

Usability evaluation of mHealth Apps for the elderly: A scoping review

Qiuyi Wang

Naval Medical University

Lanshu Zhou (✉ zhoulanshu@hotmail.com)

Naval Medical University

Jing Liu

Naval Medical University

Jing Tian

Naval Medical University

Xuemei Chen

Naval Medical University

Wei Zhang

Naval Medical University

He Wang

Naval Medical University

Wanqiong Zhou

Naval Medical University

Yitian Gao

Naval Medical University

Research Article

Keywords: mHealth, Applications, Usability, Elderly, Human factor design

Posted Date: May 3rd, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1515149/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Usability is a key factor affecting the acceptance of mobile health applications (mHealth apps) for the elderly, but traditional usability evaluation methods may not be suitable for use in this population because of aging barriers. The objectives of this study were to identify, explore, and summarize the current state of the literature on the usability evaluation of mHealth apps for older adults and to incorporate these methods into the appropriate evaluation stage.

Methods: Electronic searches were conducted in 10 databases. Inclusion criteria were articles focused on the usability evaluation of mHealth apps designed for older adults. The included studies were classified according to the mHealth app usability evaluation framework, and the suitability of evaluation methods for use among the elderly was analyzed.

Results: Ninety-six articles met the inclusion criteria. Research activity increased steeply after 2013 (n=92). The ratios of satisfaction, learnability, operability, and understandability measures were significantly related to the different stages of evaluation ($P \leq 0.05$). The methods used for usability evaluation were questionnaire (n=68), interview (n=36), concurrent thinking aloud (n=25), performance metrics (n=25), behavioral observation log (n=14), screen recording (n=3), eye tracking (n=1), retrospective thinking aloud (n=1), and feedback log (n=1). Thirty-two studies developed their own evaluation tool to assess unique design features for the elderly.

Conclusion: According to the three stages of the mHealth app usability evaluation framework, the critical measures and evaluation methods are inconsistent. Future research should focus on adapting usability evaluation methods to the elderly by improving traditional tools or introducing automated evaluation tools.

1. Background

Socioeconomic development in most regions has been accompanied by large reductions in fertility and equally dramatic increases in life expectancy, which have led to an increase in both the number and the proportion of older people [1]. The number of adults aged 65 or older worldwide is projected to grow rapidly, rising from 727 million in 2020 to 1.5 billion in 2050 [2]. As individuals age, their intrinsic capacities decline, and the risk of multimorbidity increases, resulting in the need for ongoing monitoring or treatment [3]. However, there is a disconnect between health-care needs and health-care utilization in older people that is caused by the high cost of medical expenses, the shortage of medical human resources, and the lack of access to health services due to functional constraints [4]. To break down the above barriers, Internet-based mobile health services have emerged. Mobile health (mHealth) refers to medical and public health services supported by mobile devices, and a software platform on such devices is called an mHealth app, with an estimation number of 325,000 in 2017 [5, 6].

In 2019, the adoption rate of smartphones by older adults aged 55 to 91 years was 40%~68% [7]. In this context, mHealth is a promising tool for promoting healthy aging through evidence-based self-management interventions that help older adults maintain functional ability and independence [8]. The effectiveness of mHealth in promoting healthy behavior and managing chronic diseases has been proven [9]. Nevertheless, the acceptance of mHealth tools by the elderly has been limited [10], with nearly half (43%) seniors over 70 quit using them during the first 14 days [11]. Usability is considered a vital factor influencing the adoption of mHealth by the elderly [12–14], which is defined as “the extent to which a system can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” [15]. Effectiveness, efficiency and satisfaction are the critical measures of usability and thus the key points of evaluation [16]. A usable mHealth app with an age-friendly interface has many benefits for the elderly, including enhancing their well-being, increasing accessibility and reducing risk of harm [17–19]. At present, a number of published standards have pointed out that usability evaluation is an indispensable step in the development of mHealth apps, and call for combing through the usability evaluation methods from empirical research [20–22].

Several reviews have been conducted to identify usability methods for mHealth apps. Zapata et al reviewed empirical usability methods for mHealth apps by analyzing 22 studies [23]. Four evaluation methods were identified, questionnaires, interviews, logs and thinking aloud. After four years, the review was updated to include 133 articles [24], suggesting that

further research should explore which methods are best suited for the target users according to their physiology and health conditions [24]. Considering the particularities of the disease, Davis et al provided a review of usability testing of mHealth interventions for HIV [25]. In summary, previous reviews have three limitations. First, usability methods suitable for older adults have not received attention. As the elderly generally face physical, cognitive, and perceptual barriers and have lower overall familiarity with technology [26], the evaluation methods they use may be different from those of other age groups. Inappropriate methods may increase the cognitive load of the elderly, leading to inaccurate assessment results. Second, the global mHealth app market size was valued at USD 40.05 billion in 2020, significantly higher than in 2015 [27]. It is very likely that the types of usability evaluation methods employed have been optimized or broadened. Thus, it is necessary to reinvestigate the methods currently being used. Third, user-centered design is a powerful framework for creating easy-to-use and satisfying mHealth apps, which can be divided into three phases: requirements assessment, development, and post release [28]. Choosing the appropriate usability methods at different phases can improve the cost-effectiveness of development. However, clear guidance for method selection has not been provided in the existing reviews. Based on previous literatures [29–32], the mHealth app usability evaluation framework (Table 1) was proposed to identify the evaluation timeline and focus of usability, including three stages.

Table 1
The classification criteria for the mHealth app usability evaluation framework

Stage	Classification criteria
Stage one Combining components	The evaluation took place in a laboratory setting and involved “user-task-system” interactions with a goal of diagnosing and fixing problems; typically based on small studies.
Stage two Integrating system into setting	The evaluation took place in the actual environment and involved “user-task-system-environment” interactions with a goal of testing the usability of the system in the absence of researcher influence and fixing usability problems; typically based on small studies.
Stage three Routine use	This stage was routine use of a complete or near-complete system in a realistic condition that could be used to determine whether the design met specific measurable performance and/or satisfaction goals or to establish a usability benchmark or make comparisons.

Based on the above analysis, there is a need to focus on the usability evaluation process of mHealth apps for the elderly and classify the evaluation approaches according to the mHealth app usability evaluation framework. The aim of this study includes (1) identifying, exploring, and summarizing the current state of the literature on the usability evaluation of mHealth apps for older adults and (2) incorporating evaluation methods into the appropriate stages. We performed a scoping review, as our aim is to map the literature on usability testing rather than seeking to answer a specific question by looking only for the best available information.

2. Methods

To complete this scoping review, the framework developed by Arksey and O’Malley was followed [33]. The reporting of this study followed the instructions suggested by the PRISMA extension for scoping reviews (Multimedia Appendix 1).

2.1 Identifying the Research Question

The following research questions were established to guide this review: (1) What is the current state of the literature that addresses usability evaluation for developing mHealth apps relevant to older adults? (2) What health conditions/diseases are being addressed by the apps that employ usability evaluation? (3) What critical measures of usability are dealt with in these studies? (4) What empirical methods and techniques are used to evaluate usability?

2.2 Searching for Relevant Studies

Ten databases shown in Fig. 1 of different disciplines were searched, such as medicine, nursing, allied health, computer and engineering sciences. The following keywords were identified and combined to address the research questions: (1) mobile devices, (2) the software used in the devices, (3) improving health as the main purpose, (4) mobile health, (5) usability as the research topic, and (6) the elderly as the target population. Chinese synonyms were used to maximize inclusion. Keywords and related subject headings were searched using Boolean operators. The search string is shown in Table 2. Finally, the reference lists of the included studies were reviewed to identify additional studies.

Table 2
Search string

Scope	String
Mobile device	(smartphone OR smart phone OR touchscreen OR mobile phone OR mobile device OR tablet OR phablet OR mobile) AND
Software	(application* OR app OR service OR operating system OR android OR ios OR windows OR mobile application*) AND
Health	(health* OR medic* OR clinic* OR care OR patient) AND
mHealth	(mobile health OR mHealth OR m-Health) AND
Usability	(usab* OR understandab* OR learnab* OR operab* OR attractiv* OR user experience OR user testing OR user-centered OR ease of use) AND
Elderly	(aging OR elderly OR older adult* OR elder*)

2.3 Selecting Relevant Studies to Include

The inclusion criteria were smart device-based mHealth studies that (1) focused on mHealth apps, (2) conducted usability evaluations, (3) set the target users of the apps as the elderly, and (4) were published from January 2000 until December 2020. Only articles published in 2000 or after were selected to accommodate the release of the first touchscreen phone marketed as a smartphone [23]. The exclusion criteria were as follows: (1) non-English and non-Chinese-language publications, (2) did not specifically describe the process of usability evaluation, (3) unable to obtain full-text versions, and (4) conference abstracts. Two authors (QW and JL) independently screened the titles and abstracts first, followed by a full-text review, and conflicts were resolved through judgment of a third author (JT) and team discussion.

2.4 Charting Data from the Selected Literature

The descriptive analytical method was used in this stage [34]. A data charting form was developed to guide the data extraction. Variables entered included standard bibliographical information (i.e., authors, year of publication, source of publication, country of origin), health condition/disease addressed by the app, critical measures of usability, the process of usability evaluation (methods, environment, duration, number of participants), and reflections on the evaluation methods. Full articles were imported as pdf files into NVivo software to extract, organize and search related data. Two authors (QW and JL) extracted the data independently, and the discrepancies were resolved by team consultation.

2.5 Collating, Summarizing and Reporting the Findings

This stage consisted of three substages: analyzing the data, reporting the results, and applying meaning to the results. For the first substage, a descriptive numerical summary was conducted to depict the characteristics and distribution of the included studies. Abductive approaches to qualitative content analysis, which combine the deductive and inductive phases, were used to analyze the data [35]. In the deductive phase, considering that the purpose of our research was to classify usability evaluation methods based on the development stages of mHealth apps and to recommend adopting a theoretical framework to collate and summarize the extracted data systematically [33], the three stages of the mHealth app usability evaluation framework were used as the theoretical categories (Table 1). The critical measures and evaluation methods of

usability were classified in the appropriate theoretical categories, and the frequency of each variable was counted. In the inductive phase, the text in the variable “reflecting on the evaluation methods” was read several times to extract the statement of the advantages and disadvantages of the methods. Then, these statements were condensed and abstracted to interpret whether each method was appropriate for use among the elderly. Finally, we identified possible gaps in the current studies and suggested evaluation methods that are suitable for the elderly.

3. Results

3.1 Search and Screening Results

The initial search obtained 1386 articles. After removing duplicates, reviewing the title, abstract and full text, 87 articles were selected. Nine more articles found through the reference list reviews were accepted. Finally, a total of 96 articles were included in this review. The flow diagram of the search procedure is presented in Fig. 1.

