Towards Usable E-Health

A Systematic Review of Usability Questionnaires

Vanessa E. C. Sousa[1]; Karen Dunn Lopez[1]

[1]Department of Health Systems Science, University of Illinois at Chicago, Chicago, IL, USA

Summary

Background: The use of e-health can lead to several positive outcomes. However, the potential for e-health to improve healthcare is partially dependent on its ease of use. In order to determine the usability for any technology, rigorously developed and appropriate measures must be chosen.

Objectives: To identify psychometrically tested questionnaires that measure usability of e-health tools, and to appraise their generalizability, attributes coverage, and quality.

Methods: We conducted a systematic review of studies that measured usability of e-health tools using four databases (Scopus, PubMed, CINAHL, and HAPI). Non-primary research, studies that did not report measures, studies with children or people with cognitive limitations, and studies about assistive devices or medical equipment were systematically excluded. Two authors independently extracted information including: questionnaire name, number of questions, scoring method, item generation, and psychometrics using a data extraction tool with pre-established categories and a quality appraisal scoring table.

Results: Using a broad search strategy, 5,558 potentially relevant papers were identified. After removing duplicates and applying exclusion criteria, 35 articles remained that used 15 unique questionnaires. From the 15 questionnaires, only 5 were general enough to be used across studies. Usability attributes covered by the questionnaires were: learnability (15), efficiency (12), and satisfaction (11). Memorability (1) was the least covered attribute. Quality appraisal showed that face/content (14) and construct (7) validity were the most frequent types of validity assessed. All questionnaires reported reliability measurement. Some questionnaires scored low in the quality appraisal for the following reasons: limited validity testing (7), small sample size (3), no reporting of user centeredness (9) or feasibility estimates of time, effort, and expense (7).

Conclusions: Existing questionnaires provide a foundation for research on e-health usability. However, future research is needed to broaden the coverage of the usability attributes and psychometric properties of the available questionnaires.

Correspondence to:

Vanessa E. C. Sousa, PhD, MSN
University of Illinois at Chicago, College of Nursing
Department of Health Systems Science
845 South Damen St.
Chicago, IL 60612
Email: vanessaemille@gmail.com
Phone: 773–814–0517

1. Background and Significance

E-Health is a broad term that refers a variety of technologies that facilitate healthcare, such as electronic communication among patients, providers and other stakeholders, electronic health systems and electronically distributed health services, wireless and mobile technologies for health care, telemedicine and telehealth and electronic health information exchange [1]. There has been exponential growth in the interest, funding, development and use of e-health in recent years [2–5]. Their use has led to a wide range of positive outcomes including improved: diabetes control outcomes [6], asthma lung functions [7], medication adherence [8], smoking cessation [9], sexually transmitted infection testing [10], tuberculosis cure rate [11], and reduced HIV viral load [12]. However, not all e-health demonstrate positive outcomes [13]. It is likely that even for e-health based in strong evidence based content, if the technology is difficult to use, the overall effectiveness on patient outcomes will be thwarted. In order to determine the ease of use (usability) for any new technology, rigorously developed and appropriate measures must be chosen [14, 15].

The term "usability" refers to set of concepts. Although usability is a frequently used term, it is inconsistently defined by both the research community [16] and standards organizations. The International Standards Organization (ISO) number 9241 defines usability as "the extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" [17]. Unfortunately, ISO has developed multiple usability standards each with differing terms and labels for similar characteristics [16].

In the absence of a clear consensus, we chose to use Jakob Nielsen's five usability attributes: learnability, efficiency, memorability, error rate and recovery, and satisfaction (▶Figure 1) [18]. Dr. Nielsen is highly regarded in the field of human computer interaction [19] and his usability attributes are the foundation for voluminous number of usability studies [20–25] including those of e-health [26–29]. Both ISO 9241 [17] and Nielsen's [18] definition share the concepts of efficiency and satisfaction, but the key advantage of Nielsen's definition over ISO's 9241 is the clarity and specificity of the additional three concepts of learnability, memorability and error rate and recovery over the more general concept of efficiency in ISO's definition.

