# Validation and usability study of the framework for a user needs-centered mHealth app selection

Jisan Lee [a,*], Mikyoung Angela Lee [b]

[a] Department of Nursing Science, College of Life & Health Sciences, Hoseo University, Asan, Republic of Korea
[b] Doswell Endowed Chair for Informatics and Healthcare Transformation and Professor, College of Nursing, Texas Woman's University, Dallas, TX, United States

A R T I C L E  I N F O

A B S T R A C T

*Background:* Most mHealth app users rely on an app's rankings, star ratings, or reviews, which may not reflect users' individual healthcare needs. To help healthcare providers, researchers, and users select an optimal mHealth app, the Method of App Selection based on User Needs (MASUN) 1.0[1] was developed and tested in prior research. Initial testing found the need for improvement.
*Objective:* This multiple-phase study aimed to simplify and improve MASUN 1.0, resulting in MASUN 2.0, and verify the feasibility and usability of MASUN 2.0.
*Methods:* This study was conducted in three phases: (1) modification of MASUN 1.0 to improve its importance, applicability, relevance, and clarity, in consultation with 21 experts in medical or nursing informatics; (2) validation of the draft MASUN 2.0, with 13 experts; and (3) feasibility testing of MASUN 2.0 and usability evaluation of the best app found through MASUN 2.0. Menstrual apps were used to test the framework.
*Results:* From Phases 1 and 2, MASUN 2.0, the framework for mHealth App selection, was derived with improved simplicity, usability, and applicability through a reduced number of tasks and time required. In Phase 3, after screening and scoring 2377 menstrual apps, five candidate apps were selected and evaluated by five clinical experts, five app experts, and five potential users. Finally, 194 users evaluated the usability of the app selected as the best. The best app helped users understand their health-related syndromes and patterns. Additionally, user-provided scores for impact, usefulness, and ease of use for the app were higher than for others.
*Conclusions:* This study successfully modified MASUN 1.0 into MASUN 2.0 and verified MASUN 2.0 through content validity, feasibility, and usability testing. The selected apps through MASUN 2.0 helped health consumers more easily address health discomfort. Future research should extend this work to an automated system and different medical conditions with multiple stakeholders for digital health equity.

## 1. Introduction

### 1.1. Background

With the increase in access to healthcare information via smartphones, healthcare consumers are increasingly seeking mHealth applications (apps). In fact, the mHealth market is expected to be worth approximately US $190 billion by 2025 [1], and there were 107,033 medical apps in 2021, when mHealth app downloads surpassed 3.2 billion [2,3].

Most people seeking out mHealth apps choose an app based on its ranking, star rating, or reviews in the relevant mobile app store [4].

Rankings are determined by user search terms, star ratings, reviews, update cycles, and downloads [5]. As a result, apps are usually presented in the order of popularity. Further, these stores display the score of an app's user experience based on a 5-star scale and review posts [6]. These factors influence each other, in that high ranks, star ratings, and positive reviews increase app popularity and sustainability [7]. Thus, popular apps are more likely to grow in popularity. Hence, users' attention is usually drawn to a relatively small list of highly popular apps, hindering the discovery of less popular apps that they may want to use [8].

Further, this ranking system does not inquire about users' healthcare needs and may not guide healthcare consumers to select the best mHealth app for their healthcare management [9]. In addition, the

---

* Corresponding author at: (31499) 301, RIC building, 20, Hoseo-ro 79 beon-gil, Baebang-eup, Asan-si, Chungcheongnam-do, Republic of Korea.
  *E-mail address:* saan2mari@gmail.com (J. Lee).
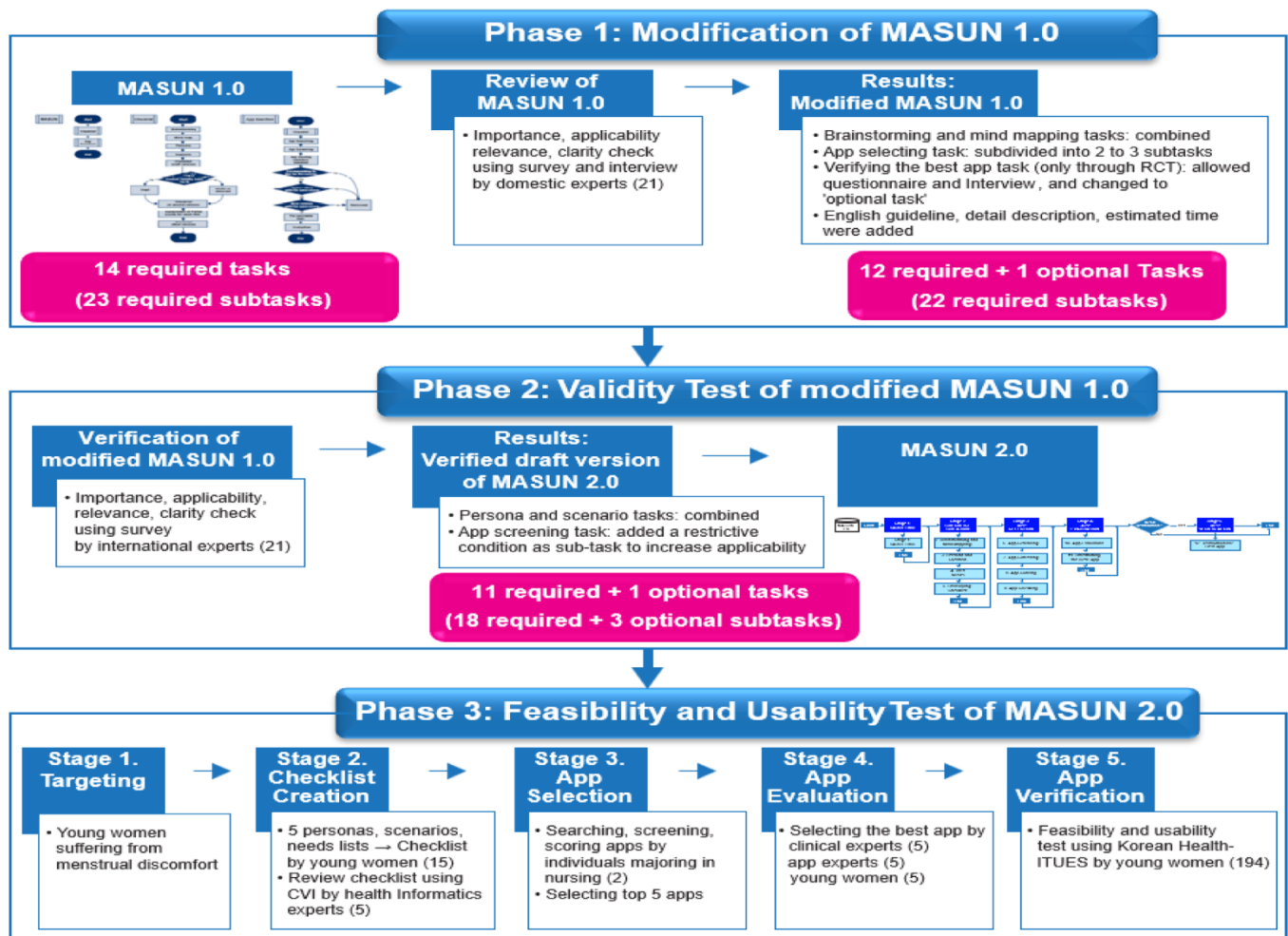  [1] MASUN: Method of App Selection based on Users' Needs.

