.

Table 5 presents all the requirements with consensus after round 3 and specifies the mean, mode, and interquartile ranges.

Table 5. Requirements with consensus in round 3

Requirement	Mean	Mode	IQR
Must:			
23	3.71	4	1
33	3.53	4	0
Must ≙ 4, Should ≙ 3, Could ≙ 2, Won't ≙ 1			

5 Discussion

After the three Delphi rounds, 28 of the 34 requirements reached the consensus threshold. Among them, not a single requirement was labeled with "Won't," so the requirements derived from the previous study can be all confirmed.

Figure 2 visualizes the development of the prioritization of the requirements over the three rounds. The inner circle refers to the first round, followed by the outward second, and finally the outermost third round. The requirement number is located in the middle of each circle.

Although the experts in the second round agree that DSMS should measure stress levels either way (see Fig. 2, Requirement 11/18: "Must"), there seems to be a strong disagreement on the suitable method of stress measurement. Requirements 15–17 present possible methods of stress measurement, with lower preference ("Could" prioritization) given to voice analyses, measurement of skin resistance, and muscular tension (Requirement 17). There is no consensus on the importance of diary and documentation functions (Requirement 15), which can be used in a broad sense to quantify stress levels. The answers range from "Must" to "Could" at this point. Moreover, the experts do not reach consensus on the continuous measurement of pulse, heart rate variability, and respiratory rate (Requirement 16). However, looking at the development of the prioritization over the three rounds, one sees an increasing focus on the categories of "Must" and "Should." In general, the results suggest that the measurement of pulse, heart rate variability, and respiratory rate is the most relevant method of stress measurement; however, the significance is not sufficient due to the prevailing disagreement. This can be attributed to the fact that stress is abstract compared to other diseases, thus rendering its measurement difficult. This uncertainty is also reflected in the market, as a lot of DSMS have not yet integrated stress measurement or only a rudimentary one, if any.

With respect to the interaction between the user and the system, it is noteworthy that although the experts agree in the second round that the reminder functions on goals, exercises, or other interactions are considered essential (Requirement 2), there is no consensus on the importance of feedback in acute stress situations (Requirement 14). Thus, it can be concluded that organizational reminders are really desired, but the importance of those in health-related content is unclear. There are hardly any recommendations on this concern in the literature either. Because all the check marks for Requirement 14 are located in the categories "Must" and "Should" in the last round, we can suppose that feedback functions are generally considered useful in acute stress situations. Some