[32–35] Usability together with utility (defined as whether a system provides the features a user needs) [18] comprise overall usefulness of a technology. Usability is so critical to the effectiveness of e-health that even applications with high utility may become unlikely to be accepted if usability is poor [30–33]. Beyond the problem of poor acceptance, e-health with compromised usability can also be harmful for patients. For example, medication errors that are facilitated by electronic health records (EHRs) with compromised efficiency affects the clinicians' ability to find needed test results, view coherent listings of medications, or review active problems which can result in delayed or incorrect treatment [34]. A computerized provider order entry system with too many windows can increase the likelihood of selecting wrong medications [35]. Other studies have shown that healthcare information technology with fragmented data (e.g. the need to open many windows to access patient data) leads to the loss of situation awareness, which compromises the quality of care and the ability to recognize and prevent impending events [36–38].

Given the importance of usability a wide range of methods for both developing usable technologies as well as assessing the usability of developed technologies have been created. Methods range from user inspection methods such as heuristic evaluations [39], qualitative think aloud interviews [39, 40], formal evaluations frameworks [41], simulated testing [42] and questionnaires. Although questionnaires have less depth than data yielded from qualitative analysis and may not be suitable when technologies are in the early stages of development, questionnaires play an important role in usability assessment. A well-tested questionnaire is generally much less expensive than using qualitative methods. In addition, unlike qualitative data, many questionnaires can be analyzed using predictive statistical analysis that can advance our understanding of technology use, acceptance and consequences.

Despite its importance, quantitative usability assessment of e-health applications have been hampered by two major problems. The first problem is that the concept of usability is often misapplied and not clearly understood. For example, some studies have reported methods of usability measurement that do not include any of its attributes [43] or include only partial assessments that do not

capture the whole meaning of usability [44–50]. It is also common to find usability being used as a synonym of acceptability [51, 52] or utility [53–56], further confounding the assessment of usability.

The second major problem is that many studies use usability measures that are very specific to an individual technology, or use questionnaires that lack psychometrics such as reliability (consistency of the measurement process) or validity (measurement of what it is supposed to be measured) [57–60]. Although we recognize that a usability assessment needs to consider specific or unique components of some e-health, we also believe that generalizable measures of usability that can be used across e-health types can be useful to the advancement of usability science. For example, to examine underuse of a patient portals, an analysis would be more robust if they measured both the patient and clinician portal interface using the same measure. There are also several benefits to having a generalizable usability measures to improve EHR usability. This includes comparative benchmarks for EHR usability across organizations, within organizations following upgrades, availability of comparable usability data prior to EHR purchase [61] and creating incentives for vendors to compete on usability performance [61]. In addition, having a generalizable measure of usability could be included as a way to operationalize Section 4002 (Transparent Reporting on Usability, Security, And Functionality) of the recently signed into law 21St Century Cures Act [62].

In sum, appropriate usability measurement is essential to give any technology implementation the best chance of success and to identify potential mediators of e-health engagement and health-related outcomes. However, questionnaires that have been used over the years to assess usability of e-health have not been systematically described or examined to guide the choice of strong measures for research in this area. The purpose of this review is to identify psychometrically tested questionnaires to measure usability of e-health tools, and to appraise their generalizability, attributes coverage, and quality.

2. Methods

2.1. Search Strategy and Study Selection

We conducted a review of four databases: Scopus, MEDLINE via PubMed, the Cumulative Index to Nursing Allied Health (CINAHL) via Ebscohost, and Health and Psychosocial Instruments (HaPI), from October 6th to November 3rd, 2015. Search terms included: "usability", "survey", "measure", "instrument", and "questionnaire", combined by the Boolean operators AND and OR. We used different search strategies in each database to facilitate the retrieval of studies that included measures of usability developed with users of e-health tools (▶Supplementary online File 1). Our search yielded 5,558 articles. After 2,206 duplicates were removed, a sample of 3,352 articles were considered for inclusion.

Papers were identified, screened and selected based on specific inclusion and exclusion criteria applied in three stages: 1) title review (2282 studies excluded), 2) abstract review (877 studies excluded), and 3) full article review (158 studies excluded) (▶Figure 2). Title exclusions rules included articles that were: not in English, not primary research, not inclusive of usability measures, not e-health tool related, conducted with children and people with cognitive limitations, and testing of assistive devices and medical equipment. We chose to exclude studies with children and people with cognitive limitations as subjects because questionnaires used with these populations may not be applicable to general adult users of e-health. We excluded assistive devices and medical equipment because of their specific features (e.g. user satisfaction with feedback for moving away from obstacles), which cannot be compared with questions applied to other types of technology.