3.2 Characteristics of Source Documents

Figure 2 shows the number of articles published per year and the types of journals. The articles were published between 2010 and 2020, with only 4 articles before 2014 [36–39], after which the growth rate increased and reached a peaked of 27 articles in 2020. Health informatics journals were the main publication channel, accounting for 42% (n = 40) of the selected articles. Of the 96 studies included (Fig. 3), 41 were from Europe, 30 from America, 21 from Asia, and 4 from Australia. According to the mHealth app usability evaluation framework, the distribution of articles under 3 stages was combining the components (n = 61, 63.6%), integrating the system into the setting (n = 30, 31.4%), and routine use (n = 19, 19.8%). It is worth noting that the assessment process of 13 studies (13.5%) involved multiple stages. Slightly less than one-third (n = 29, 30.2%) of the studies reported the iterative design-evaluation process of mHealth applications by involving end users and stakeholders. Multimedia Appendix 2 provides an overview of the articles included in the scoping review.

3.3 Functions of the mHealth Application

The function of mHealth apps in the selected studies can be divided into four categories: wellness management (n = 40), disease management (n = 33), health-care services (n = 19), and social contact (n = 4). Nearly half of the mHealth apps were used to improve the general wellness of older adults rather than focusing on specific diseases, which contained a variety of solutions, including fall prevention [40, 41], fitness [42, 43], lifestyle modification [44, 45], medication adherence [46, 47], health monitoring [39, 48], nutrition [49, 50], and cognitive stimulation [51, 52]. Approximately one-third of the mHealth apps focused on the management of chronic diseases in the elderly such as diabetes [53, 54], mental illness [55, 56], and cardiovascular disease [57, 58]. In order to achieve the continuous care, mHealth apps have been a useful tool for delivering health-care services and empowering users to manage their health during care transitions [59], including postoperative care [60], stroke rehabilitation therapy [61], and medical advisory services [62, 63]. Four mHealth apps aimed to reduce social isolation and loneliness in older adults by encouraging social participation and strengthening ties with family members [64–67]. In addition, the target users of the mHealth apps in 78 (81.4%) articles were the elderly (aged 50/55/60/65 years or older), while others were aimed mostly at people with chronic diseases and were tested to see whether these apps suitable for use by older people. The complete range of functions, health conditions and target users can be found in Multimedia Appendix 2.

3.5 Critical Measures of Usability Evaluation for mHealth Applications

Following the usability definitions of ISO 9241-11, ISO 25010, and Nielsen, nine critical measures of usability evaluation were extracted from the selected articles: effectiveness, efficiency, satisfaction, learnability, memorability, errors, attractiveness, operability, and understandability [15, 68, 69]. It is worth noting that effectiveness, efficiency, and satisfaction focus on the impact on users when they interact with the system, while the others concern the characteristics of the system and whether they can compensate for the decline of intrinsic capacity in the elderly. As shown in Table 3, the two most frequently evaluated measures are satisfaction and learnability, consistent with the dimensions of the Systems Usability Scale (SUS)

[70], which was applied in 40 papers. The aspects of usability that were considered least often in the articles reviewed were errors and memorability. The assessment ratios of some critical measures were significantly related to the different stages of evaluation, as more of the articles focused on satisfaction and learnability in the third stage than in the first and second stages, and the articles on the third stage paid less attention to operability and understandability than those on the first two stages.

Table 3
The critical measures of usability evaluation for the mHealth application

Critical measures n (%)	Definition	Three stages of evaluation n (%) a			
		1 (n = 61)	2 (n = 30)	3 (n = 19)	P value b
Satisfaction 74 (77.1)	The extent to which the user's physical, cognitive and emotional responses that result from the use of an app meet the user's needs and expectations and can be expressed as interest in the app, willingness to continue using it, and initiative to share it.	43 (70.5)	25 (83.3)	19 (100)	.018
Learnability 60 (62.5)	The app should be easy to learn by the class of users, which can be reflected in the introduction/instruction documents helping users to reach a reasonable level of usage proficiency within a short time.	34 (55.7)	22 (73.3)	16 (84.2)	.04
Operability 50 (52.1)	The app should be easy to operate and control, which can be expressed as navigable and manipulable on the touchscreen to address the decline of cognitive ability, dexterity and muscle control in the elderly.	36 (59.0)	19 (63.3)	5 (26.3)	.02
Understandability 40 (41.7)	The interaction information of the app should be easy to understand, which can be embodied in the clarity of the provided explanations and the graphical interface to compensate for the cognitive decline of the elderly.	33 (54.1)	9 (30.0)	4 (21.1)	.01
Attractiveness 38 (39.6)	The interface of the app should enable pleasing and satisfying interaction for the user, for example, in terms of color use and graphic design, to meet the aesthetic needs of the elderly and accommodate their age-related perceptual resources.	29 (47.4)	12 (40.0)	4 (21.1)	.12
Efficiency 33 (34.4)	The extent to which external resources, including time, human effort, money, and materials, are consumed when achieving goals by using the app.	24 (39.3)	9 (30.0)	6 (31.6)	.63
Effectiveness 26 (27.1)	The extent to which actual outcomes match intended outcomes and can be measured by the accuracy and completeness with which users achieve specified goals using the app.	21 (34.4)	8 (26.7)	4 (21.1)	.48
Errors 22 (22.9)	The app should have a low error rate and protect users against making errors, for example, providing error messages or help documentation to tell users how to fix problems.	18 (29.5)	4 (13.3)	4 (21.1)	.22
Memorability 13 (13.5)	The operational flow of the app should be easy to remember, which can be embodied by reducing the demand on working memory through supporting recognition rather than recall.	11 (18.0)	4 (13.3)	3 (15.8)	.94
a First stage: combining components, second stage: integrating the system into the setting, third stage: routine use.					
b Chi-Squared Test were conducted to reflect the statistical significance of the intergroup difference.					

3.6 Empirical Methods of Usability Evaluation for mHealth Applications

Usability evaluation approaches can be classified in two categories: usability inspection and usability testing. Usability inspection is a general name for a set of methods that are all based on having experienced practitioners inspect the system using the predetermined principles with the aim of identifying usability problems [71]. In contrast, usability testing involves observing and recording the objective performance and subjective opinions of the target users when interacting with the product in order to diagnose usability issues or establish benchmarks [72].

Usability inspection methods

Fifteen articles used usability inspection methods to assess mHealth applications, which included two approaches: heuristic evaluation (n = 14) and cognitive walkthrough (n = 2).

The heuristic evaluation method requires one or more reviewers to compare the app to a list of principles that must be taken into account when designing and identify where the app does not follow those principles [73]. In the 14 heuristic evaluation articles, the evaluators usually had different research backgrounds, such as human-computer interaction, gerontology, and specific disease areas, so that a multidisciplinary perspective could be obtained [53, 64]. The number of evaluators was in the range of 2 to 8, which generally referred to the suggestion by Nielsen that 'three to five evaluators can identify 85% of the usability problems' [68]. The heuristics could be divided into two types: generic and specific. Six studies used Nielsen's ten principles, which are the most utilized generic heuristics [32, 39, 68]. However, the traditional generic heuristics were not created for small touchscreen devices, which were the main type of app carrier, and did not consider design features that were appropriate for older adults to address their age-related functional decline in terms of perception, cognition, and movement [73]. To ensure that usability issues in these specific domains were not overlooked, the remaining eight studies extended the generic heuristics by adding usability requirements specific to the elderly, such as dexterity, navigation, and visual design, and finally established new heuristic checklists to evaluate the apps targeting older adults [53, 64]. Nevertheless, there was a lack of reliability analysis and expert validation for these tools except for a checklist developed by Silva [74].

Cognitive walkthrough involves one or more evaluators working through a series of tasks using the apps and describing their thought process while doing so as if they are a first-time user [75]. The focus of this method is on understanding the app's learnability for new users [30]. The evaluators in these two studies were usability practitioners and health-care professionals [76, 77]. Before the assessment, the researchers prepared the users' personals and the task lists [77]. During the walkthrough, the evaluators were encouraged to think aloud, and their performance was recorded by usability metrics, such as task duration and completion rate [76].

Usability testing methods

Almost 93% (89/96) of the studies used usability testing to evaluate mobile applications. Test participants were the target users of the apps, and they were all elderly. Some studies (n = 52) investigated the experiences of evaluators with mobile devices or their level of eHealth literacy to obtain the testing results for experts, intermediates, and novices [40, 46, 57]. The number of participants varied according to the stage and purpose of the evaluation. The average sample sizes of the first two stages were 22.8 (ranging from 2 to 189) and 15.2 (ranging from 3 to 50), respectively, with the purpose of identifying usability problems in the laboratory or real-life environment. Most of the above studies referred to Nielsen's recommendations, which can come close to the maximum benefit-cost ratio, that is, testing three to five subjects, modifying the application, and then retesting three to five new subjects iteratively until no new major problems are identified [78]. Some studies determined the sample sizes according to the type of study design, including RCT and qualitative research [56, 79, 80]. In stage three, usability testing was usually part of a feasibility or pilot study, and the sample size was therefore based on these design types, with an average of 60.1 (ranging from 8 to 450) [58, 81, 82].

During usability testing, the objective performance and subjective opinions of the participants were collected with the corresponding data collection methods. Thirty-four studies presented objective performance data that came from observation of operational behavior, body movements and facial expressions and could be collected by performance metrics, behavioral observation logs, screen recordings, and eye tracking [46, 55, 76, 83]. Eighty-five studies gathered the subjective opinions of the participants, which involved the users' experience with the app and their design preferences for each part of the interface and could be investigated by means of concurrent thinking aloud, retrospective thinking aloud, questionnaires, interviews, and feedback logs [36, 40, 51, 57, 84]. The details and descriptive statistics of each data collection method are presented in Table 4.

The most frequently used collection method was questionnaires (n = 68). Of the studies, 51 used well-validated usability questionnaires, which were flexible enough to assess a wide range of technology interfaces. Frequently used usability questionnaires were the SUS (n = 40), the NPS (n = 4) and the NASA-TLX (n = 3). However, considering the lack of specificity of the standardized tool, self-designed questionnaires that lacked a reliable psychometric analysis were used in 24 studies to assess the unique features of the apps, including navigation, interface layout, and font size [44, 56, 85]. A combination of these two types of questionnaires was employed in 8 studies [56, 64, 86].