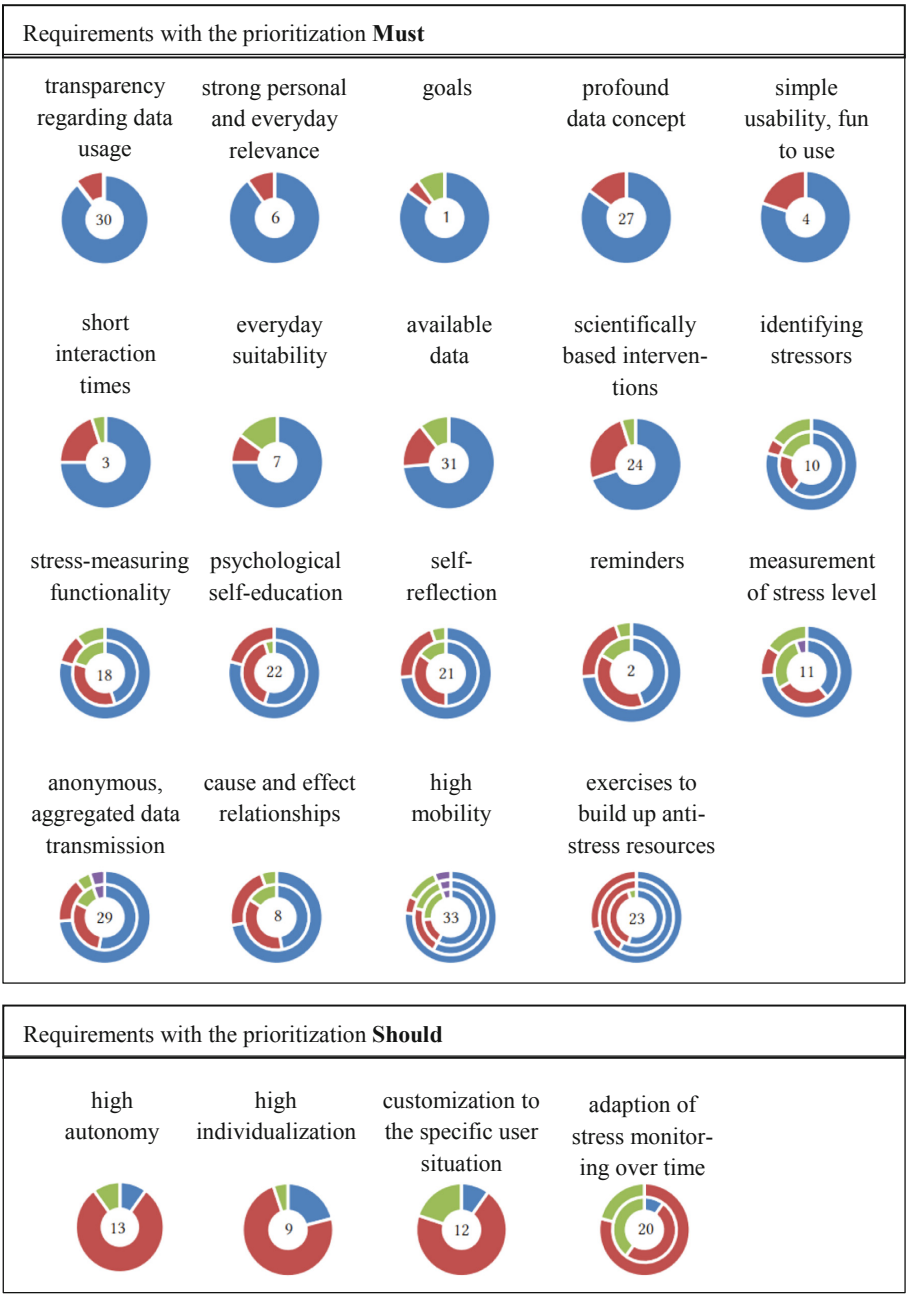

Fig. 2. Results

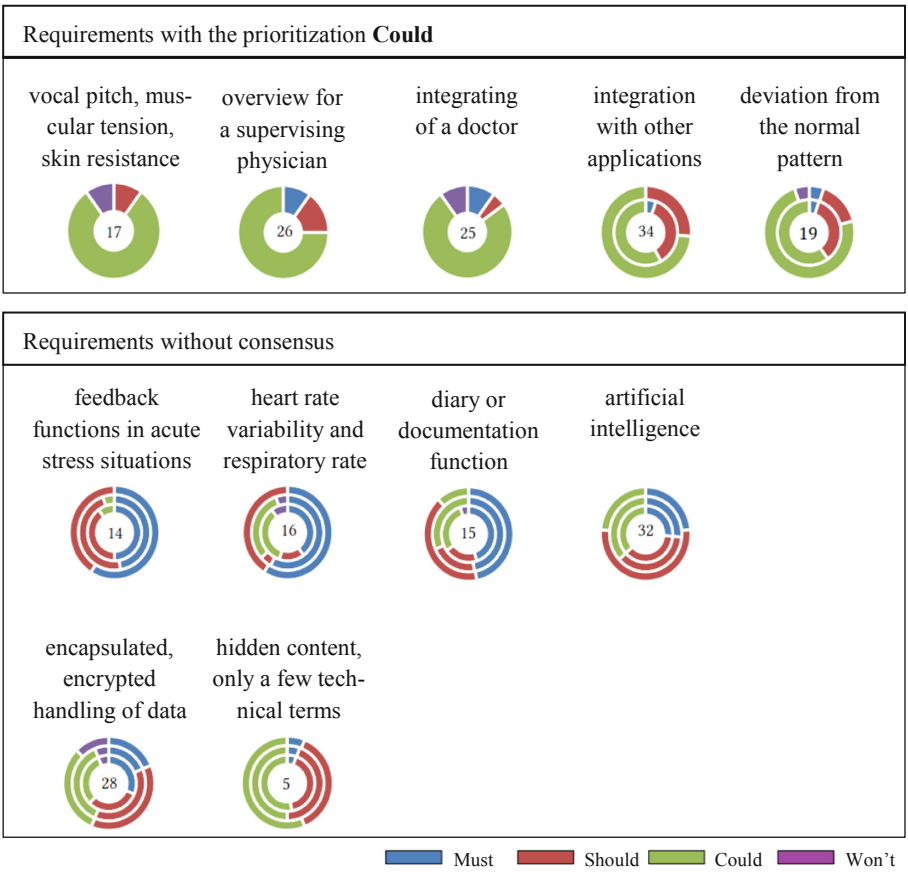

Fig. 2. (*continued*)

experts state that it is better not to present any feedback directly in the stress situation itself but with a slight delay. This minimizes the risk of aggravating the stress level due to the given notification, whereas the user can more easily reflect on the stress-inducing situation afterward.