Abstract based exclusion rules included articles that: used fully qualitative approach or met previous title exclusion rules. Full article based exclusion rules included articles that: did not assess any usability attribute, with usability measures unavailable, or with psychometrics unavailable in the article or citations. This provided a final sample of 35 articles.

Each round of exclusion (title, abstract and full article) was individually assessed by two authors using a sample of 5% of the articles. When disagreements arose, they were resolved by discussion. The use of two assessors continued for 3–6 rounds until inter-rater agreement was established for

each step at 85% (90% for titles review after 3 rounds, 85% for abstracts review after 6 rounds, and 97% for full articles review after 1 round).

2.2. Data Extraction and Analysis

Two investigators extracted data from papers using a data extraction tool built using Google Forms with variables, categories and definitions. We derived categories for variables using our general knowledge of usability, testing the categories using a subsample of articles, and adding new labels inductively, as necessary to accommodate important information that did not fall into any of the existing categories. Data entered was automatically stored in an online spreadsheet and assessed for agreement reaching the goal of >85% at the first round.

We extracted data into two stages. Stage 1 extracted general data from each of the 35 studies that met our inclusion and exclusion rules. This included: authors, place and year of publication, type of e-health technology evaluated, questionnaire used and origin of the questions. Because some of the questionnaires were used in more than one article, Stage 2 focused on the 15 unique questionnaires identified in the sample of 35 articles. For this stage, we extracted more specific data about the questionnaires' development and psychometric testing. When necessary, we extracted information about the questionnaires' development and psychometric assessment from the reference list of the original studies, or by the questionnaires' original authors.

2.3. Generalizability, Attributes Coverage and Quality Appraisal

Generalizability was assessed by one question asking if the questionnaires' items were generic or technology-related, i.e., whether the questionnaires included items referring to specific features that are not common across e-health applications. Attributes coverage was evaluated for each usability attribute (learnability, efficiency, memorability, error rate and recovery, and satisfaction) using Nielsen's definitions [18] (▶Figure 1). The quality assessment was comprised of the five criteria used in Hesselink, Kuis, Pijnenburg and Wollersheim [63]: validity, reliability, user centeredness, sample size, and feasibility (▶Supplementary online File 2). We did not include the criteria of responsiveness (the ability of a questionnaire to detect important changes over time) from the original tool because none of the studies assessed usability over time. The total score possible for each questionnaire is 10.

Definitions from the literature [57] for the specific types of validity and reliability can be seen in ▶Supplementary online File 3. From these definitions, we emphasize that user centeredness refers to "the inclusion of users' opinions or views to define or modify items". Thus, we carefully searched for information in each study about the participation of potential users during the questionnaires' development stage to rate this construct and based our ratings on whether this information was present or not in the published article.

3. Results

Our inclusion and exclusion rules yielded 35 unique articles that used 15 unique questionnaires.

3.1. Descriptive Analysis of Studies

The majority of the 35 studies were conducted in the United States (n=13) [64–76] and in European countries (n=8) [77–84]. The primary subjects were health workers (n=17) [59, 65, 66, 70–72, 75, 78, 80, 81, 85–91] and patients (n=11) [69, 73, 74, 76, 78, 79, 84, 91–94]. The types of e-health tools tested in the studies were comprised of: clinician-focused systems (n=15) [59, 65, 66, 70–72, 75, 80, 81, 85–88, 90, 95], patient-focused systems (n=10) [64, 67, 69, 74, 76, 83, 91–94], wellness/fitness applications (n=4) [77, 79, 84, 96], electronic surveys (n=3) [73, 78, 97], and digital learning objects (n=3) [68, 82, 89] (▶Supplementary online File 4).