Table 4
Data collection methods for usability testing

Data collection method n (%)	Description	Metrics/Tools	Comments
Performance metrics 25 (26.1))	Collecting quantifiable measurements of participants' actions during the test to understand the impacts of usability issues, usually focusing on effectiveness and efficiency.	Effectiveness: number of errors, number of tasks that can be completed successfully; efficiency: task duration, number of times asking for assistance or hints, time spent recovering from errors.	These quantitative indicators can be compared in young adults and seniors to reflect differences in performance. [79, 83]
Behavior observation log 14 (14.6)	Observing and recording the participant's mood and body gestures during the test.	Sometimes, the observation is structured and based on predefined classifications of user behavior, such as delay or pause of > 5 s in locating the answer button. [42]	This method is often used in conjunction with thinking aloud and performance metrics to improve triangulation. [32]
Screen recording 3 (3.1)	Capturing the touches and actions performed on the mobile device.	Screen recording software and video coding software (Behavioral Observation Research Interactive Software)	—
Eye tracking 1 (1.1)	Monitoring and recording the visual activity of the participants by tracing pupil movement within the eye.	Fixations: the number of views of the area of interest; Saccade: the number of repeated visits to the specific area.[52]	Because of the drooping eyelids of the elderly, the eye tracker may not scan their pupil accurately.
Concurrent thinking aloud 25 (26.1)	Encouraging the participants to continuously verbalize their ideas, beliefs, expectations, doubts, and discoveries while performing tasks in order to understand their thoughts as they interact with the app.	—	This method relies heavily on the cognitive capacities of participants, whereas these capacities decline with age; thus, it may cause reporter bias. [85]
Retrospective thinking aloud 1 (1.1)	Asking the participants to view the recording of their actions and verbalize their thoughts about the tasks and the difficulties they encountered in completing the tasks.	—	1. This method will increase the overall length of the evaluation and may cause the elderly to lose focus. [87] 2. This method will not increase the cognitive load of the elderly compared with concurrent thinking aloud. [85]

a SUS: System Usability Scale [88], USE: Usefulness Satisfaction and Ease of Use questionnaire [89], UEQ: User Experience Questionnaire [90], ASQ: After Scenario Questionnaire [91], NASA-TLX: National Aeronautics and Space Administration task load index [92], NPS: Net Promoter Score [81], Health-ITUES: Health Information Technology Usability Evaluation Scale [93], QUIS: Questionnaire for User Interaction Satisfaction [94], PSSUQ: Post-Study System Usability Questionnaire [95], ICF-US: International Classification of Functioning based Usability Scale [96], MARS: Mobile Application Rating Scale [97]

Data collection method n (%)	Description	Metrics/Tools	Comments
Questionnaire 68 (70.8)	Gathering the participants' opinions about, preferences for and satisfaction with the user interface on a predefined scale after they completed the tasks.	Validated questionnaires ^a : SUS, USE, UEQ, ASQ, NASA-TLX, NPS, Health-ITUES, QUIS, PSSUQ, ICF-US, MARS, Ruland's eight-item adaptation of Davis' ease-of-use survey, self-made questionnaires according to the unique features of a specific app.	<p>1. A larger sample size can be investigated by this method. [81]</p> <p>2. Some items have to be answered by an expert rather than the elderly because they are either beyond the scope of the test or based on experiencing rare occurrences. [87]</p> <p>3. To prevent the response burden of the elderly and</p> <p>improve the understandability of the questionnaire, some items are removed or the language is modified. [42, 82, 84]</p>
Interview 36 (37.5)	Collecting the data in the form of face-to-face oral conversations with the participants, including individual interviews and focus group interviews.	The interview outline: opinions on unique features, product satisfaction, and difficulties encountered during the test as well as suggestions for improvement.	<p>1. This method can obtain more new insights from the participants.</p> <p>2. This method is often combined with a questionnaire to collect the explanations of answers to the questionnaire.</p>

^a SUS: System Usability Scale [88], USE: Usefulness Satisfaction and Ease of Use questionnaire [89], UEQ: User Experience Questionnaire [90], ASQ: After Scenario Questionnaire [91], NASA-TLX: National Aeronautics and Space Administration task load index [92], NPS: Net Promoter Score [81], Health-ITUES: Health Information Technology Usability Evaluation Scale [93], QUIS: Questionnaire for User Interaction Satisfaction [94], PSSUQ: Post-Study System Usability Questionnaire [95], ICF-US: International Classification of Functioning based Usability Scale [96], MARS: Mobile Application Rating Scale [97]

Data collection method n (%)	Description	Metrics/Tools	Comments
Feedback log 1 (1.1)	Asking the participants to record their experiences on a provided form when using the app.	–	This method is suitable for long-term usability testing, as it can record the participant's experience dynamically. [51]
a SUS: System Usability Scale [88], USE: Usefulness Satisfaction and Ease of Use questionnaire [89], UEQ: User Experience Questionnaire [90], ASQ: After Scenario Questionnaire [91], NASA-TLX: National Aeronautics and Space Administration task load index [92], NPS: Net Promoter Score [81], Health-ITUES: Health Information Technology Usability Evaluation Scale [93], QUIS: Questionnaire for User Interaction Satisfaction [94], PSSUQ: Post-Study System Usability Questionnaire [95], ICF-US: International Classification of Functioning based Usability Scale [96], MARS: Mobile Application Rating Scale [97]			

The intersection of these methods is presented in Fig. 4. Seven studies conducted both usability inspection and usability testing. Thirty studies analyzed the results of testing based on both the objective performance and subjective perceptions. Figure 5 demonstrates the distribution of the three types of evaluation methods in each stage of the mHealth app usability evaluation framework. In the three stages, most of the studies captured the subjective opinions during or after the user testing process, which was most prominent in the “routine use” stage (90.5%). The objective performance of the users was also collected at all stages, which accounted for the highest proportion in the “combining components” stage (29.3%). The usability inspection conducted by the experts was applied only in the first stage (16.3%). Table 5 illustrates the statistical description of each evaluation approach in the three stages.

Table 5
Usability evaluation approaches in three stages of the mHealth app usability evaluation framework

Evaluation approach	Three stages of evaluation n (%)		
	1 (n = 61)	2 (n = 30)	3 n = 19)
Heuristic evaluation	14 (23.0)	0 (0)	0 (0)
Cognitive walkthrough	2 (3.3)	0 (0)	0 (0)
Performance metrics	19 (31.1)	5 (16.7)	2 (10.5)
Behavioral observation log	10 (16.4)	5 (16.7)	0 (0)
Screen recording	3 (4.9)	0 (0)	0 (0)
Eye tracking	1 (1.6)	0 (0)	0 (0)
Concurrent thinking aloud	22 (36.1)	4 (13.3)	0 (0)
Retrospective thinking aloud	1 (1.6)	0 (0)	0 (0)
Questionnaire	34 (55.7)	22 (73.3)	18 (94.7)
Interview	22 (36.1)	16 (53.3)	3 (15.8)
Feedback log	0 (0)	1 (3.3)	0 (0)

4. Discussion

4.1 Principal Findings

This review identified 9 usability critical measures and 11 unique methods of usability evaluation, and analyzed their distribution in the mHealth app usability evaluation framework. The results can assist researchers in the field of mHealth for the elderly in identifying the appropriate critical measures and choosing evaluation methods that are suitable for each usability assessment stage in the life cycle of development.

Emerging trends in mHealth apps to support wellness and disease management for the elderly

Overall, usability evaluation research on mHealth for the elderly has been on the rise, with a noticeable increase in 2016, and the number of articles published in 2020 was higher than that between 2010 and 2016. However, the growth rate of usability studies is far lower than the increasing number of mHealth apps. The total global mHealth market is predicted to reach nearly USD 100 billion in 2021, which would be a fivefold increase from approximately 21 billion dollars in 2016. In addition, 68% of health-care organizations in Europe reported that they were targeting elderly people for telehealth solutions. There may be two reasons for this unequal increase. First, researchers may not realize the importance of improving the usability of mHealth apps to help the elderly overcome the digital divide [26]. Second, commercial companies developing mHealth apps are reluctant to expose usability problems to the public because of the risk of losing competitive [98]. In terms of app functions, wellness management and disease management have become the main types, which is consistent with recommendations for healthy aging, suggesting the prevention strategies according to dynamic changes in the intrinsic abilities of the elderly [99]

Stage one: Combining components

Approximately 64% of the studies evaluated the usability of mHealth apps at stage one, which means that most of the digital health technologies for the elderly were still in development and needed to be optimized iteratively in a controlled environment. The critical measures chosen in this phase tended to evaluate the design attributes of the system, such as understandability, operability, and attractiveness. The reason for this choice may derive from the primary purpose of this stage, which focuses on identifying usability problems rather than collecting users' perceived ease of use or satisfaction [23, 29]. Additionally, usability inspection methods were used only in this stage. Some researchers pointed out that this type of approach should be used in the early stage of development because it is important not to expose a prototype with potential ergonomic quality control and safety problems to a vulnerable user group, such as older adults, until it has been fully inspected by experts [73, 76, 100].

Stage two: Integrating the system into the setting

Even if an mHealth app is usable in a laboratory setting, implementation in a real environment may have different results. Therefore, stage two was carried out in realistic situations to evaluate the usability under the influence of uncontrolled environmental variables. Approximately 30% of the studies involved stage two, and eight were conducted on the basis of the optimized results in stage one. In terms of the critical measures, more research focused on the user's subjective feelings; for example, 83.3% assessed user satisfaction in stage two and only 70.5% in stage one. The operation of the apps by the elderly was also highlighted in this stage. Age-related cognitive changes, including processing speed, executive function, and visuomotor skills, may negatively influence interactions with apps [101]. Recent design guidelines for mobile phones suggested that improving the operability of the interface, such as a simple navigation structure, could help minimize users' cognitive load [102]. In terms of evaluation methods, most studies used questionnaires and/or interviews to collect users' subjective opinion, and only 20% collect objective performance data. This phenomenon may be caused by the function of mHealth apps, most of which require the elderly to use them for a period of time for self-management. However, it is unrealistic and inconvenient for researchers to observe usage performance over a long period; thus, collecting perceptions after self-exploration is a viable evaluation method.

Stage three: Routine use

After the first two stages, researchers used complete mHealth apps to conduct pilot or feasibility studies among the target population, and the usability evaluation was part of them [103]. Perceptions of satisfaction and learnability were most often evaluated, probably because almost 60% of the studies at stage three used the SUS, including two dimensions: satisfaction and learnability. In the 96 articles, there was no research to establish a usability benchmark for an app. This may be due to the large sample size required for this type of study and is usually conducted by commercial companies through market research [104].

4.2 Gaps and Potential for Future Research

The use of multiple usability evaluation methods

Several design guidelines state that a usability evaluation should include both inspection and testing methods, and inspection should be carried out before testing [30, 68, 105]. However, only two studies met the above recommendations. There are two reasons for using multiple evaluation methods. First, usability inspection methods do not have the problem of the participants in usability testing possibly not representing the pronounced heterogeneity of the target users [53]. Second, the evaluators in usability inspection are experts, thereby limiting the potential of the assessment results to provide the views of the elderly who are the end users of the app [106].