**Fig. 1.** Study procedures for Phases 1–3.

literature emphasizes the need for apps with a user-centered design and addressing specific healthcare conditions, such as cystic fibrosis or depression [10,11].

To help users select the optimal mHealth apps for their healthcare needs, the Method of App Selection based on User Needs 1.0 (MASUN 1.0) was developed in 2018 [8]. MASUN 1.0 comprises 14 required tasks and 23 subtasks including brainstorming, mind mapping, and persona and scenario creation; all tasks and subtasks to be conducted by clinical experts, app experts, and potential users. The usefulness of MASUN 1.0 was evaluated through a randomized control trial (RCT) with young women who sought help in the management of menstrual discomfort, which is defined as common menstrual-related discomfort symptoms among women of reproductive age: primary dysmenorrhea and/or premenstrual symptom (PMS) [12]. The young women in the experimental group explored the MASUN 1.0 process to find the best app while the control group found the app based on the market ratings only. Overall, participants in the previous study were satisfied with the apps selected through MASUN 1.0, giving higher scores for app outcome expectancy, app social influence, intent to recommend, number of recording days in the menstrual app, and potential for behavioral or cognitive changes in symptom management than the control group (which used the related app with the largest number of users worldwide) [12]. Nonetheless, the clinicians and researchers described the process of MASUN 1.0 as complex by many required tasks and subtasks and their uncertain understanding of the guidelines for performing the tasks. Thus, they called for a simpler approach with clearer instructions [12].

### 1.2. Aim

This study aimed to simplify and advance MASUN 1.0 and verify the feasibility and usability of the modified version, MASUN 2.0.

### 2. Materials and methods

#### 2.1. Study design

This multiple-phase research comprised three phases. Phase 1 entailed the modification of MASUN 1.0. Phase 2 served to conduct the validity testing of the modified MASUN 1.0 and develop the draft of MASUN 2.0. Phase 3 served for the feasibility and usability testing of MASUN 2.0 by clinical experts, app experts, and potential users (Fig. 1).

#### 2.2. Sample

Convenience and snowball sampling were used for recruitment, which occurred through email and online messaging (e.g., through KakaoTalk) in South Korea and the United States. For Phases 1 and 2, the inclusion criteria were healthcare providers, researchers, and graduate students who majored in medical or nursing informatics and either had clinical or research experience in mHealth app usage or planned to use such apps. In Phase 3, we recruited a total of 214 young women who experienced the target symptom: menstrual discomfort. We also recruited five health informatics experts, five clinical experts (e.g., doctors, nurses), and five app experts (graduate students and professors

majoring in informatics). Inclusion criteria for this study were unmarried women between greater than 20 and < 39 years of age. This is because age, marital status, and delivery status could affect dysmenorrhea and PMS [13,14]. In addition, in Stage 5 of Phase 3, young women users needed previous or current App A usage experience of more than one month.

## 2.3. Data collection

### 2.3.1. Phase 1: Modification of MASUN 1.0

The MASUN 1.0 comprised five stages, 14 required tasks, and 23 subtasks (Fig. 1). In a questionnaire, all the MASUN 1.0 tasks were listed (e.g., "1. Targeting - Determine the [target disease or symptom]"), and participants were asked to rate the importance, applicability, and relevance of each task on a 4-point Likert scale (1 = not at all; 4 = very important). The higher the scale value, the higher the importance, applicability, relevance, and clarity of the task. Each rating was calculated as a content validity index (CVI), and if the value was less than or equal to 0.78, the task was considered in need of modification [15]. Furthermore, when the researchers needed additional explanations from respondents, individual semi-structured telephonic interviews were conducted using the questionnaire items.

### 2.3.2. Phase 2: Validity test of the modified MASUN 1.0 and the draft of MASUN 2.0

The modified MASUN 1.0 was derived from the results of Phase 1. In Phase 2, international experts also analyzed the four aspects of tasks identical to Phase 1 using the English version: importance, applicability, relevance, and clarity.

### 2.3.3. Phase 3: Feasibility and usability test of MASUN 2.0

The results of Phase 2 allowed us to develop MASUN 2.0. In Phase 3, the following procedures were conducted using MASUN 2.0.

In *Stage 1: Targeting*, menstrual discomfort was selected as the target symptom because MASUN 1.0 was validated for menstrual apps [12], and the market for women's health apps is expected to be worth US $4.8 billion by 2025 [16].

In *Stage 2: Checklist Creation*, 15 young women (divided into five groups) conducted group brainstorming and designed a mind map of menstrual discomfort. From the scenario, the participants prepared a checklist of the user's needs in order of priority. The CVI of the checklist was checked by five health informatics experts. This stage also comprised one open-ended question, which asked them to write a comment freely sharing their opinion on the checklist items and the respective scores.

In *Stage 3: App Selection*, the following Medical Subject Headings (MeSH) [17] and general search terms were used in both English and Korean to search for apps in Apple App Store and Google Play Store: "period," "dysmenorrhea," "PMS," "menstruation," "menstrual," "menstrual pain," and "premenstrual syndrome." Top 100 apps were retrieved. Then, the top five apps (candidate apps) were selected based on checklist score, star rating, number of reviews in the app store, whether the app was paid, and whether it existed in both Google Play Store and Apple App Store.

In *Stage 4: App Evaluation*, five clinical experts, five app experts, and five young women (potential users) suffering from menstrual discomfort and using menstrual apps were asked to evaluate the candidate apps by completing the 20-item Korean Health Information Technology Usability Evaluation Scale (ITUES) [18]. It is measured on a five-point Likert scale (1 = strongly disagree; 5 = strongly agree); the higher the score, the higher the usability of the app. In the case of potential users, in-depth interviews were conducted based on questionnaires, to understand app preference better.

In *Stage 5: App Verification*, 194 young women users completed an online questionnaire comprising the following sections: (1) general characteristics (3 items), (2) smartphone experience along with menstrual app usage and influence (11 items), (3) the Korean Health-ITUES [18], (4) dysmenorrhea (2 items), with questions on pain on the first and second days of the menstrual period, using visual analog scales [12], and (5) a PMS (11 items) [12,19]—the higher the score, the severer the symptoms. All questions were multiple-choice. Moreover, the questions on the usage and influence of the menstrual app were presented in a multiple-choice format with reference to previous studies [8,12], and the open question 'Other' was added to allow users to elaborate their answers.

After completing the questionnaire, a 2000 KRW (approximately USD 2) online coffee coupon was provided to participants.

## 2.4. Ethical considerations

This study was approved by the Institutional Review Board (CUPIRB-2018-014, CUPIRB-2019-003). Written study descriptions were provided to and informed consent was gathered from all participants.

## 2.5. Statistical analysis

All analyses were performed using IBM SPSS Statistics 24.0. Descriptive statistics were used to analyze participants' demographic characteristics and mean scores for dysmenorrhea, PMS, and the Korean Health-ITUES. An item-level CVI of 0.78 or higher provided by experts was deemed evidence of good content validity [15]. Two-sided $P$-values <0.05 were considered statistically significant.

## 3. Results

### 3.1. Phase 1: Modification of MASUN 1.0

In phase 1, MASUN 1.0 was modified and yielded the draft of MASUN 2.0. Most scores were above 0.78 (importance: 0.81–1.00; applicability: 0.76–0.95; relevance: 0.86–1.00; and clarity: 0.90–1.00). The experts were concerned about the applicability of the persona, app screening, and app scores crosscheck tasks. Comments collected in the interview are presented in Table 1. The modified MASUN 1.0 comprised 12 required tasks, 22 required subtasks, and one optional task.