3.2 Descriptive Analysis of Questionnaires

We identified 15 unique questionnaires across the 35 studies that measured individual perceptions of usability. The number of items in the questionnaires ranged from 3 to 38. Fourteen out of 15 questionnaires were Likert type, and 1 used a visual analog scale [92]. Eleven of the questionnaires [59, 82, 85, 86, 88, 90, 96, 98–101] had subscales. Most questionnaires were derived from empirical studies (pilot testing with human subjects) [59, 82, 85, 90, 98–100, 102, 103]. The others were derived from theories or models and from the literature (▶Table 1).

3.3. Generalizability

Generalizability assessments revealed that 10 questionnaires were created by the studies' authors specifically for their e-health and contain several items that may be too specific to be generalized (e.g. "*The built-in hot keys on the CPOE system facilitate the prescription of physician orders*"). Only 5 questionnaires include items that could potentially be applied across different types of e-health tools: the System Usability Scale (SUS) [103], the Questionnaire for User Interaction Satisfaction (QUIS) [100], the After-Scenario Questionnaire (ASQ) [104], the Post-Study System Usability Questionnaire (PSSUQ) [99], and the Computer System Usability Questionnaire (CSUQ) [105].

3.4. Attributes Coverage

Each questionnaire was evaluated in terms of usability attributes coverage. Learnability was the most covered usability attribute (all questionnaires). The least assessed usability attributes were: (1) error rate/recovery, that was included in only 6 questionnaires: PSSUQ [99], CSUQ [105], QUIS [100], Lee, Mills, Bausell and Lu [88], Oztekin, Kong and Uysal [82], and Peikari, Shah, Zakaria, Yasin and Elhissi [90]; and (2) memorability, that was accessed only by Oztekin, Kong and Uysal [82]. The 4 questionnaires that had the highest attribute coverage were: QUIS [100], Lee, Mills, Bausell and Lu [88], PSSUQ [99], and CSUQ [105] (▶Figure 3 and ▶Figure 4). All but 2 questionnaires [59, 78, 82, 85, 86, 88, 90, 92, 96, 98–101] also included items that measured the e-healths' utility.

3.5. Quality Assessment of Studies

An overview of the quality appraisal is shown in ▶Table 2. Quality scores ranged from 1–7 of a possible 10 points, and the average score was 4.1 (SD 1.9). The maximum score (7 points) was achieved by only 2 questionnaires: SUS [103] and Peikari, Shah, Zakaria, Yasin and Elhissi [90].

Face or content validity, often used as interchangeable terms, were addressed in 14 questionnaires. Construct validity, performed by a series of hypothesis tests to determine if the measure reflects the unobservable constructs [106], was established by exploratory factor analysis in 7 questionnaires [82, 86, 98–100, 102, 107], and by confirmatory factor analysis in 2 questionnaires [82, 90]. Criterion validity, assessed by correlating the new measure with a well-established or "gold standard" measure [106], was addressed by only 1 study [102].

Reliability, a measure of reproducibility [106], was assessed for all questionnaires by a Cronbach's α coefficient or by composite reliability. All questionnaires had an acceptable or high reliability based on the 0.70 threshold [108]. Inter-item correlations were reported for only 1 questionnaire [109]. The questionnaire's correlation coefficients were weak for some items, but strong for others, based on a 0.50 threshold [110]. Test-retest reliability, a determination of the consistency of the responses over time, was assessed for 1 questionnaire [89] and resulted in a correlation coefficient above the minimum threshold of 0.70 [111].

In addition to classic psychometric evaluation of the measure's quality, Hesselink and colleagues [63] quality assessment method also includes sample size, feasibility and user centeredness. Three studies [82, 86, 99] had small samples (below 5 participants per item) and 5 studies [90, 98, 100, 102, 109] had acceptable samples (5–10 participants per item) based on Kass and Tinsley [112] guidelines. The remaining studies did not perform factor analysis so there was not a sample size standard to evaluate. Feasibility, related to respondent burden, was assessed in 8 studies, but this was measured in disparate ways. Six studies reported difficulties perceived by users [59, 85, 88, 90, 100,

103], 2 studies reported time needed for completion [99, 104], and another one measured training needs [85]. User centeredness, defined as taking users' perceptions into account during instrument development, was identified in 6 studies [59, 85, 88, 90, 100, 103].