With regard to usability testing, collecting only the subjective experience of users is inadequate for identifying usability problems accurately and comprehensively [107]. However, among the 89 articles involving usability testing, 37% (n = 33) employed one evaluation method, and questionnaires were chosen in 23 of them. Specific reasons for using multiple methods to collect both subjective and objective data may be as follows. First, varying results may be obtained from different evaluation methods. One study by Richard and colleagues conducted a questionnaire survey (ASQ and NASA-TLX) from elderly users to evaluate a fall detection app [76]. The ASQ scores indicated that the users were satisfied with the product, while the NASA-TLXA and objective metrics results suggested that the app was creating a large mental burden for the users. The possible reason for these conflicting results was that the users judged the app as easy and satisfactory because they completed the task successfully without considering the difficulties encountered and the time spent [76]. Second, the advantages and disadvantages of each method can supplement each other. Observational performance data collect objective behavioral characteristics of users, which cannot explain the internal mechanism of such behavior [30]. This disadvantage can be solved by analyzing user experience and preferences, which identify the cognitive process during interaction with the app [108]. Additionally, the subjective opinion data are self-reported and often affected by acquiescence bias, social desirability bias, and recency bias, which leads to the underestimation of results [109]. If objective evaluation methods are also used in the test, these biases may be balanced [110]. However, using multiple evaluation methods may increase the length of testing, ultimately adding to the test burden of the elderly [32]. Thus, researchers should use the appropriate number of evaluation methods to collect subjective and objective data according to the stage of assessment, testing goals, and the workload that the participants can accept.

Adapting usability evaluation methods to the elderly

In the context of mHealth apps for the elderly, it is necessary to adjust the standardized usability evaluation methods to accommodate the end users' abilities. Standardized usability evaluation tools, such as Nielsen's heuristics and the SUS, usually overlook specific usability issues to compensate for the decline in cognition, perception, and mobility among the elderly [100]. Thirty-two articles in this review developed their own assessment tools, of which 8 were heuristic checklists and 24 were questionnaires. However, these tools still need rigorous psychometric analysis [64].

In usability testing with older adults, researchers should choose the appropriate data collection methods according to their physiological characteristics [32]. For example, the concurrent think-aloud method requires too much attention from elderly participants with cognitive limitations, resulting in reporter bias and task execution failed [85]; thus, one study used the retrospective think-aloud method to enable the participants to explain their behavior after completing the tasks [84]. Automated usability evaluation (AUE) methods are a promising area of usability research and can improve the accuracy and

efficiency of the test; thus, they may be suitable for the elderly because of the shorter timeline, preventing participants from losing focus [111, 112]. In this review, 3 papers employed an automatic capture method (screen recording and eye tracking) [76, 83, 113], and one paper used the automatic analysis method (natural language processing) [114]. In some studies, the language of the original scales is modified to match the understandability of the elderly and avoid increasing the response burden, for example, by removing a double negative from an item in the SUS or changing “cumbersome” in the SUS to “awkward” [84, 115].

Deciding the sample size of usability evaluation

Our review found that in the first two stages of the usability evaluation framework, the articles focused on detecting usability problems, and the sample size generally referred to the suggestions by Nielsen [68]. However, if the products under investigation have many problems available for discovery with probabilities of occurrence that are markedly different from the 0.31 proposed by Nielsen, then there is no guarantee that observing five participants will lead to the discovery of 85% of the problems [98]. Some researchers have suggested using complex alternative models instead of the simple binomial model to calculate the sample size [116]. However, the feasibility of such a model needs to be verified.

4.3 Study Limitations

This study may have some threats to its validity. (1) Conclusion validity: Relevant research questions may have been overlooked. Considering that this review focuses mainly on the evaluation methods, the results of the usability assessment were not summarized. In future studies, the severity of usability problems in each study can be classified and rated through the user action framework (UAF) and Nielsen’s severity rating [117]. (2) Construct validity: Although the PICO criteria were used to guide the search strategies, we did not include the gray literature or literature other than Chinese and English.

Given the nature of the scoping review, this study did not synthesize evidence to determine the effectiveness of usability evaluation methods. Instead, it captured the diversity of the available literature with its varied objectives, critical measures, populations, and methods. Consequently, this study was primarily exploratory and suggestive of future research directions.

5. Conclusions

This scoping review provides a descriptive map of the literature on the methods used for usability evaluation of mHealth apps for the elderly. With the widespread popularity of mHealth applications for the elderly, the number of articles evaluating the usability of these techniques has grown rapidly in the past five years. mHealth apps are often used as an auxiliary means of self-management to help the elderly manage their wellness and disease. However, according to the different stages of the mHealth app usability evaluation framework, the critical measures and evaluation methods are inconsistent. Future research should focus on adapting usability evaluation methods to the elderly by improving traditional tools or introducing automated evaluation tools.

Abbreviations

mHealth: Mobile health; mHealth apps: Mobile health applications; SUS: Systems Usability Scale; USE: Usefulness Satisfaction and Ease of Use questionnaire; UEQ: User Experience Questionnaire; ASQ: After Scenario Questionnaire; NASA-TLX: National Aeronautics and Space Administration task load index; NPS: Net Promoter Score; Health-ITUES: Health Information Technology Usability Evaluation Scale; QUIS: Questionnaire for User Interaction Satisfaction; PSSUQ: Post-Study System Usability Questionnaire; ICF-US: International Classification of Functioning based Usability Scale; MARS: Mobile Application Rating Scale; AUE: Automated usability evaluation; UAF: User action framework

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors report no declarations of interest.

Funding

This work was funded by the Shanghai Science and Technology Committee under Grant [16XD1403200].

Authors' Contributions

QW and LZ formulated and designed the study topic. QW participated in designing, coordinating, and conducting the study and drafting the manuscript. JL participated in the definition of the review plan, monitored all phases of the review, participated in the decision making, and reviewed the manuscript. TJ, XC, and Wei Z contributed to the data analysis and the discussion of the results. HW, Wan Z, and YG contributed to manuscript development and refinement.

Acknowledgments

We would like to thank Ms. Xuemei Li, Naval Medical University, who assisted with developing the search strategy.

References

1. Beard JR, Officer A, de Carvalho IA, Sadana R, Pot AM, Michel J, et al. The World report on ageing and health: A policy framework for healthy ageing. *The Lancet* 2016 May 21;387(10033):2145–2154. [doi: 10.1016/S0140-6736(15)00516-4] [Medline: 26520231]
2. United Nations. World population ageing 2020 highlights: Living arrangements of older persons. 2020. URL: <https://www.un.org/development/desa/pd/news/world-population-ageing-2020-highlights> [accessed 2021-05-21]
3. Chhetri J, Xue Q, Chan P. Intrinsic capacity as a determinant of physical resilience in older adults. *J Nutr Health Aging* 2021 Feb;25(8):1006–1011. [doi: 10.1007/s12603-021-1629-z] [Medline: 34545921]
4. Liao Q, Lau W, MCGhee S, Yap M, Sum R, Liang J, et al. Barriers to preventive care utilization among Hong Kong community-dwelling older people and their views on using financial incentives to improve preventive care utilization. *Health Expect* 2021 Aug;24(4):1242–1253. [doi: 10.1111/hex.13256] [Medline: 33949749]
5. World Health Organization. MHealth: New horizons for health through mobile technologies. 2011. URL: https://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2021-10-27]

6. Research2Guidance. MHealth App Economics 2017. 2017. URL: <https://research2guidance.com/wp-content/uploads/2017/10/1-mHealth-Status-And-Trends-Reports.pdf> [accessed 2021-10-27]
7. Pea research center. Millennials stand out for their technology use, but older generations also embrace digital life. 2019 Sep 09. URL: <https://www.benton.org/headlines/millennials-stand-out-their-technology-use-older-generations-also-embrace-digital-life-0> [accessed 2021-10-27].
8. Siaw-Teng L, Ansari S, Jonnagaddala J, Narasimhan P, Ashraf MM, Harris-Roxas B, et al. Use of mHealth for promoting healthy ageing and supporting delivery of age-friendly care services: A systematic review. *Int J Integr Care* 2019 Aug 08;19(S1):1–2. [doi: 10.5334/ijic.s3147]
9. Elavsky S, Knapova L, Klocek A, Smahel D. Mobile health interventions for physical activity, sedentary behavior, and sleep in adults aged 50 years and older: A systematic literature review. *J Aging Phys Activ* 2019 Aug;27(4):565–593. [doi: 10.1123/japa.2017-0410] [Medline: 30507266]
10. Peek STM, Luijckx KG, Vrijhoef HJM, Nieboer ME, Aarts S, van der Voort CS, et al. Understanding changes and stability in the long-term use of technologies by seniors who are aging in place: A dynamical framework. *Bmc Geriatr* 2019 Aug 28;19(1):e236. [doi: 10.1186/s12877-019-1241-9] [Medline: 31462214]
11. Puri A, Kim B, Nguyen O, Stolee P, Tung J, Lee J. User acceptance of Wrist-Worn activity trackers among Community-Dwelling older adults: Mixed method study. *Jmir Mhealth Uhealth* 2017 Nov 15;5(11):e173. [doi: 10.2196/mhealth.8211] [Medline: 29141837]
12. Hoque R, Sorwar G. Understanding factors influencing the adoption of mHealth by the elderly: An extension of the UTAUT model. *Int J Med Inform* 2017 May;101:75–84. [doi: 10.1016/j.ijmedinf.2017.02.002] [Medline: 28347450]
13. Lan Z, Liu H, Yang C, Liu X, Sorwar G. Investigating influencing factors of chinese elderly users' intention to adopt MHealth based on the UTAUT2 model. Association for Computing Machinery; BIBE 2020: Proceedings of the Fourth International Conference on Biological Information and Biomedical Engineering; 2020; July 21-July 23; Chengdu, China.
14. Chen K, Chan AHS. Gerontechnology acceptance by elderly Hong Kong Chinese: A senior technology acceptance model (STAM). *Ergonomics* 2014 Mar 24;57(5):635–652. [doi: 10.1080/00140139.2014.895855] [Medline: 24655221]
15. International Organization for Standardization. Ergonomics of human-system interaction-Part 11: Usability: Definitions and concepts. 2018. URL: <https://www.iso.org/standard/63500.html>[accessed 2021-10-28]
16. Bevan N, Carter J, Earthy J, Geis T, Harker S. New ISO standards for usability, usability reports and usability measures. In: Proceedings of the 18th International Conference on Human-Computer Interaction. Toronto, Canada; Springer International Publishing; 2016. p. 268–278.
17. Narasimha S, Madathil KC, Agnisarman S, Rogers H, Welch B, Ashok A, et al. Designing telemedicine systems for geriatric patients: A review of the usability studies. *Telemed E-Health* 2017 Jun;23(6):459–472. [doi: 10.1089/tmj.2016.0178] [Medline: 27875667]
18. Harte R, Quinlan LR, Glynn L, Rodríguez-Molinero A, Baker PM, Scharf T, et al. Human-Centered design study: Enhancing the usability of a mobile phone app in an integrated falls risk detection system for use by older adult users. *Jmir Mhealth Uhealth* 2017 May 30;5(5):e71-e71. [doi: 10.2196/mhealth.7046] [Medline: 28559227]
19. Ruzic L, Sanfod JA. Universal design mobile interface guidelines (UDMIG) for an aging population. In: H Marston, S Freeman, C Musselwhite, editors. Mobile e-Health. Human-Computer Interaction Series. Springer International Publishing; 2017: 17–37.
20. Sadegh SS, Khakshour Saadat P, Sepehri MM, Assadi V. A framework for m-health service development and success evaluation. *Int J Med Inform* 2018 Apr;112:123–130. [doi: 10.1016/j.ijmedinf.2018.01.003] [Medline: 29500009]
21. Harte R, Glynn L, Rodríguez-Molinero A, Baker PM, Scharf T, Quinlan LR, et al. A Human-Centered design methodology to enhance the usability, human factors, and user experience of connected health systems: A Three-Phase methodology. *JMIR human factors* 2017 Mar 16;4(1):e8-e8. [doi: 10.2196/humanfactors.5443] [Medline: 28302594]
22. National Institute of Standards and Technology. NIST guide to the processes approach for improving the usability of electronic health records. 2010. URL: <https://www.nist.gov/publications/nistir-7741-nist-guide-processes-approach>