### 3.2. Phase 2: Validity test of the modified MASUN 1.0 and the draft of MASUN 2.0

The survey results showed that three tasks needed modifying (importance: 0.69–1.00, applicability: 0.75–1.00, relevance: 0.85–1.00, and clarity: 0.92–1.00). The questionnaire results show that international experts thought that the subtask where two reviewers scored apps independently based on the checklist was not important, and that the persona and scenario creation and app-screening tasks were too hard to apply. After reconfirming the importance of app scoring independently by two researchers in previous research [20] and the Cochran guide [21], we decided to keep it. Moreover, the persona and scenario creation tasks were combined into one task, and detailed instructions were added. A restrictive condition that selects only the top 100 apps with more than 500 reviews by each search term was added as a subtask for applicability in the app-screening task. Regarding the open-ended question, the participants' statements are in Table 1.

Consequently, MASUN 2.0 included 11 required tasks and one optional task; the 21 subtasks included three optional subtasks (Fig. 2 and Table 2).

### 3.3. Phase 3: Feasibility and usability test of MASUN 2.0

#### 3.3.1. Stage 1: Targeting

The target symptom was menstrual discomfort, defined as primary dysmenorrhea and/or PMS.

**Table 1**
Statements in open-ended questions and interview in phases 1 and 2.

| Phase | Statements | Participants | Apply to MASUN 2.0 |
|---|---|---|---|
| Phase 1 | "It seems that MASUN's stakeholders (researchers, medical consumers, etc.) may have difficulties in performing their tasks and its applicability may be reduced. Improving MASUN's performance is important, but to complete the tasks of stakeholders, they will need to maximize their satisfaction [by eliminating unnecessary difficulties]. Also, detailed and clear descriptions in Korean/English are needed." | Domestic expert ID 2, faculty in a Human-Computer Interaction college with Ph.D. in medical informatics | When MASUN 1.0 was implemented, the candidate apps were in the top 20. Consequently, it is set to target top 30 apps, not top 100, when making an app list and scoring. The description of the Korean and English MASUN guidelines has been revised. |
| | "In the process of creating [a checklist] and evaluating [the apps] by groups of experts and consumers, I am concerned that the number and nature of reviews provided by users about the app may be detrimental to objectivity, as it is affected not only by the quality of the actual app but also by variables such as app release date (i.e., a construct not directly related to app quality)." | Domestic expert ID 7, faculty in a nursing college with Ph.D. in nursing informatics | Re-investigate the necessity of the number of reviews and star rating in the app stores |
| | "Stage: 2 Checklist Creation and 4: App Evaluation should consider and reflect the user's needs when evaluating the app. However, should the method for app validation be RCT? RCTs are not easily applied. What about evaluating usability through surveys or interviews, or eliminating them [Stages 2 and 4] altogether?" | Domestic expert ID 10, nurse informatist with Ph.D. in nursing informatics | Stage 5 is changed to optional task and includes the same survey methodology using questionnaires and interviews other than RCT. |
| | "What would be the solution when the opinions of the 4th stage expert group differ?" "I wonder how long it takes to implement each task. And how about combining the brainstorming and mind mapping steps?" | Domestic expert ID 12, faculty in a nursing college with Ph.D. in nursing who wants to do research with mHealth app. | If the evaluators of the candidate apps disagree, the researcher can choose another app from the potential user's preferred apps. At this time, if the opinions of clinical and app experts differ, it will be necessary to reach an agreement, such as through interviews. Combine the brainstorming and mind-mapping tasks into a single task Provide time estimations for completing each task |
| Phase 2 | "Most items seem to be necessary to select apps. However, it [MASUN] can combine 'creating user personas' and 'creating a scenario' into one task. In addition, the best apps [for people] may vary depending on their [women's] preferences, skills, and ability to use information. For example, some people may like apps that provide as much detail as possible, while others prefer to keep things neat and simply tell them what is important. Therefore, in addition to selecting one of the best apps, it would be good to grasp the characteristics of each app well, inform the user of the characteristics of each app, and suggest using the desired app." | International expert ID 6 with Ph.D. in nursing | Combine the creating persona and scenario tasks into a single task In a follow-up study, a system that can select apps according to individual users' preferences was planned. |
| | "Reviews are subjective; it might not be a fair way of rating an app based on number of reviews. " "I feel that seeing a doctor is still the best way to be diagnosed with a medical problem and [this] should not be replaced by an app. However, the app is a great idea to track disease, improve medical follow-up (such as recommended vaccination), compiling a person's medication list with a link to medication education." | International expert ID 2, nursing informatics graduate | Re-investigate the necessity of the number of reviews in the app store |

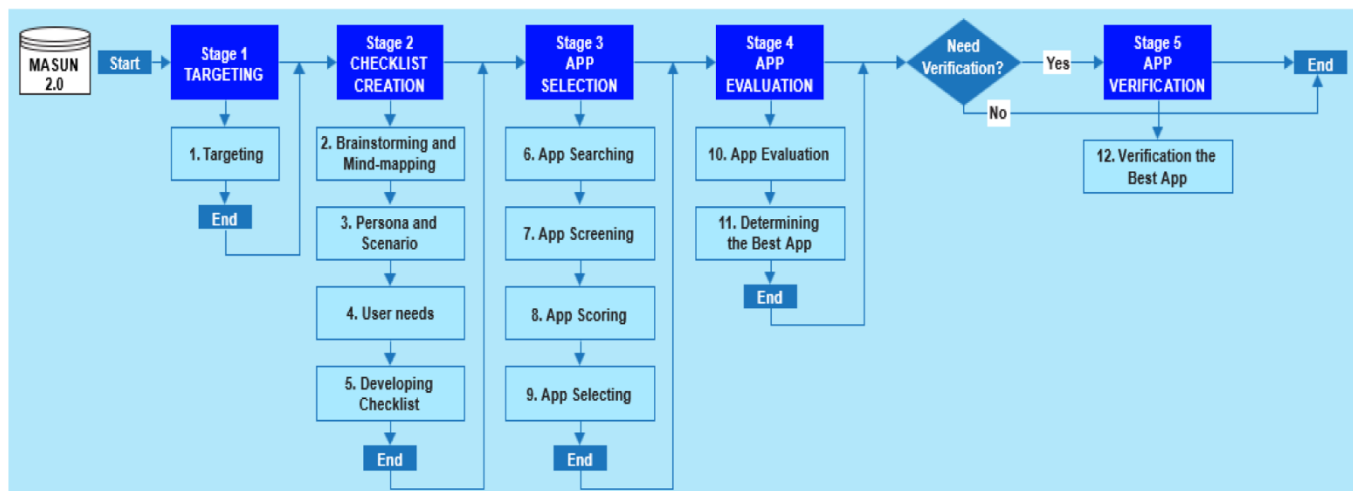*Information within square brackets serves only to enhance clarity.

**Fig. 2.** MASUN 2.0 Flowchart.

*3.3.2. Stage 2: Checklist Creation*

A mind map of menstrual discomfort was created from *Task 2: Brainstorming and Mind-Mapping*. Then, the participants created five personas and scenarios managing menstrual discomfort and wrote down a checklist of the user's needs in order of priority through *Tasks 3–5*. The checklist included 23 items (app features and requirements) and a scoring matrix for each item: ① app design, ② widget, ③ alarm, ④ separate features by sex (female or male), ⑤ app attendance point, ⑥ indication of menstrual cycle and fertile period (ovulation date), ⑦ period tracking and symptom recording, ⑧menstruation management including dysmenorrhea, ⑨ PMS management, ⑩ visualization (graphical chart), ⑪ visualize menstruation status on the calendar, ⑫ contraception management, ⑬ medication, ⑭ women's health information, ⑮ healthcare providers' consultant, women's health checkup, ⑯ information sharing, ⑰ user community menu, ⑱ anonymous user community, ⑲ information sources, ⑳ login, ㉑ unlock mode, ㉒ privacy policy, and additional features such as ㉓ easy delivery of women's items (e.g., sanitary napkin and menstrual cup). The CVI values for the scoring matrices of each item from each of the five experts were between 0.8 and 1; therefore, no items had to be modified.