Furthermore, there is no agreement on how important it is to use only hidden content in a playful way and with few technical terms. On the one hand, healthy people without cognitive impairments comprise the DSMS target group, and a facilitation of interventions is usually not necessary. Thus, Requirement 5 can be a low priority. Conversely, users of DSMS often have heavy workloads and little free time, which necessitates a presentation of easily understandable content, as per Requirement 5. On an average, it is prioritized by the experts with the value of 2.5.

By already agreeing in the first round on the "Should" prioritization for Requirements 9 and 12, the experts highlight that an individual tailoring of the DSMS to the user and his environment is desirable. In order to implement it, the use of artificial intelligence (Requirement 32) would basically be conceivable. However, this aspect does

not reach any consensus. The mixed responses can be attributed to the fact that artificial intelligence is a promising but new technology which is not yet mature for a wide range of applications. Expert 10 comments regarding the Requirement 32:

"Should, not must. It's very complicated with AI and it requires extremely much data to be correct."

Further annotations from the experts suggest that working with artificial intelligence has great potential and is promising for the future. Expert 13 states:

"There is no doubt that AI is the future."

In sum, the results of the Delphi study reveal that strong motivational elements (Requirement 1, 2, 4, 8) and the conditions of usage (Requirement 3, 6, 7) are crucial to successful DSMS. This gives priority to the user first and emphasizes that DSMS are being developed for healthy people, who possibly do not feel such symptoms as compared to those who are already suffering from a condition like burnout. Therefore, great attention should be paid to motivational aspects so that preventive programs, such as DSMS, are used permanently.

In addition, the three types of DSMS exercises (Requirements 21–23) are also validated and given high priority. This aligns with the previous study, in which all the interviewed participants (health insurance companies, care providers, users, and the private sector) recommended their availability. These are also strongly addressed and examined in the literature [10, 11, 20]. Furthermore, the high prioritization of goalsetting (Requirement 1, "Must" prioritization in the first round) is consistent with findings from the literature [8, 35]. On the other hand, it is notable that everyday suitability (Requirement 7) and scientific evidence of offered interventions (Requirement 24) were already given extremely high importance in the first round, but these aspects are hardly considered in the literature. This underlines that some of the requirements identified in the preliminary study and prioritized in this study remain little known to date.

Some results of this study do not align with the literature. The experts consider the integration of other digital systems or external stakeholders, such as physicians, of little importance, although research has proven that guided interventions are more effective than unguided ones [10]. This can be attributed to a poor cost-benefit ratio of guidance or to the visionary focus of the experts on successful self-management and the well-empowered user.

In conclusion, a focus on the key features of DSMS, such as providing appropriate measurement methods and interventions to reduce excessive stress and its motivating presentation, is far more important than the integration of multiple additional functions.

6 Limitations and Future Work

There were several limitations to this study. There was a risk of ambiguity and conditional statements provided with the list of requirements. The experts may have interpreted the statements differently, and this could lead to a distortion in the results. In order to minimize this risk to the greatest extent possible, explanations of several of the requirements were also provided to the experts.

Although we executed a considerable search to identify experts, it is possible that some areas of expertise, geographic regions, and disciplines were not as well represented as others. For example, the experts only came from Europe and the United States, with a special focus on Germany. Thus, the findings may have limited the generalizability to other countries and regions, where the culture, handling of stress, or the work environment are very likely to be different. Another potential limitation of the study is that we only employed a panel of experts rather than the users themselves.

The knowledge presented by the experts in this study can be accessed by users who wish to test DSMS and directly report their findings and experience. Moreover, the additional involvement of employers as a further stakeholder group would be of interest, as they are closely related to the stress-causing source of burnout.

Furthermore, it will be beneficial to run comprehensive validations to evaluate the upcoming costs and the equipment required.

7 Conclusion

This article confirms and prioritizes the given list of requirements for DSMS. After the three rounds of this Delphi study, 82% of all the requirements had reached consensus. The experts agree that the offering of psycho-educative and reflective exercises as well as those to build up anti-stress resources must be given high priority.