improving-usability-electronic-health-records [accessed 2021-10-28]

23. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: A systematic literature review. *J Med Syst* 2015 Feb;39(2):1–19. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
24. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: A scoping review. *Int J Med Inform* 2019 Jun;126:95–104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]
25. Davis R, Gardner J, Schnall R. A review of usability evaluation methods and their use for testing eHealth HIV interventions. *Curr Hiv-Aids Rep* 2020 Jun;17(3):203–218. [doi: 10.1007/s11904-020-00493-3] [Medline: 32390078]
26. Wildenbos GA, Peute L, Jaspers M. Aging barriers influencing mobile health usability for older adults: A literature based framework (MOLD-US). *Int J Med Inform* 2018 Jun;114:66–75. [doi: 10.1016/j.ijmedinf.2018.03.012] [Medline: 29673606]
27. Grand View Research. MHealth Apps Market Size, Share & Trends Analysis Report by Type (Fitness, Medical), by Region (North America, APAC, Europe, MEA, Latin America), and Segment Forecasts, 2021–2028. 2021 Feb. URL:<https://www.marketresearch.com/Grand-View-Research-v4060/mHealth-Apps-Size-Share-Trends-14163352/> [accessed 2021-10-28].
28. International Organization for Standardization. Ergonomics of human-system interaction-Part 210: Human-Centred design for interactive systems. 2019. URL: <https://www.iso.org/standard/77520.html>[accessed 2021-10-28]
29. Yen P, Bakken S. Review of health information technology usability study methodologies. *J Am Med Inform Assn* 2012 May;19(3):413–422. [doi: 10.1136/amiajnl-2010-000020] [Medline: 21828224]
30. International Organization for Standardization. Ergonomics of human-system interaction-Usability methods supporting human-centred design. 2002. URL: <https://www.iso.org/standard/31176.html>[accessed 2021-10-28]
31. Molina-Recio G, Molina-Luque R, Jiménez-García AM, Ventura-Puertos PE, Hernández-Reyes A, Romero-Saldaña M. Proposal for the User-Centered design approach for health apps based on successful experiences: Integrative review. *Jmir Mhealth Uhealth* 2020 Apr 22;8(4):e14376. [doi: 10.2196/14376] [Medline: 32319965]
32. Cornet VP, Toscos T, Bolchini D, Rohani GR, Ahmed R, Daley C, et al. Untold stories in User-Centered design of mobile health: Practical challenges and strategies learned from the design and evaluation of an app for older adults with heart failure. *Jmir Mhealth Uhealth* 2020 Jul 21;8(7):e17703. [doi: 10.2196/17703] [Medline: 32706745]
33. Levac D, Colquhoun H, O'Brien KK. Scoping studies: Advancing the methodology. *Implement Sci* 2010 Sep 20;5:69–78. [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
34. Anderson JK, Howarth E, Vainre M, Humphrey A, Jones PB, Ford TJ. Advancing methodology for scoping reviews: Recommendations arising from a scoping literature review (SLR) to inform transformation of Children and Adolescent Mental Health Services. *Bmc Med Res Methodol* 2020 Sep 29;20(1) [doi: 10.1186/s12874-020-01127-3] [Medline: 32993505]
35. Graneheim UH, Lindgren B, Lundman B. Methodological challenges in qualitative content analysis: A discussion paper. *Nurs Educ Today* 2017 Sep;56:29–34. [doi: 10.1016/j.nedt.2017.06.002] [Medline: 28651100]
36. Siek KA, Ross SE, Khan DU, Haverhals LM, Cali SR, Meyers J. Colorado Care Tablet: The design of an interoperable Personal Health Application to help older adults with multimorbidity manage their medications. *J Biomed Inform* 2010 Oct;43(5):S22-S26. [doi: 10.1016/j.jbi.2010.05.007] [Medline: 20937480]
37. Fromme EK, Kenworthy-Heinige T, Hribar M. Developing an easy-to-use tablet computer application for assessing patient-reported outcomes in patients with cancer. *Support Care Cancer* 2011 Jun;19(6):815–22. [doi: 10.1007/s00520-010-0905-y] [Medline: 20512360]
38. Hamid A, Sym FP. Designing for patient-centred factors in medical adherence technology. Proceedings of the 6th International Conference on Rehabilitation Engineering & Assistive Technology; 2012; January 1; Tampines, Singapore.
39. Vermeulen J, Neyens JC, Spreeuwenberg MD, van Rossum E, Sipers W, Habets H, et al. User-centered development and testing of a monitoring system that provides feedback regarding physical functioning to elderly people. *Patient Prefer*

- Adherence 2013 Aug 27;7:843–854. [doi: 10.2147/PPA.S45897] [Medline: 24039407]
40. Hamm J, Money AG, Atwal A. Enabling older adults to carry out paperless falls-risk self-assessments using gidgetomeasure-3D: A mixed methods study. *J Biomed Inform* 2019 Apr;92(103–135) [doi: 10.1016/j.jbi.2019.103135] [Medline: 30826542]
 41. Hawley-Hague H, Tacconi C, Mellone S, Martinez E, Ford C, Chiari L, et al. Smartphone apps to support falls rehabilitation exercise: App development and usability and acceptability study. *Jmir Mhealth Uhealth* 2020 Sep 28;8(9):e15460. [doi: 10.2196/15460] [Medline: 32985992]
 42. Nikitina S, Didino D, Baez M, Casati F. Feasibility of virtual Tablet-Based group exercise among older adults in siberia: Findings from two pilot trials. *Jmir Mhealth Uhealth* 2018 Feb 27;6(2):e40. [doi: 10.2196/mhealth.7531] [Medline: 29487045]
 43. Ehn M, Eriksson LC, Akerberg N, Johansson A. Activity monitors as support for older persons' physical activity in daily life: Qualitative study of the users' experiences. *Jmir Mhealth Uhealth* 2018 Feb 1;6(2):e34. [doi: 10.2196/mhealth.8345] [Medline: 29391342]
 44. Compernelle S, Cardon G, van der Ploeg HP, Van Nassau F, De Bourdeaudhuij I, Jelsma JJ, et al. Engagement, acceptability, usability, and preliminary efficacy of a Self-Monitoring mobile health intervention to reduce sedentary behavior in belgian older adults: Mixed methods study. *Jmir Mhealth Uhealth* 2020 Oct 29;8(10):e18653. [doi: 10.2196/18653] [Medline: 33118951]
 45. Quinn CC, Staub S, Barr E, Gruber-Baldini A. Mobile support for older adults and their caregivers: Dyad usability study. *JMIR Aging* 2019 May 23;2(1):e12276. [doi: 10.2196/12276] [Medline: 31518271]
 46. Guay M, Latulippe K, Auger C, Giroux D, Seguin-Tremblay N, Gauthier J, et al. Self-Selection of Bathroom-Assistive Technology: Development of an Electronic Decision Support System (Hygiene 2.0). *J Med Internet Res* 2020 Aug 10;22(8):e161758. [doi: 10.2196/16175] [Medline: 32773380]
 47. Chung K, Kim S, Lee E, Park JY. Mobile app use for insomnia Self-Management in urban Community-Dwelling older korean adults: Retrospective intervention study. *Jmir Mhealth Uhealth* 2020 Aug 24;8(8):e17755. [doi: 10.2196/17755] [Medline: 32831177]
 48. Panagopoulos C, Kalatha E, Tsanakas P, Maglogiannis I. Evaluation of a mobile home care platform lessons learned and practical guidelines. In: B DeRuyter, A Kameas, P Chatzimisios, I Mavrommati, editors. *Lecture Notes in Computer Science*. Springer International Publishing; 2015: 328–343.
 49. Liu YC, Chen CH, Lin YS, Chen HY, Irianti D, Jen TN, et al. Design and usability evaluation of mobile Voice-Added food reporting for elderly people: Randomized controlled trial. *Jmir Mhealth Uhealth* 2020 Sep 28;8(9):e20317. [doi: 10.2196/20317] [Medline: 32985999]
 50. Ribeiro D, Machado J, Ribeiro J, Vasconcelos M, Vieira EF, de Banos AC. SousChef: Mobile meal recommender system for older adults. In: C Rocker, J ODonoghue, M Ziefle, L Maciaszek, W Molloy, editors. *Communications in Computer and Information Science*. Springer International Publishing; 2017: 36–45.
 51. Hill NL, Mogle J, Wion R, Kitt-Lewis E, Hannan J, Dick R, et al. App-based attention training: Incorporating older adults' feedback to facilitate home-based use. *Int J Older People Nurs* 2018 Mar;13(1):1–10. [doi: 10.1111/opn.12163] [Medline: 28752612]
 52. Boyd A, Synnott J, Nugent C, Elliott D, Kelly J. Community-based trials of mobile solutions for the detection and management of cognitive decline. *Healthc Technol Lett* 2017 May 5;4(3):93–96. [doi: 10.1049/htl.2016.0102] [Medline: 28706726]
 53. Arnhold M, Quade M, Kirch W. Mobile applications for diabetics: A systematic review and Expert-Based usability evaluation considering the special requirements of diabetes patients age 50 years or older. *J Med Internet Res* 2014 Apr 9;16(4):e104. [doi: 10.2196/jmir.2968] [Medline: 24718852]
 54. Gao C, Zhou L, Liu Z, Wang H, Bowers B. Mobile application for diabetes self-management in China: Do they fit for older adults? *Int J Med Inform* 2017 May;101:68–74. [doi: 10.1016/j.ijmedinf.2017.02.005] [Medline: 28347449]