*3.3.3. Stage 3: App selection*

Initially, 2,377 apps were found in the Korean and American Apple App Stores and Google Play Stores (Fig. 3). After removing duplicate apps, apps that were not suitable for the target symptoms were also removed. In Phase 2, an international expert said that *Subtask 10: Remove all unrelated apps based on the app name or description* was relatively insufficient regarding its applicability. However, it was relatively easily completed when the task was operationalized using Excel's sorting function.

Two bilingual (in Korean and English) nursing students (raters) independently scored each of the 100 apps using the checklist, after using each app for at least 10 min. For crosschecking, the score was compared; if the scores differed between the raters, they discussed its features until a consensus was reached on the score. Then, the top five apps (the candidate apps) were selected based on checklist score, star rating, the number of reviews in the app store, whether the app was paid, and whether it existed in both app stores.

*3.3.4. Stage 4: App Evaluation*

According to the app experts', clinical experts', and potential users' evaluation of the five candidate apps using the pre-defined assessment tool (Korean Health-IUUES), App E scored the highest, followed by App A. However, App E was in English, so most evaluators stated that it was not suitable for Korean women to use in *Stage 5: App Verification* and later.

Therefore, App A, which showed the second-highest scores when evaluated by clinical and app experts, was the preferred app among the five potential users. It was provided in Korean and was deemed the best app based on MASUN 2.0 (Fig. 4).

*3.3.5. Stage 5: App Verification*

Table 3 shows the general and menstrual characteristics of the 194 young, unmarried women who experienced menstrual discomfort, used App A for at least a month, and completed the questionnaire.

Table 4 shows the results of App A use. Most participants used the app to track their menstruation cycle, and it helped them understand the cycle. They wanted features to manage their menstrual experience, provide personalized information recommendations, and recommend menstrual products.

**4. Discussion**

MASUN 1.0 was developed to select reliable apps that meet users' health needs. This multiphase study successfully modified MASUN 1.0 into MASUN 2.0 and verified the usability and feasibility of MASUN 2.0 with domestic and international healthcare providers, researchers, graduate students, and potential users. Users want health apps to do many things [22]. Identifying the target users' health needs is critical for a customized mHealth app selection. MASUN 2.0 considers users' needs and opinions through its needs list, star ratings, and the number of reviews, which can predict usability and long-term usage [23–25]. Therefore, we developed personas based on typical users with menstrual discomfort. Through a time- and cost-effective persona technique, we tried to understand the user's goal and user experience, including inconvenience [26]. Specifically, the needs list derived from personas and scenarios that clearly reflected users' health needs became a criterion for screening mHealth apps, enabling the selection of apps based on user needs [27]. Thus far, persona and scenario techniques have been aimed at understanding users and deriving needs [26,27]. This study is meaningful because it lists the user's needs derived through the technology and utilizes them as conditions for finding health apps.

While considering users' needs is essential for mHealth app selection, evaluating the selected app is also important. The impact of the app should include an objective and transparent evaluation by clinical and app experts [28]. Recently, there has been a movement to promote meaningful and equitable engagement of all relevant stakeholders (e.g., health consumers, providers, payers, purchasers, policymakers, product makers, and researchers) in the development and implementation process of healthcare guidelines [29]. Accordingly, we endeavored to incorporate users' needs and the evaluations of clinical and app experts

**Table 2**
MASUN 2.0 English Guideline.

| Tasks [a] | Descriptions of Subtask | Stakeholders | Estimated Time |
|---|---|---|---|
| *Stage 1. Targeting* | | | |
| **1. Targeting** | 1) Determine the [target disease or symptom] | Researcher who participated in the current study | – |
| *Stage 2. Checklist Creation* | | | |
| **2. Brainstorming & Mind Mapping** | 2) Conducting group brainstorming sessions on the [target disease or symptom] and designing a **mind map based on the brainstorming activity** | Individual (patient, healthcare provider, or researcher) who had experienced the target disease or symptom, or had majored in health informatics (e.g., *in medical or nursing informatics*) (number of groups $\geq$ 2; number of individuals in a group $\geq$ 2) | 1 h |
| **3. Persona & Scenario Creation** | 3) **Design a persona** who uses an app to manage the [target disease and symptom; either target disease or symptom] (the number of personas per group $\geq$ 1) and **create a scenario** for each persona (the number of scenarios per group $\geq$ 1) | | 2 h |
| **4. User Needs** | 4) **List of user needs** (expected features and functions of an app: requirements) derived from each scenario of each persona | | 0.5 h |
| **5. Developing the Checklist** | 5) **Draft of the checklist** in a table format: compile all user needs (requirements) included in the scenario into the checklist and develop a scoring matrix for each item [8]. (e.g., *If the user need is "recording the pain in the app," 0 points = no need for a pain recording menu, 1 point = need for a pain recording menu with a dichotomous (Yes/No) scale, 2 points = need for a pain recording menu with a Likert scale, and 3 points = need for a pain recording menu with scale and memo or emoticon or graph)* | [b] The same individual who conducted tasks 1 or 2 can conduct this task (i.e., researcher who participated in the current study) | 3 h |
| | 6) **Modifying the checklist**: each stakeholder is involved in the content validity index (CVI) of the items and scoring matrix in the checklist using a four-point scale. For each item, the CVI is computed as the number of stakeholders who gave a score of 3 or 4 points divided by the total number of experts. Those responsible for this procedure should consider modifying or deleting items with a score of <0.78, and then develop a second version of the checklist. | [c] This task should be conducted by an individual who has not participated in any of the previous tasks. Individual (healthcare provider or researcher) who has experience with mHealth apps or majored in health informatics (e.g., *medical or nursing informatics professional; ≥3 individuals*) Healthcare providers include doctors and nurses, among others. Researchers include graduate students and faculty with a master's degree or higher. | 1 h |
| *Stage 3. App Selection* | | | |
| **6. App Searching** | 7) **Find search terms**: use all possible synonyms for [target disease and symptom; either target disease or symptom]. | [b] The same individual who conducted tasks 1 or 5 can conduct this task. Individual (healthcare provider, researcher, or research assistant) who participated in the current study or can understand this research and related languages (e.g., *undergraduate nursing students who understand the languages in which the search terms are provided; ≥2 individuals*) | 1 h |
| | 8) **App search**: search for top 30 apps in one or more app stores using the derived search terms (multiple languages as needed) and make a list (app name, developer, star rating, the number of reviews, etc.). | | 3 h for 1 search term |
| **7. App Screening** | 9) **Deduplication**: compare the lists of searched apps, compile them into a single list, and remove duplicates. | | 1 h for 100 apps |
| | 10) **Remove apps unrelated** to the [target disease and symptom; either target disease or symptom] based on the app name or description. | | 1 h for 30 apps |
| | 11) **[ Optional subtask] Select the Top 30 apps with more than 500 reviews:** When there are more than 30 related apps, check star rating and the number of reviews in the app store | | 1 h |
| **8. App Scoring** | 12) **Score the apps** using the checklist and scoring matrix after exploring each app for at least 10 min. Those exploring the apps should complete the scoring process **independently**, not in a group. | | 1 h for 5 apps |
| | 13) **Crosscheck**: compare the scores given by each rater for the **Top 30 apps**. If the total scores of the app or the scores given for a specific item differ between raters, they should review the app and discuss it to reach a consensus on the score. | | 1 h for 10 apps |