In addition, DSMS should provide reminder functions, goal setting features, and stress measurement methods. A clear reference to everyday life and high suitability for everyday use as well as high mobility are also vital characteristics. Some of those less important ones are, in the experts' opinion, the involvement of medical professionals, the possibility to integrate other applications into the DSMS, and the use of some specific stress measurement methods (voice analyses, skin resistance, muscular tensions). Overall, the study demonstrates that there is strong agreement on the importance of individual functions and features for successful DSMS. The results support developers in profitably selecting and using functions and characteristics for better DSMS in order to counteract excessive stress. We believe that the improvement of DSMS is an effective way to reach more working individuals with psychological interventions in order to reduce stress and help strike a healthy balance. This could be important from both public health and societal perspectives.

References

1. Möltner, H., Leve, J., Esch, T.: Burnout-Prävention und mobile Achtsamkeit: Evaluation eines appbasierten Gesundheitstrainings bei Berufstätigen (Burnout Prevention and Mobile Mindfulness: Evaluation of an App-Based Health Training Program for Employees). Gesundheitswesen (Bundesverband der Ärzte des Öffentlichen Gesundheitsdienstes (Germany)) **80**(3), 295–300 (2018)
2. Kuster, A.T., Dalsbø, T.K., Luong Thanh, B.Y., Agarwal, A., Durand-Moreau, Q.V., Kirkehei, I.: Computer-based versus in-person interventions for preventing and reducing stress in workers. Cochrane Database Syst. Rev. **8**, CD011899 (2017)

3. Ly, K.H., Asplund, K., Andersson, G.: Stress management for middle managers via an acceptance and commitment-based smartphone application. A randomized controlled trial. Internet Interv. **1**(3), 95–101 (2014)
4. Jayewardene, W.P., Lohrmann, D.K., Erbe, R.G., Torabi, M.R.: Effects of preventive online mindfulness interventions on stress and mindfulness: A meta-analysis of randomized controlled trials. Prev. Med. Rep. **5**, 150–159 (2017)
5. Pisani, A.R., et al.: Human Subjects Protection and Technology in Prevention Science: Selected Opportunities and Challenges. Prev. Sci. Official J. Soc. Prev. Res. **17**(6), 765–778 (2016)
6. Walter, U., Krugmann, C.S., Plaumann, M.: Burn-out wirksam prävenieren? Ein systematischer Review zur Effektivität individuumbezogener und kombinierter Ansätze (Preventing burnout? A systematic review of effectiveness of individual and combined approaches). Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz **55**(2), 172–182 (2012)
7. Günthner, A., Batra, A.: Stressmanagement als Burn-out-Prophylaxe (Prevention of burnout by stress management). Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz **55**(2), 183–189 (2012)
8. Billings, D.W., Cook, R.F., Hendrickson, A., Dove, D.C.: A web-based approach to managing stress and mood disorders in the workforce. J. Occup. Environ. Med. **50**(8), 960–968 (2008)
9. Ebert, D.D., Lehr, D., Heber, E., Riper, H., Cuijpers, P., Berking, M.: Internet- and mobile-based stress management for employees with adherence-focused guidance: efficacy and mechanism of change. Scand. J. Work Environ. Health **42**(5), 382–394 (2016)
10. Heber, E., et al.: The benefit of web- and computer-based interventions for stress: a systematic review and meta-analysis. J. Med. Internet Res. **19**(2), e32 (2017)
11. Jonas, B., Leuschner, F., Tossmann, P.: Efficacy of an internet-based intervention for burnout: a randomized controlled trial in the German working population. Anxiety Stress Coping **30**(2), 133–144 (2017)
12. Ahtinen, A., et al.: Mobile mental wellness training for stress management: feasibility and design implications based on a one-month field study. JMIR mHealth and uHealth **1**(2), e11 (2013)
13. Pospos, S., et al.: Web-based tools and mobile applications to mitigate burnout, depression, and suicidality among healthcare students and professionals: a systematic review. Acad. Psychiatry **42**(1), 109–120 (2018). The Journal of the American Association of Directors of Psychiatric Residency Training and the Association for Academic Psychiatry
14. Amutio, A., Martínez-Taboada, C., Delgado, L.C., Hermosilla, D., Mozaz, M.J.: Acceptability and effectiveness of a long-term educational intervention to reduce physicians' stress-related conditions. J. Continuing Educ. Health Prof. **35**(4), 255–260 (2015)
15. Allexandre, D., Bernstein, A.M., Walker, E., Hunter, J., Roizen, M.F., Morledge, T.J.: A web-based mindfulness stress management program in a corporate call center: a randomized clinical trial to evaluate the added benefit of onsite group support. J. Occup. Environ. Med. **58**(3), 254–264 (2016)
16. Lange, A., van de Ven, J.-P., Schrieken, B., Smit, M.: 'Interapy' Burn-out. Prävention und Behandlung von Burn-out über das Internet. Verhaltenstherapie **14**(3), 190–199 (2004)
17. Morledge, T.J., et al.: Feasibility of an online mindfulness program for stress management – a randomized, controlled trial. Ann. Behav. Med. Publ. Soc. Behav. Med. **46**(2), 137–148 (2013)
18. Heber, E., Lehr, D., Ebert, D.D., Berking, M., Riper, H.: Web-based and mobile stress management intervention for employees: a randomized controlled trial. J. Med. Internet Res. **18**(1), e21 (2016)
19. Hintz, S., Frazier, P.A., Meredith, L.: Evaluating an online stress management intervention for college students. J. Couns. Psychol. **62**(2), 137–147 (2015)