55. Alvarez EA, Garrido M, Ponce DP, Pizarro G, Cordova AA, Vera F, et al. A software to prevent delirium in hospitalised older adults: Development and feasibility assessment. *Age Ageing* 2020 Feb 27;49(2):239–245. [doi: 10.1093/ageing/afz166] [Medline: 31957783]
56. Lunardini F, Luperto M, Romeo M, Basilico N, Daniele K, Azzolino D, et al. Supervised digital neuropsychological tests for cognitive decline in older adults: Usability and clinical validity study. *Jmir Mhealth Uhealth* 2020 Sep 21;8(9):e17963. [doi: 10.2196/17963] [Medline: 32955442]
57. Reading TM, Grossman LV, Baik D, Lee CS, Maurer MS, Goyal P, et al. Older adults can successfully monitor symptoms using an inclusively designed mobile application. *J Am Geriatr Soc* 2020 Jun;68(6):1313–1318. [doi: 10.1111/jgs.16403] [Medline: 32157679]
58. Guo Y, Chen Y, Lane DA, Liu L, Wang Y, Lip G. Mobile health technology for atrial fibrillation management integrating decision support, education, and patient involvement: MAF app trial. *Am J Med* 2017 Dec;130(12):1388–1396.e6. [doi: 10.1016/j.amjmed.2017.07.003] [Medline: 28847546]
59. World Health Organization. A handbook on how to implement mAgeing. 2018. URL: <https://www.who.int/home/cms-decommissioning> [accessed 2021-10-28]
60. Jonker LT, Plas M, de Bock GH, Buskens E, van Leeuwen BL, Lahr MMH. Remote home monitoring of older surgical cancer patients: Perspective on study implementation and feasibility. *Ann Surg Oncol* 2021 Jan;28(1):67–78. [doi: 10.1245/s10434-020-08705-1] [Medline: 32602060]
61. Simpson DB, Bird M, English C, Gall SL, Breslin M, Smith S, et al. "Connecting patients and therapists remotely using technology is feasible and facilitates exercise adherence after stroke". *Top Stroke Rehabil* 2020 Mar;27(2):93–102. [doi: 10.1080/10749357.2019.1690779] [Medline: 31762412]
62. Jing C. Design for elderly mobile medical App based on user experience. (Master Thesis) North China University of Technology, Beijing, China, 2018
63. Liu H, Li Y, Zhu L. Optimal design of APP user experience for the elderly based on conjoint analysis. *Packaging Engineering* 2018 Dec;39(24):264–270. [doi: 10.19554/j.cnki.1001-3563.2018.24.047]
64. Petrovcic A, Rogelj A, Dolnicar V. Smart but not adapted enough: Heuristic evaluation of smartphone launchers with an adapted interface and assistive technologies for older adults. *Comput Hum Behav* 2018 Feb;79:123–136. [doi: 10.1016/j.chb.2017.10.021]
65. Ray P, Li J, Ariani A, Kapadia V. Tablet-Based Well-Being check for the elderly: Development and evaluation of usability and acceptability. *JMIR Hum Factors* 2017 May 12;4(2):e12. [doi: 10.2196/humanfactors.7240] [Medline: 28500017]
66. Jansen-Kosterink S, Varenbrink P, Naafs A. The GezelschApp a dutch mobile application to reduce social isolation and loneliness. The 4th International Conference on Information and Communication Technologies for Ageing Well and e-Health; 2018; March 22-March 23; Funchal, Madeira.
67. Jansen-Kosterink SM, Bergsma J, Francissen A, Naafs A. The first evaluation of a Mobile application to encourage social participation for community-dwelling older adults. *Health and Technology* 2020 May 02;10(5):1107–1113. [doi: 10.1007/s12553-020-00430-9]
68. Nielsen J. Usability engineering. Cambridge, MA: AP Professional; 1994.
69. International Organization for Standardization. Systems and software engineering-Systems and software Quality Requirements and Evaluation (SQuaRE)-System and software quality models. 2011. URL: <https://www.iso.org/standard/35733.html>[accessed 2021-10-28]
70. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. *Int J Hum-Comput Int* 2008 Jul 30;24(6):574–594. [doi: 10.1080/10447310802205776]
71. Nielsen J. Usability inspection methods. Conference Companion on Human Factors in Computing Systems. Massachusetts, USA; 1994.
72. Barnum C. Usability testing essentials. Morgan Kaufmann; 2018.

73. Salman HM, Ahmad WFW, Sulaiman S. Usability evaluation of the smartphone user interface in supporting elderly users from experts' perspective. *Ieee Access* 2018 Apr 17;6:22578–22591. [doi: 10.1109/ACCESS.2018.2827358]
74. Silva PA, Holden K, Nii A. Smartphones, smart seniors, but Not-So-Smart apps: A heuristic evaluation of fitness apps. *Foundations of Augmented Cognition*. Springer International Publishing; 2014: 347–358.
75. Beauchemin M, Gradilla M, Baik D, Cho H, Schnall R. A multi-step usability evaluation of a Self-Management app to support medication adherence in persons living with HIV. *Int J Med Inform* 2019 Feb;122:37–44. [doi: 10.1016/j.ijmedinf.2018.11.012] [Medline: 30623782]
76. Harte R, Quinlan LR, Glynn L, Rodriguez-Molinero A, Baker PM, Scharf T, et al. Human-Centered design study: Enhancing the usability of a mobile phone app in an integrated falls risk detection system for use by older adult users. *Jmir Mhealth Uhealth* 2017 May 30;5(5):e71. [doi: 10.2196/mhealth.7046] [Medline: 28559227]
77. Morey SA, Stuck RE, Chong AW, Barg-Walkow LH, Mitzner TL, Rogers WA. Mobile health apps: Improving usability for older adult users. *Ergonomics in Design* 2019 Apr 01;27(4):4–13. [doi: 10.1177/1064804619840731]
78. Why You Only Need to Test with 5 Users. 2000 March 18. URL:<https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/> [accessed 2021-10-28].
79. Liu YC, Chen CH, Tsou YC, Lin YS, Chen HY, Yeh JY, et al. Evaluating mobile health apps for customized dietary recording for young adults and seniors: Randomized controlled trial. *Jmir Mhealth Uhealth* 2019 Feb 15;7(2):e10931. [doi: 10.2196/10931] [Medline: 30767906]
80. Grindrod KA, Li M, Gates A. Evaluating user perceptions of mobile medication management applications with older adults: A usability study. *Jmir Mhealth Uhealth* 2014 Mar 14;2(1):e11. [doi: 10.2196/mhealth.3048] [Medline: 25099993]
81. de Batlle J, Massip M, Vargiu E, Nadal N, Fuentes A, Ortega BM, et al. Implementing mobile Health-Enabled integrated care for complex chronic patients: Patients and professionals' acceptability study. *Jmir Mhealth Uhealth* 2020 Nov 20;8(11):e22136. [doi: 10.2196/22136] [Medline: 33216004]
82. Jr. Miller DP, Weaver KE, Case LD, Babcock D, Lawler D, Denizard-Thompson N, et al. Usability of a novel mobile health iPad app by vulnerable populations. *Jmir Mhealth Uhealth* 2017 Apr 11;5(4):e434. [doi: 10.2196/mhealth.7268] [Medline: 28400354]
83. Jankowski J, Saganowski S, Brodka P. Evaluation of TRANSFoRm mobile eHealth solution for remote patient monitoring during clinical trials. *Moile Information Systems* 2020 Jul 30;22(7):e18598. [doi: 10.1155/2016/1029368] [Medline: 32729843]
84. Hamm J, Money A, Atwal A. Fall prevention Self-Assessments via mobile 3D visualization technologies: Community dwelling older adults' perceptions of opportunities and challenges. *JMIR Hum Factors* 2017 Jun 19;4(2):e15. [doi: 10.2196/humanfactors.7161] [Medline: 28630034]
85. Wildenbos GA, Jaspers M, Schijven MP, Dusseljee-Peute LW. Mobile health for older adult patients: Using an aging barriers framework to classify usability problems. *Int J Med Inform* 2019 Apr;124:68–77. [doi: 10.1016/j.ijmedinf.2019.01.006] [Medline: 30784429]
86. Zmily A, Mowafi Y, Mashal E. Study of the usability of spaced retrieval exercise using mobile devices for alzheimer's disease rehabilitation. *Jmir Mhealth Uhealth* 2014 Aug 14;2(3):e31. [doi: 10.2196/mhealth.3136] [Medline: 25124077]
87. Isakovic M, Sedlar U, Volk M, Bester J. Usability pitfalls of diabetes mHealth apps for the elderly. *J Diabetes Res* 2016;2016:1–9. [doi: 10.1155/2016/1604609] [Medline: 27034957]
88. Brooke J. SUS - a quick and dirty usability scale. 1996. URL: https://digital.ahrq.gov/sites/default/files/docs/survey/systemusabilityscale%2528sus%2529_comp%255B1%255D.pdf [accessed 2021-10-28]
89. Gao M, Kortum P, Oswald F. Psychometric evaluation of the USE (Usefulness, satisfaction, and ease of use) questionnaire for reliability and validity. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 2018 Sep 27;62(1):1414–1418. [doi: 10.1177/1541931218621322] [Medline: 30784429]