**Table 2** (*continued*)

| Tasks [a] | Descriptions of Subtask | Stakeholders | Estimated Time |
|---|---|---|---|
| **9. App Selecting** 16) **[Optional subtask]** When multiple apps receive the same score and star rating, the app with the highest number of reviews should be selected. | 14) **Select the top 5 candidate apps** based on scores. 15) **[Optional subtask]** When multiple apps receive the same score, the app with the highest star rating should be selected. | | 1 h |
| *Stage 4. App Evaluation* **10. App Evaluation** | 17–19) **Evaluate the five candidate apps** using an assessment tool (e.g., Health Information Technology Usability Evaluation Scale) and indicate the preferred app. If necessary, an interview can be conducted. | [c] **Clinical expert** experienced in the treatment and research of [target disease and symptom; either target disease or symptom] on patients/consumers (≥5 individuals); or a **clinical expert** experienced in either the treatment or research of [target disease and symptom; either target disease or symptom] on patients/consumers (≥5 individuals) [c] **App expert** experienced in developing and designing health-related apps (≥5 individuals); or an **app expert** experienced in either developing or designing health-related apps (≥5 individuals) [c] **Potential user** who has experienced the [target disease and symptom; either target disease or symptom] (≥5 individuals) | 2 h |
| **11. Determining the Best App** | 20) **Determine the best app** based on the aggregated evaluations. | [b] The same individual who conducted tasks 1 or 5 or any of 6–9 can conduct this task. Researcher who participated in the current study. | |
| *Stage 5. App Verification* **[Optional task] 12. Verifying the Best App** | 21) **Conduct the usability test** by an intervention trial (e.g., *randomized control trial*), survey, or interview. The participants use the best app selected in the previous stage. | [c] This task should be conducted by an individual who has not participated in any of the previous tasks. Person who has experienced the [target disease and symptom; either target disease or symptom] | |

[a] What is MASUN 2.0? MASUN is the Method of App Selection based on Users' Needs. MASUN 1.0 was designed and developed in 2018 [8] and was verified through a randomized controlled trial with 61 users in 2019 [12]. It was modified to MASUN 2.0 through this study in 2021, and its domestic/international applicability and feasibility were verified.

[b] Individuals who participated in the previous task **can** participate in this task.

[c] Individuals who participated in the previous task **cannot** participate in this task, so a new participant or researcher should perform this task.
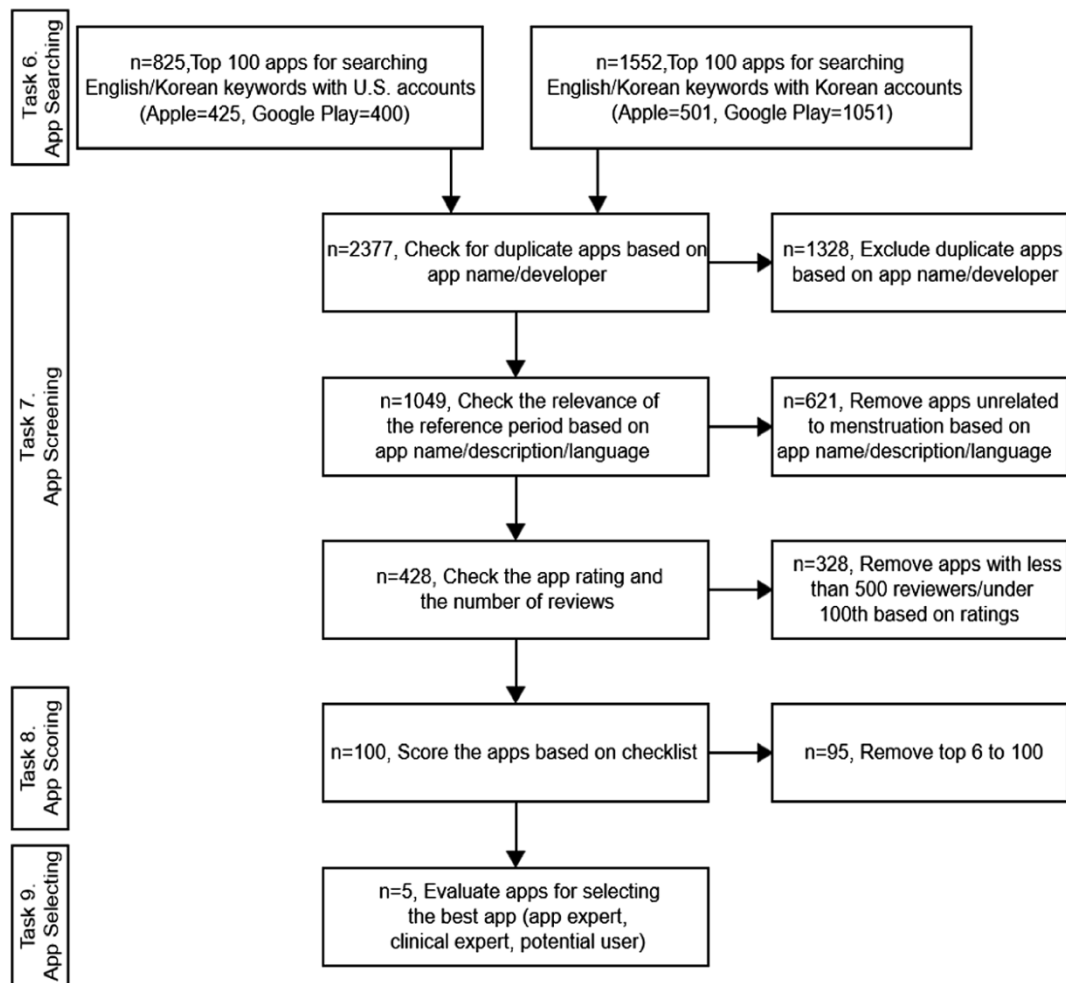
**Fig. 3.** An overview of the app selection and screening processes under *Stage 3 App Selection* of MASUN 2.0.
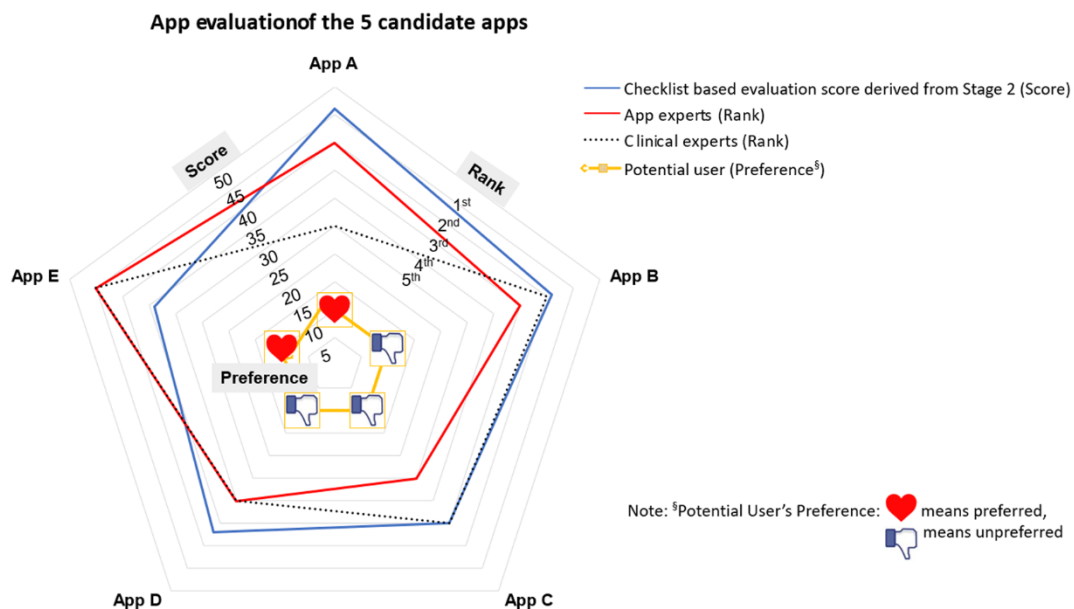