20. Hasson, D., Anderberg, U.M., Theorell, T., Arnetz, B.B.: Psychophysiological effects of a web-based stress management system: a prospective, randomized controlled intervention study of IT and media workers ISRCTN54254861. BMC Public Health **5**, 78 (2005)
21. Blankenhagel, K.J., Theilig, M.-M., Koch, H., Witte, A.-K., Zarnekow, R.: Challenges for preventive digital stress management systems - identifying requirements by conducting qualitative interviews. In: Proceedings of the 52nd Hawaii International Conference on System Sciences 2019, pp. 3810–3819 (2019)
22. Flick, U.: Qualitative Sozialforschung. Eine Einführung. In: 8th edn. Rororo Rowohlts Enzyklopädie, vol. 55694. Rowohlts Enzyklopädie im Rowohlt Taschenbuch Verlag, Reinbek bei Hamburg (2017)
23. Döring, N., Bortz, J.: Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften, 5th edn. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-642-41089-5
24. Mayring, P.: Qualitative Inhaltsanalyse. Grundlagen und Techniken. 12th edn. Beltz Pädagogik. Beltz, Weinheim (2015)
25. de Villiers, M.R., de Villiers, P.J.T., Kent, A.P.: The Delphi technique in health sciences education research. Med. Teach. **27**(7), 639–643 (2005)
26. Efstathiou, N., Ameen, J., Coll, A.-M.: A Delphi study to identify healthcare users' priorities for cancer care in Greece. Eur. J. Oncol. Nurs. Official J. Eur. Oncol. Nurs. Soc. **12**(4), 362–371 (2008)
27. Zelmer, J., van Hoof, K., Notarianni, M., van Mierlo, T., Schellenberg, M., Tannenbaum, C.: An assessment Framework for e-Mental health apps in canada: results of a modified Delphi process. JMIR mHealth uHealth **6**(7), e10016 (2018)
28. Keeney, S., Hasson, F., McKenna, H.P.: The Delphi Technique in Nursing and Health Research. Wiley-Blackwell, Chichester (2011)
29. Häder, M., Häder, S.: Die Delphi-Technik in den Sozialwissenschaften. VS Verlag für Sozialwissenschaften, Wiesbaden (2000)
30. Okoli, C., Pawlowski, S.D.: The Delphi method as a research tool. An example, design considerations and applications. Inf. Manag. **42**(1), 15–29 (2004)
31. Skinner, R., Nelson, R.R., Chin, W.W., Land, L.: The Delphi method research strategy in studies of information systems. CAIS **37**, 31–63 (2015)
32. Woudenberg, F.: An evaluation of Delphi. Technol. Forecast. Soc. Chang. **40**(2), 131–150 (1991)
33. Brooks, K.W.: Delphi technique. Expanding applications. North Central Assoc. Q. **53**(3), 377–385 (1979)
34. Duffield, C.: The Delphi technique A comparison of results obtained using two expert panels. Int. J. Nurs. Stud. **30**(3), 227–237 (1993)
35. Ebert, D.D., et al.: Efficacy and cost-effectiveness of minimal guided and unguided internet-based mobile supported stress-management in employees with occupational stress: a three-armed randomised controlled trial. BMC Public Health **14**, 807 (2014)