90. Schrepp M. User experience questionnaire handbook. 2019. URL: <https://www.ueq-online.org/Material/Handbook.pdf> [accessed 2021-10-28]
91. Lewis JR. Psychometric evaluation of an after-scenario questionnaire for computer usability studies: The ASQ. ACM SIGCHI bulletin 1991 Jan 01;23(1):78–81. [doi: 10.1145/122672.122692] [Medline:
92. Hart SG, Staveland LE. Development of NASA-TLX (Task load index): Results of empirical and theoretical research. 52:139–183. [doi: 10.1016/S0166-4115(08)62386-9]
93. Schnall R, Cho H, Liu J. Health information technology usability evaluation scale (Health-ITUES) for usability assessment of mobile health technology: Validation study. Jmir Mhealth Uhealth 2018 Jan 5;6(1):e4. [doi: 10.2196/mhealth.8851] [Medline: 29305343]
94. The University of Maryland. QUIS: The questionnaire for user interaction satisfaction. 1999 March 09. URL:<https://www.umventures.org/technologies/quis%E2%84%A2-questionnaire-user-interaction-satisfaction-0> [accessed 2022-1-7].
95. Lewis JR. Psychometric evaluation of the Post-Study system usability questionnaire: The PSSUQ. Proceedings of the Human Factors Society annual meeting 1992 Oct 01;36(16):1259–1260. [doi: 10.1177/154193129203601617]
96. Teixeira A, Ferreira F, Almeida N, Silva S, Rosa AF, Pereira JC, et al. Design and development of Medication Assistant: Older adults centred design to go beyond simple medication reminders. Universal Access Inf 2016 Jul 11;16(3):545–560. [doi: 10.1007/s10209-016-0487-7]
97. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: A new tool for assessing the quality of health mobile apps. Jmir Mhealth Uhealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
98. Lewis JR. Usability: Lessons learned. And yet to be learned. Int J Hum-Comput Int 2014 Jul 10;30(9):663–684. [doi: 10.1080/10447318.2014.930311]
99. World Health Organization. World report on ageing and health. 2015. URL: <https://www.who.int/publications/i/item/9789241565042> [accessed 2021-10-28]
100. Hermawati S, Lawson G. Establishing usability heuristics for heuristics evaluation in a specific domain: Is there a consensus? Appl Ergon 2016 Sep;56:34–51. [doi: 10.1016/j.apergo.2015.11.016] [Medline: 27184309]
101. Li C, Neugroschl J, Zhu CW, Aloysi A, Schimming CA, Cai D, et al. Design considerations for mobile health applications targeting older adults. J Alzheimers Dis 2020 Nov 17;79(1):1–8. [doi: 10.3233/JAD-200485] [Medline: 33216024]
102. Ruzic L, Harrington C, Sanford J. Universal design mobile interface guidelines for mobile health and wellness apps for an aging population including people aging with disabilities. International Journal on Advances in Software 2017 Jan;10(3):372–384.
103. Wilson K, Bell C, Wilson L, Witteman H. Agile research to complement agile development: A proposal for an mHealth research lifecycle. NPJ Digit Med 2018 Sep 13;1:46–52. [doi: 10.1038/s41746-018-0053-1] [Medline: 31304326]
104. User Experience Professionals' Association. Usability benchmark. 2012 Dec. URL:<http://usabilitybok.org/usability-benchmark> [accessed 2021-10-28].
105. Barnum C. Big U and little u usability. Usability Testing Essentials. Morgan Kaufmann; 2020: 53–81.
106. Bhattarai P, Newton-John TRO, Phillips JL. Quality and usability of arthritic pain Self-Management apps for older adults: A systematic review. Pain Med 2018 Mar 1;19(3):471–484. [doi: 10.1093/pm/pnx090] [Medline: 28541464]
107. Broekhuis M, van Velsen L, Hermens H. Assessing usability of eHealth technology: A comparison of usability benchmarking instruments. Int J Med Inform 2019 Aug;128:24–31. [doi: 10.1016/j.ijmedinf.2019.05.001] [Medline: 31160008]
108. Alwashmi MF, Hawboldt J, Davis E, Feters MD. The iterative convergent design for mobile health usability testing: Mixed methods approach. Jmir Mhealth Uhealth 2019 Apr 26;7(4):e11656. [doi: 10.2196/11656] [Medline: 31025951]

109. Cho H, Powell D, Pichon A, Kuhns LM, Garofalo R, Schnall R. Eye-tracking retrospective think-aloud as a novel approach for a usability evaluation. *Int J Med Inform* 2019 Sep;129:366–373. [doi: 10.1016/j.ijmedinf.2019.07.010] [Medline: 31445278]
110. Wang J, Antonenko P, Celepkolu M, Jimenez Y, Fieldman E, Fieldman A. Exploring relationships between eye tracking and traditional usability testing data. *Int J Hum-Comput Int* 2018 Apr 30;35(6):483–494. [doi: 10.1080/10447318.2018.1464776]
111. Fabo P, Durikovic R. Automated usability measurement of arbitrary desktop application with eyetracking. 16th International Conference on Information Visualisation; 2012; July 11-July 13, 2012; Montprillier, France.
112. Harms P. Automated usability evaluation of virtual reality applications. *Acm T Comput-Hum Int* 2019 Jun;26(3):1–36. [doi: 10.1145/3301423]
113. Quintana Y, Fahy D, Abdelfattah AM, Henao J, Safran C. The design and methodology of a usability protocol for the management of medications by families for aging older adults. *BMC Med Inform Decis Mak* 2019 Sep 5;19(1):181. [doi: 10.1186/s12911-019-0907-8] [Medline: 31488134]
114. Petersen CL, Halter R, Kotz D, Loeb L, Cook S, Pidgeon D, et al. Using natural language processing and sentiment analysis to augment traditional User-Centered design: Development and usability study. *Jmir Mhealth Uhealth* 2020 Aug 7;8(8):e16862. [doi: 10.2196/16862] [Medline: 32540843]
115. Quintana M, Anderberg P, Sanmartin Berglund J, Frogren J, Cano N, Celtek S, et al. Feasibility-Usability study of a tablet app adapted specifically for persons with cognitive Impairment-SMART4MD (Support monitoring and reminder technology for mild dementia). *Int J Env Res Pub He* 2020 Sep 18;17(18):e6816. [doi: 10.3390/ijerph17186816] [Medline: 32961993]
116. Schmettow M. Sample size in usability studies. *Commun Acm* 2012 Apr;55(4):64–70. [doi: 10.1145/2133806.2133824]
117. Khajouei R, Peute LWP, Hasman A, Jaspers MWM. Classification and prioritization of usability problems using an augmented classification scheme. *J Biomed Inform* 2011 Dec;44(6):948–957. [doi: 10.1016/j.jbi.2011.07.002] [Medline: 21782036]