**Fig. 4.** Results of *Stage 5 App Evaluation* of MASUN 2.0 for the five candidate apps.

**Table 3**

The general and menstrual characteristics of respondents (N = 194).

| Characteristics | Options | |
|---|---|---|
| Age, mean (SD[b]) | | 22.43 (2.98) |
| Residency, n (%) | Seoul | 65 (33.5) |
| | Gyeonggi Province | 52 (26.8) |
| | Chungcheong Province | 38 (19.6) |
| | Gyeongsang Province | 34 (17.5) |
| | Others | 5 (2.6) |
| Occupation, n (%) | Related to health science | 96 (49.5) |
| | Unrelated to health science | 98 (50.5) |
| Experience of dysmenorrhea, n (%) | Yes | 168 (86.6) |
| | No | 18 (9.3) |
| | Unsure | 8 (4.1) |
| Dysmenorrhea pain score, mean (SD[b]) | Pain: first day of period | 5.40 (2.81) |
| | Pain: second day of period | 5.21 (2.89) |
| Experience of PMS[a], n (%) | Yes | 133 (68.6) |
| | No | 21 (10.8) |
| | Unsure | 40 (20.6) |
| PMS score, mean (SD[b]) | Yes | 36.45 (10.84) |
| | No | 23.00 (9.19) |
| | Unsure | 26.48 (9.53) |

[a] PMS, premenstrual syndrome.

[b] SD, standard deviation.

into MASUN 2.0 to find the most suitable app based on health needs, usability, and expert evaluation results. In the future, an updated version of MASUN 2.0 with more stakeholders may be needed to select optimal mHealth apps through inclusive, comprehensive, and equitable approaches. We believe that this will help advance mHealth app selection and the recommendation framework.

In *Stage 3: App Selection*, MASUN 2.0 uses all possible synonyms during search. In our feasibility test, we used both free-text keywords (general terms) and MeSH terms for effectiveness, compared with using only free-text [30]. The women's health app market is vast, but a relatively large number of these apps have <500 reviews [31]. Moreover, while MASUN 1.0 requires scoring the top 100, MASUN 2.0 has been updated to score the top 30 apps with more than 500 reviews. Since it took approximately an hour to score five apps, this change was helpful in MASUN 2.0 exploration. Hence, we hope that more results will be accumulated so that the tasks to be executed by MASUN users can be further simplified.

Initially, App E was identified as the best app but the translated Korean seemed unnatural. Recently, many mHealth apps offer a variety of language settings, but they are not always easy to use because some translated features are limited or unnatural. Therefore, if the language is translated more carefully when developing an mHealth app, the app can be used as an actual health care tool in more countries.

The evaluation and verification of mHealth app usability were conducted by experimental evaluations (e.g., RCTs), expert-based evaluation, narrative-based user reviews (from app stores), and observational

**Table 4**

App A usage, its influence on users, the features that users would like to see in the app, and the results of the Korean Health-ITUES (N = 194).

| Characteristics | Options | |
|---|---|---|
| Smartphone type, n (%) | iOS | 133 (68.6) |
| | Android | 61 (31.4) |
| Frequency of App A monthly use, n (%) | <1 | 8 (4.1) |
| | 1–2 | 76 (39.2) |
| | 3–5 | 75 (38.7) |
| | >5 | 35 (18.0) |
| App usage period, n (%) | <1 year | 87 (44.8) |
| | >1 year | 107 (55.2) |
| Used menus, n (%)[a] | Menstruation tracking | 188 (96.9) |
| | Ovulation cycle check | 96 (49.5) |
| | Contraception | 61 (31.4) |
| | Menstrual data backup | 45 (23.3) |
| | Symptom recording | 24 (12.4) |
| | Alarm for symptoms | 24 (12.4) |
| | Community | 21 (10.8) |
| | Collecting information | 19 (9.8) |
| | Sharing cycle | 7 (3.6) |
| | Others (e.g., intercourse record) | 1 (1.9) |
| Influence, n (%)[a] | Understanding the menstruation cycle | 187 (96.4) |
| | Understanding PMS [b] patterns | 53 (27.3) |
| | Learning about methods for dysmenorrhea relief | 7 (3.6) |
| | Learning about methods for PMS relief | 3 (1.5) |
| | Other (e.g., birth control pill check) | 4 (3.8) |
| Feature wanted to be added, n (%)[a] | Menstrual experience (dysmenorrhea/PMS[b]) | 91 (46.9) |
| | Personalized information recommendation | 84 (43.3) |
| | Recommend menstrual products | 82 (42.3) |
| | Sex education (related birth control and sexual experience) | 69 (35.6) |
| | Women's health information | 59 (30.4) |
| | Q&A about users' experiences | 45 (23.2) |
| | Sex education (related physical topics and diseases) | 40 (20.6) |
| Features wanted to be deleted, n (%)[a] | Advertising | 154 (79.4) |
| | Not applicable/unnecessary information | 57 (29.4) |
| | Related sexual experiences | 11 (5.7) |
| | Women's health information | 7 (3.6) |
| | Q&A about user's experiences | 3 (1.5) |
| Korean Health-ITUES, mean (SD) | Total score | 3.79 (0.64) |
| | Impact | 3.64 (0.82) |
| | Perceived Usefulness | 3.86 (0.71) |
| | Perceived ease of use | 4.00 (0.77) |
| | User control | 3.40 (0.78) |

[a] Items allowed for multiple answers.

[b] PMS, premenstrual syndrome.

approaches (surveys) [31,32]. RCTs, although powerful, are time- and resource-intensive and problematic when applied to commercial apps, because researchers often have no control over app features (e.g., feature changes and app updates) [33]. In MASUN, various experts' evaluations, star ratings, and the number of reviews are sufficient to validate the best app. Therefore, following previous studies and the opinions of the expert in this study, *Stage 5: App Verification* was changed to optional.

We implemented *Phase 3: Feasibility and Usability Test of MASUN 2.0* comprehensively and conservatively to verify app suitability; however, we recommend that future MASUN users follow the MASUN 2.0 Guidelines. For example, in this study, five groups, not two, were included to create a persona scenario; five experts, not three, were included to evaluate the scoring matrix of the checklist; and 100 apps, not 30, were included in the selection pool. However, such conservative approaches will no longer be required because the adequacy of the guidelines has been verified in various studies [8,12,15,34]. Nevertheless, two independent raters are still required in Stage 3, as the error rate for each reviewer is approximately 10%, and two independent raters are needed to minimize such errors [35,36]. In addition, as described in the Background section, there is a limitation that the star rating and a number of reviews cannot meet expectations of high compliance because they do not reflect the needs of the user, and thus cannot be the only criterion for health app selection. However, stars and reviews have the advantage of simplifying the experience of users and app features [37]. Therefore, they were used to filter out apps that were rejected by several users when setting the Top 5 (candidate apps) in Stage 3 of MASUN 2.0.

The most notable difference between MASUN 1.0 and MASUN 2.0 is the reduced time taken to complete the procedure because tasks have been simplified and organized. In MASUN 2.0, if the optional tasks and subtasks are not necessary, the selection of the best app may be achieved with up to three tasks and five subtasks fewer than MASUN 1.0. In particular, Stage 5, which required only time- and cost-intensive RCTs for app verification, became an optional stage through feedback from experts. Moreover, new methods such as surveys were confirmed through the feasibility and usability test, increasing the likelihood of using MASUN.

Indeed, these modifications for selecting the optimal mHealth app in MASUN 2.0 are supported by the following findings in an ad-hoc review. The best app chosen based on MASUN 1.0 ranked 20th in the Apple App Store at the time of the study [8] and entered the top 10 in February 2022. Also, the best app selected through MASUN 2.0 ranked 3rd in both Google and Apple App Stores in February 2022, while it was ranked 30th in the Apple App Store in 2019. Hence, the use of MASUN 2.0 is valuable and reliable for selecting apps that are best for users and may help estimate the sustainability of the selected app. We look forward to the analyses of these assumptions in future research.

In the usability test, participants used the menstruation-related tracking and recording feature of the best app (App A) for more than 1 month and 107 respondents (55.2%) used the app for more than 1 year, and it helped them understand their menstruation and PMS patterns better. PMS varies widely in types and degrees of symptoms, making it difficult for women to recognize PMS [38]. Therefore, it is meaningful that the app chosen via MASUN 2.0 helped women understand and recognize their PMS patterns.

To increase the availability and ease and efficient use of MASUN 2.0, we are designing an electronic MASUN (eMASUN) website. We are also developing a weighting algorithm for prioritizing user needs. The effectiveness of MASUN 2.0 using this weighting algorithm needs to be examined in future research.

There are many mHealth apps in the market, but most are not evidence-based; hence, further verification is required [24,29,39,40]. The main limitation of this study is that only unmarried women's menstrual discomfort was verified, and we thus could not capture data on *trans*-men or other gender-diverse people who menstruate. Further

exploration of MASUN 2.0 is needed to examine whether it is applicable in choosing an app that best suits the needs of people suffering from different conditions.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

*Data Statement*

The data sets generated during and/or analyzed during the current study are not publicly available as they contain information that could compromise research participant privacy/consent but are available from the corresponding author on reasonable request.

## References

[1] C. Stewart, Global Digital Health Market Size between 2015 and 2025, by Major Segment, 2022. <https://www.statista.com/statistics/387867/value-of-worldwide-digital-health-market-forecast-by-segment/> (Accessed 1 March 2022).

[2] L. Ceci, Number of mHealth apps available in the Google Play Store from 1st quarter 2015 to 1st quarter 2021, 2021. <https://www.statista.com/statistics/779919/health-apps-available-google-play-worldwide/> (Accessed 1 March 2022).

[3] Data, ai (formerly App Annie). App Annie's State of Mobile, 2021, 2013. <https://f.hubspotusercontent20.net/hubfs/8885028/App%20Annie%20-The%20State%20Of%20Mobile%202021%20.pdf (Accessed 1 March 2022).

[4] W. Peng, S. Kanthawala, S. Yuan, S.A. Hussain, A qualitative study of user perceptions of mobile health apps, BMC Public Health. 16 (2016) 1158, https://doi.org/10.1186/s12889-016-3808-0, http://www.ncbi.nlm.nih.gov/pubmed/27842533.

[5] App Store Ranking Factors: App Store vs. Google Play. <https://appradar.com/academy/aso-basics/app-store-ranking-factors> (Accessed 1 March 2022).

[6] Apple Inc, Ratings, reviews, and responses. <https://developer.apple.com/app-store/ratings-and-reviews/> (Accessed 1 March 2022).

[7] G. Lee, T.S. Raghu, Determinants of mobile apps' success: evidence from the app store market, J. Manag. Inf. Syst. 31 (2014) 133–170, https://doi.org/10.2753/MIS0742-1222310206.

[8] J. Lee, J. Kim, Method of app selection for healthcare providers based on consumer needs, Comput. Inform. Nurs. 36 (2018) 45–54, https://doi.org/10.1097/CIN.0000000000000399, http://www.ncbi.nlm.nih.gov/pubmed/29084030.

[9] E. Lee, S. Han, S.H. Jo, Consumer choice of on-demand mHealth app services: context and contents values using structural equation modeling, Int. J. Med. Inform. 97 (2017) 229–238, https://doi.org/10.1016/j.ijmedinf.2016.10.016, http://www.ncbi.nlm.nih.gov/pubmed/27919381.

[10] J. Floch, A. Zettl, L. Fricke, T. Weisser, L. Grut, T. Vilarinho, E. Stav, A. Ascolese, C. Schauber, User needs in the development of a health app ecosystem for self-management of cystic fibrosis: user-centered development approach, JMIR MHealth UHealth. 6 (2018) e113, https://doi.org/10.2196/mhealth.8236, http://www.ncbi.nlm.nih.gov/pubmed/29739742.

[11] M.R. Emerson, S. Watanabe-Galloway, D. Dinkel, S. Kotcherlakota, L. Fok, Lessons learned in selection and review of depression apps for primary care settings, J. Technol. Behav. Sci. 6 (2021) 42–53, https://doi.org/10.1007/s41347-020-00156-5.

[12] J. Lee, J. Kim, Can menstrual health apps selected based on users' needs change health-related factors? A double-blind randomized controlled trial, J. Am. Med. Inform. Assoc. 26 (2019) 655–666, https://doi.org/10.1093/jamia/ocz019, http://www.ncbi.nlm.nih.gov/pubmed/30946478.

[13] H. Ju, M. Jones, G. Mishra, The prevalence and risk factors of dysmenorrhea, Epidemiol. Rev. 36 (2014) 104–113, https://doi.org/10.1093/epirev/mxt009.

[14] A.E. Arafa, S.A. Senosy, H.K. Helmy, A.A. Mohamed, Prevalence and patterns of dysmenorrhea and premenstrual syndrome among Egyptian girls (12–25 years), Middle East Fertil. Soc. J. 23 (2018) 486–490.

[15] D.F. Polit, C.T. Beck, S.V. Owen, Is the CVI an acceptable indicator of content validity? Appraisal and recommendations, Res. Nurs. Health. 30 (2007) 459–467, https://doi.org/10.1002/nur.20199, http://www.ncbi.nlm.nih.gov/pubmed/17654487.

[16] C. Stewart, Mobile health apps market forecast in the United States from 2018 to 2025, by type (in million U.S. dollars), 2020. <https://www.statista.com/statistics/889582/mhealth-apps-market-size-forecast-by-type-in-the-us/> (Accessed 1 March 2022).

[17] N.L.o. Medicine, Medical Subject Headings, 2021. <https://www.nlm.nih.gov/mesh/meshhome.html> (Accessed 1 March 2022).

[18] J. Lee, R. Schnall, Validity and reliability of the Korean Version of the Health Information Technology Usability Evaluation Scale: psychometric evaluation, JMIR Med. Inform. 10 (2022) e28621, https://doi.org/10.2196/28621, http://www.ncbi.nlm.nih.gov/pubmed/35072630.

[19] S.S. Allen, C.M. McBride, P.L. Pirie, The shortened premenstrual assessment form, J. Reprod. Med. 36 (1991) 769–772. http://www.ncbi.nlm.nih.gov/pubmed/1765953.