Figures

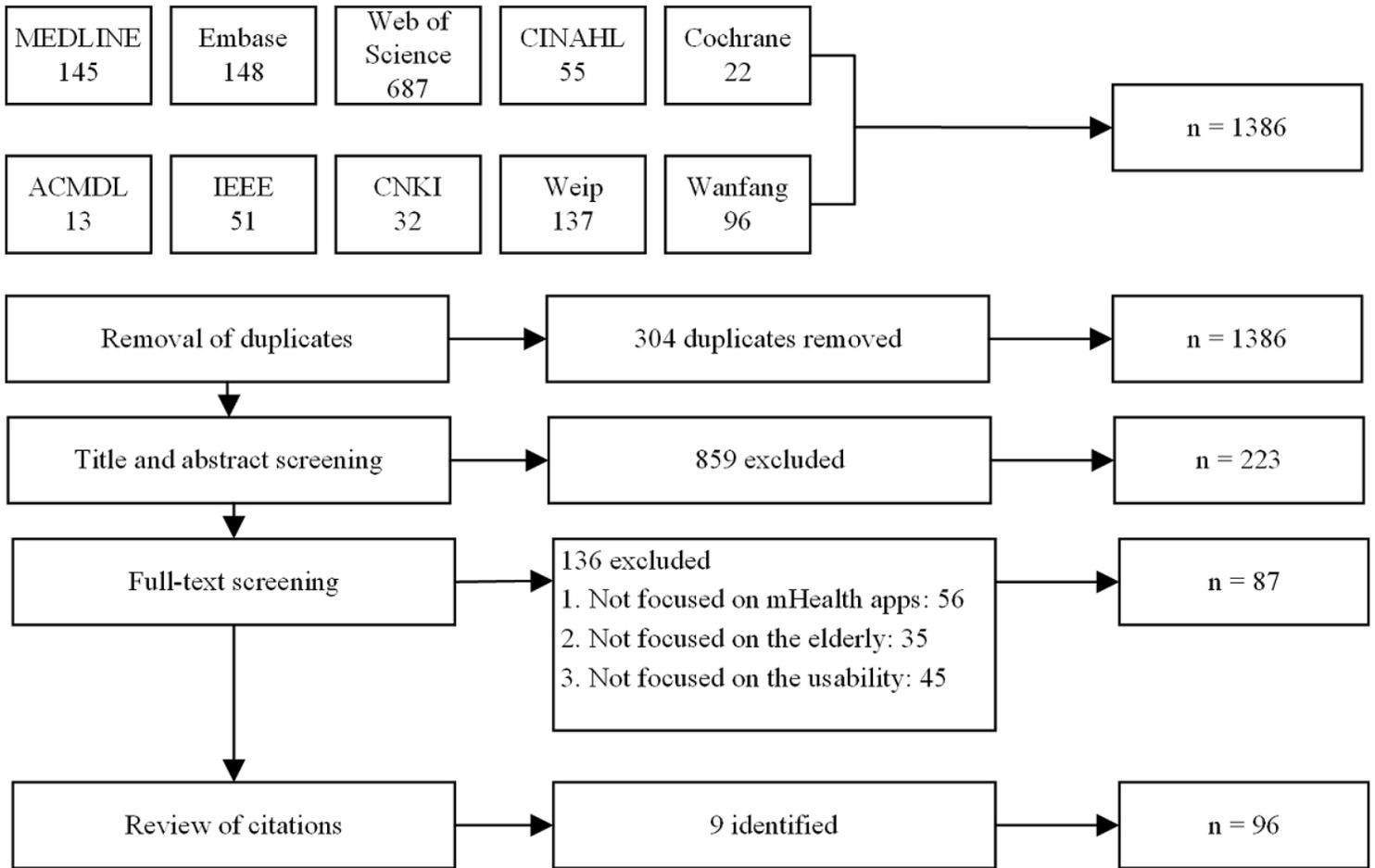


Figure 1

Flowchart of the study selection process

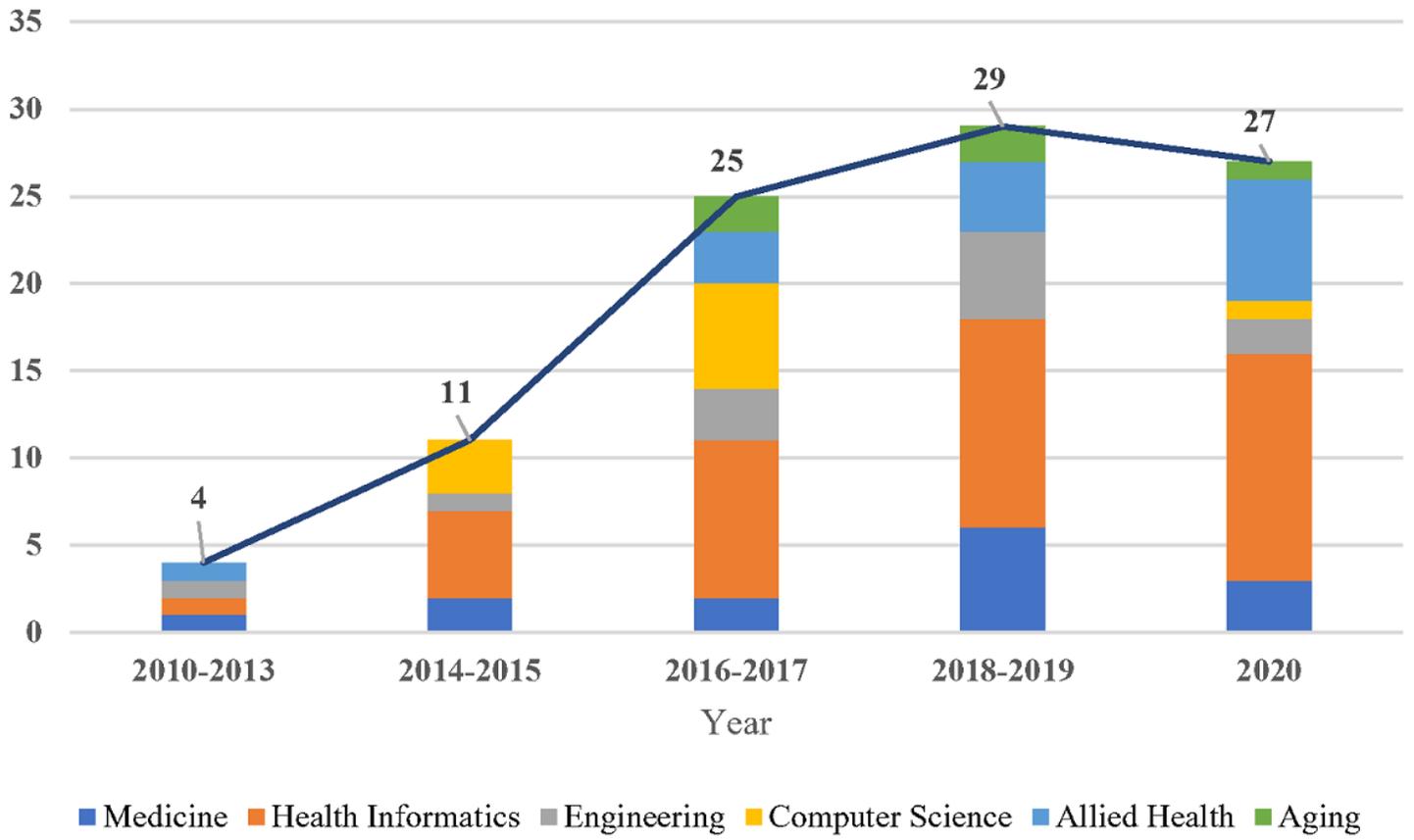


Figure 2

Number of articles published by year and type of journal

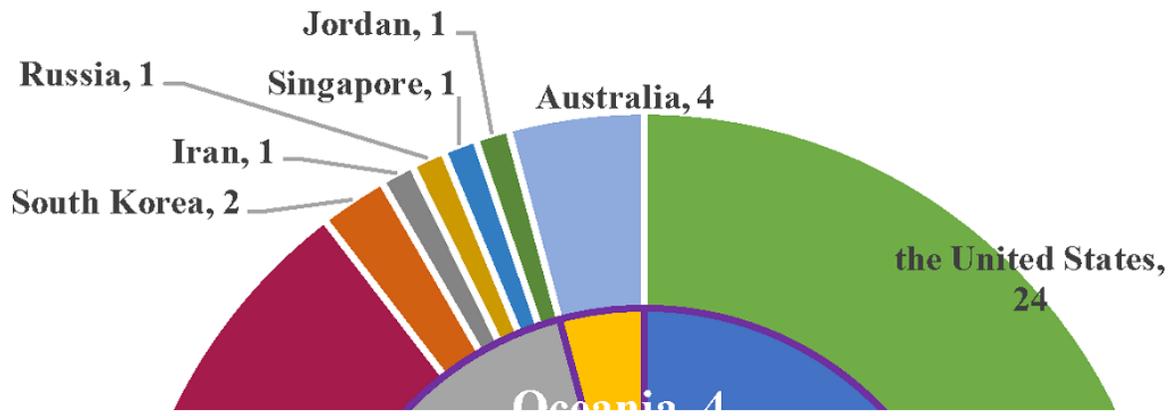


Figure 3

Country distribution of publications

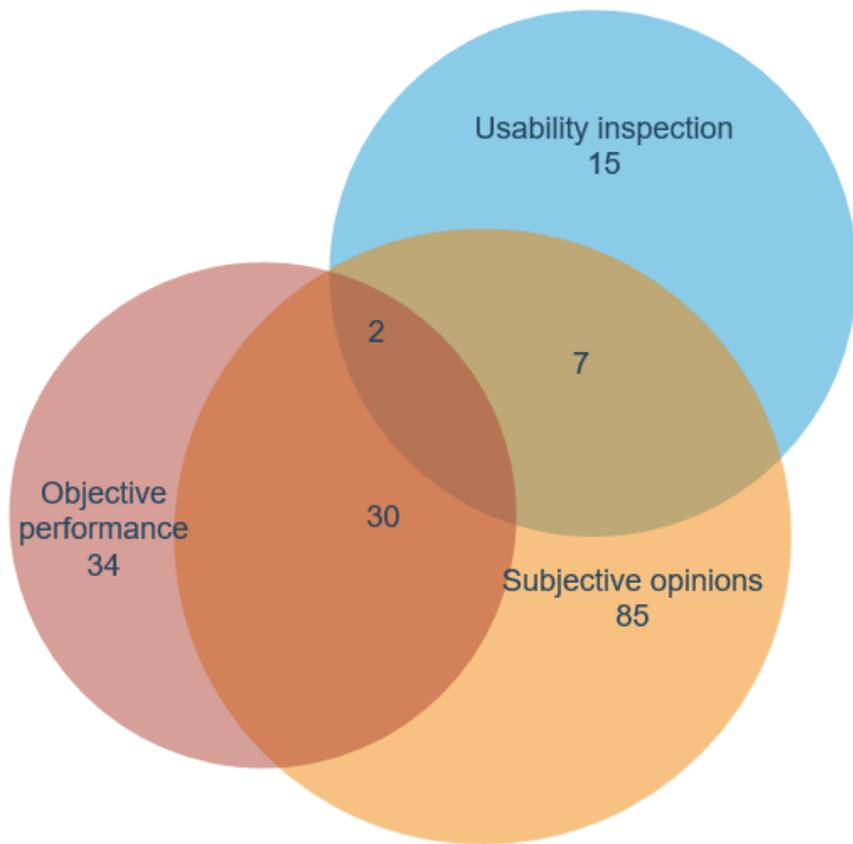


Figure 4

Categories of usability evaluation methods

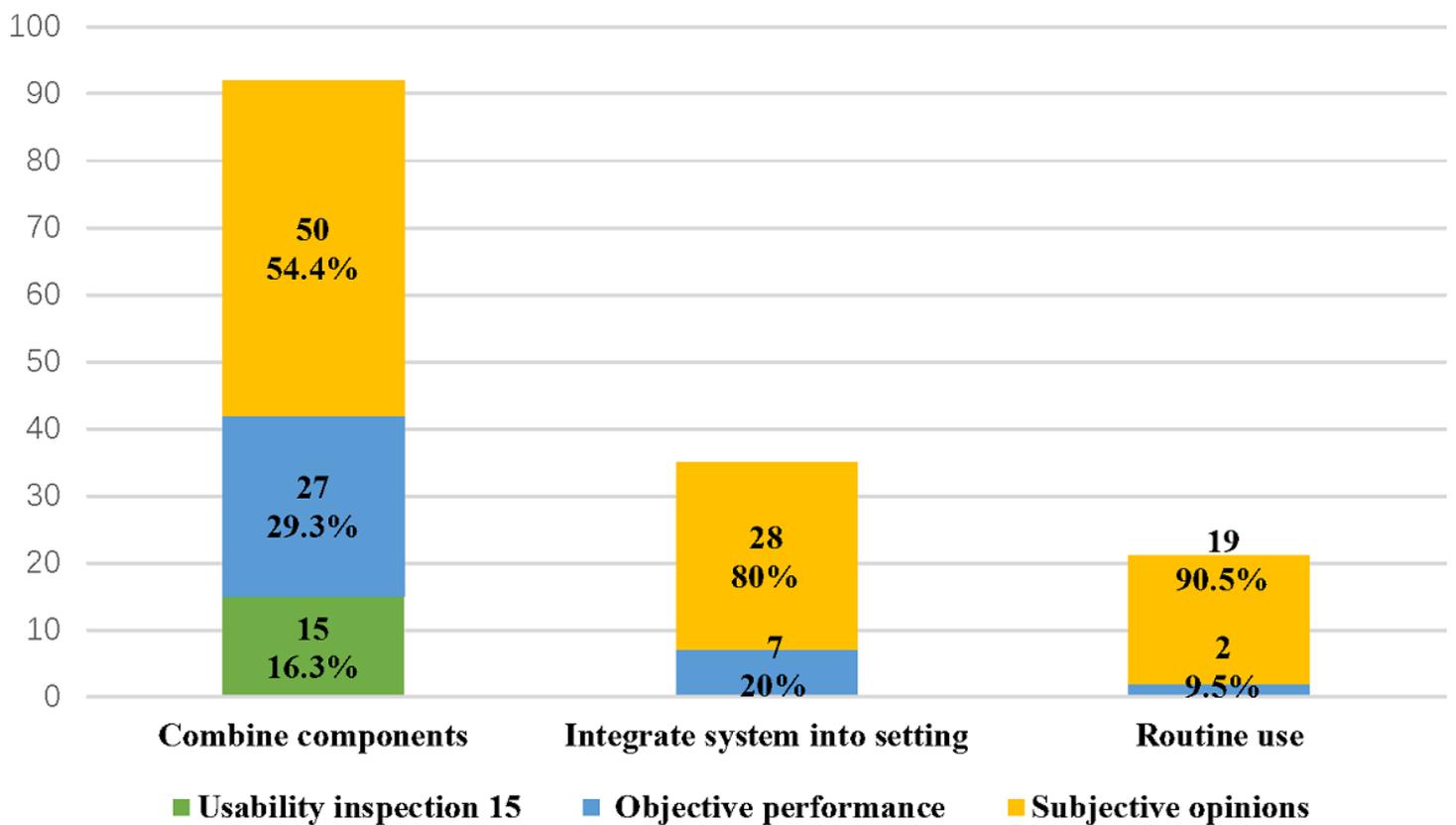


Figure 5

Distribution of the evaluation method types according to the mHealth app usability evaluation framework

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [MultimediaAppendix1.PRISMAScRFillableChecklist.docx](#)
- [MultimediaAppendix2.96includedarticles.docx](#)