[20] T. Mathes, P. Klaßen, D. Pieper, Frequency of data extraction errors and methods to increase data extraction quality: a methodological review, BMC Med. Res. Methodol. 17 (2017) 152, https://doi.org/10.1186/s12874-017-0431-4, http://www.ncbi.nlm.nih.gov/pubmed/29179685.

[21] J.P.T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M.J. Page, V.A. Welch (Eds.), Cochrane Handbook for Systematic Reviews of Interventions, Wiley, 2019.

[22] H. Duan, Z. Wang, Y. Ji, L. Ma, F. Liu, M. Chi, N. Deng, J. An, Using goal-directed design to create a mobile health app to improve patient compliance with hypertension self-management: development and deployment, JMIR MHealth UHealth. 8 (2020) e14466, https://doi.org/10.2196/14466, http://www.ncbi.nlm.nih.gov/pubmed/32130161.

[23] W. Luiz, F. Viegas, R. Alencar, F. Mourão, T. Salles, D. Carvalho, M.A. Gonçalves, L. Rocha, A feature-oriented sentiment rating for mobile app reviews, in: Proceedings of the 2018 World Wide Web Conference, ACM Press, 2018, pp. 1909–1918. <https://doi.org/10.1145/3178876.3186168>.

[24] F. Alqahtani, R. Orji, Insights from user reviews to improve mental health apps, Health Informatics J. 26 (2020) 2042–2066, https://doi.org/10.1177/1460458219896492, http://www.ncbi.nlm.nih.gov/pubmed/31920160.

[25] S. Davalbhakta, S. Advani, S. Kumar, V. Agarwal, S. Bhoyar, E. Fedirko, D.P. Misra, A. Goel, L. Gupta, V. Agarwal, A systematic review of smartphone applications available for corona virus disease 2019 (COVID19) and the assessment of their quality using the mobile application rating scale (MARS), J. Med. Syst. 44 (2020) 164, https://doi.org/10.1007/s10916-020-01633-3.

[26] K.S. Ha, An empirical test for applying the rapid persona method for the service design, J. Korea Contents Assoc. 16 (3) (2016) 210–218.

[27] R. Valaitis, J. Longaphy, J. Ploeg, G. Agarwal, D. Oliver, K. Nair, M. Kastner, E. Avilla, L. Dolovich, Health TAPESTRY: co-designing interprofessional primary care programs for older adults using the persona-scenario method, BMC Fam. Pract. 20 (2019) 122, https://doi.org/10.1186/s12875-019-1013-9.

[28] M.Á. Amor-García, R. Collado-Borrell, V. Escudero-Vilaplana, A. Melgarejo-Ortuño, A. Herranz-Alonso, J.Á.A. Arranz Arija, M. Sanjurjo-Sáez, Assessing apps for patients with genitourinary tumors using the Mobile Application Rating Scale (MARS): systematic search in app stores and content analysis, JMIR MHealth UHealth. 8 (2020) e17609. http://www.ncbi.nlm.nih.gov/pubmed/32706737, https://doi.org/10.2196/17609.

[29] J. Petkovic, A. Riddle, E.A. Akl, J. Khabsa, L. Lytvyn, P. Atwere, P. Campbell, K. Chalkidou, S.M. Chang, S. Crowe, L. Dans, F.E. Jardali, D. Ghersi, I.D. Graham, S. Grant, R. Greer-Smith, J.-M. Guise, G. Hazlewood, J. Jull, S.V. Katikireddi, E. V. Langlois, A. Lyddiatt, L. Maxwell, R. Morley, R.A. Mustafa, F. Nonino, J. P. Pardo, A. Pollock, K. Pottie, J. Riva, H. Schünemann, R. Simeon, M. Smith, A. T. Stein, A. Synnot, J. Tufte, H. White, V. Welch, T.W. Concannon, P. Tugwell, Protocol for the development of guidance for stakeholder engagement in health and healthcare guideline development and implementation, Syst. Rev. 9 (1) (2020), https://doi.org/10.1186/s13643-020-1272-5.

[30] S. Wang, H. Li, H. Scells, D. Locke, G. Zuccon, MeSH Term Suggestion for Systematic Review Literature Search. arXiv Preprint ArXiv:2112.00277, 2021. <https://doi.org/10.48550/arXiv.2112.00277>.

[31] Grand View Research, Women's Health App Market Size, Share & Trends Report Women's Health App Market Size, Share & Trends Analysis Report by Type (Fitness & Nutrition, Pregnancy Tracking & Postpartum Care, Menopause), by Region (North America, Europe, APAC, Latin America, MEA), and Segment Forecasts, 2021–2028, 2021. <https://www.grandviewresearch.com/industry-analysis/womens-health-app-market> (Accessed 1 March 2022).

[32] P. Wicks, E. Chiauzzi, 'Trust but verify'-five approaches to ensure safe medical apps, BMC Med. 13 (2015) 205, https://doi.org/10.1186/s12916-015-0451-z.

[33] D.E. Jake-Schoffman, V.J. Silfee, M.E. Waring, E.D. Boudreaux, R.S. Sadasivam, S. P. Mullen, J.L. Carey, R.B. Hayes, E.Y. Ding, G.G. Bennett, S.L. Pagoto, Methods for evaluating the content, usability, and efficacy of commercial mobile health apps, JMIR MHealth UHealth. 5 (2017) e190, https://doi.org/10.2196/mhealth.8758, http://www.ncbi.nlm.nih.gov/pubmed/29254914.

[34] M. LaNoue, A. Gentsch, A. Cunningham, G. Mills, A.M.B. Doty, J.E. Hollander, B. G. Carr, L. Loebell, G. Weingarten, K.L. Rising, Eliciting patient-important outcomes through group brainstorming: when is saturation reached? J. Patient Rep. Outcomes. 3 (2019) 9, https://doi.org/10.1186/s41687-019-0097-2.

[35] Z. Wang, T. Nayfeh, J. Tetzlaff, P. O'Blenis, M.H. Murad, S. Bencharit, Error rates of human reviewers during abstract screening in systematic reviews, PLoS ONE 15 (1) (2020) e0227742.

[36] C.R.T. Stoll, S. Izadi, S. Fowler, P. Green, J. Suls, G.A. Colditz, The value of a second reviewer for study selection in systematic reviews, Res. Synth. Methods. 10 (2019) 539–545, https://doi.org/10.1002/jrsm.1369.

[37] F. Sarro, M. Harman, Y. Jia, Y. Zhang, Customer rating reactions can be predicted purely using app features, in: 2018 IEEE 26th International Requirements Engineering Conference (RE), 2018, pp. 76–87, https://doi.org/10.1109/RE.2018.00018.

[38] T. Matsumoto, M. Egawa, T. Kimura, T. Hayashi, A potential relation between premenstrual symptoms and subjective perception of health and stress among college students: a cross-sectional study, Biopsychosoc. Med. 13 (2019) 26, https://doi.org/10.1186/s13030-019-0167-y.

[39] R.A. Karasneh, S.I. Al-Azzam, K.H. Alzoubi, S.M. Muflih, S.S. Hawamdeh, Smartphone applications for period tracking: rating and behavioral change among women users, Obstet. Gynecol. Int. 2020 (2020) 2192387, https://doi.org/10.1155/2020/2192387.

[40] L. Tucker, A.C. Villagomez, T. Krishnamurti, Comprehensively addressing postpartum maternal health: a content and image review of commercially available mobile health apps, BMC Pregn. Childbirth. 21 (2021) 311, https://doi.org/10.1186/s12884-021-03785-7.