4. Discussion

The aim of this study was to appraise the generalizability, attributes coverage, and quality of questionnaires used to assess usability of e-health tools. We were surprised to find that none of the questionnaires cover all of the usability attributes or achieved the highest possible quality score using Hesselink, Kuis, Pijnenburg and Wollersheim [63]. However, by combining the generalizability, attributes coverage, and quality criteria, we believe the strongest of the currently available the questionnaires are the SUS, the QUIS, the PSSUQ, and the CSUQ. Although the SUS does not cover efficiency, memorability or errors, it is a widely-used questionnaire [113] with general questions that can be applied to a wide range of e-health. In addition, the SUS achieved the highest quality score of the identified questionnaires. The QUIS, the PSSUQ, and the CSUQ also include measures that are general to many types of e-health and have the advantage of covering additional usability attributes (efficiency and errors) when compared with SUS. However, we emphasize that researchers should define which usability measures are the best fit for the intent of the study, technology being assessed, and context of use. For example, an EHR developer may be more concerned about creating a system with low error rate than user satisfaction, while the developer of an "optional" technology (e.g. patient portal or exercise tracking) is likely to need a measure of satisfaction. In addition, we acknowledge that there may be specific factors in some e-health (e.g. size and weight of a hand-held consumer focused EKG device) that need to be measured along with the general usability attributes. In these instances, we recommend using a high-quality general usability questionnaire along with well tested and e-health specific questions.

There were notable weaknesses found across many of the questionnaires. We were surprised to learn that most the questionnaires (10 out of 15) were specific to a single technology, as their items focus on aspects specific to unique e-health tools. Others have noted the challenges associated with a decision to use specific or generalizable usability measures and advocate for modifying items to address specific systems and user tasks under evaluation [114, 115]. Although this approach can be helpful, it can affect the research subjects' comprehension of the questions and can change the psychometric properties of the questionnaire. Thus, questionnaires having adapted or modified items need to be tested before use [116].

We believe there is value to having the community of usability scientists use common questionnaires that allow comparisons across technology types. For examples knowing the usability ratings of commerce focused applications versus e-health can help set benchmarks for raising the standards of e-health usability. In addition, having common questionnaires can help pin point usability issues in an underused e-health. One potential questionnaire to promote for such purposes is the SUS. The SUS is the only usability questionnaire we identified that allows researchers to "grade" their e-health on the familiar A-F grade range often used in education. It has also has been cited in more than 1,200 publications and translated in eight languages becoming one of the most widely used usability questionnaires [103, 107, 113, 117]. This is not to say that SUS does not have weaknesses. In particular, we believe this measurement tool could be strengthened by further validity testing and by expanding the usability coverage to include efficiency, memorability and errors.

We were a bit disheartened to find that no single questionnaire enables the assessment of all usability attributes defined by Nielsen [18]. Most questionnaires have items covering only one or two aspects of usability (learnability and efficiency were the most common attributes), while other important aspects (such as memorability and error rate/recovery) are left behind. Incomplete or inconsistent assessments of the usability of e-health technologies are problematic and can be harmful. For example, a technology can be subjectively pleasing but have poor learnability and memorability, requiring extra mental effort from its users [118]. An EHR can be easy to learn but at the same time can require the distribution of information over several screens (poor efficiency), resulting in increased workload and documentation time [119]. A computerized physician order entry system that make it difficult to detect errors can increase the probability of prescribing errors [35]. These

examples also serve to illustrate importance of usability to e-health and the fact that adoption is not a measure of success of e-health, especially for technologies that are mandated by organizations or strongly encourage through national policies.

We found that most questionnaires measured utility along with usability, and were especially surprised to find that these terms were used interchangeably. Our search also yielded studies (excluded using our exclusion rules) that did not measure any of the usability attribute despite using this term either in their titles, abstracts, or within the text [120, 121]. Together these findings suggests that despite several decades of measurement, usability is an immature concept that is not consistently defined [122] or universally understood.

The quality appraisal showed that most questionnaires lack robust validity testing despite being widely used. While all questionnaires have been accessed for internal consistency and face/content validity (assessed subjectively and considered the weakest form of validity), other types of validity (content, construct and criterion) are missing. Face and content validity are not enough to ensure that a questionnaire is valid because they do not necessarily refer to what the test actually measures, but rather to a cursory judgment of what the questions appear to measure [123]. Lacking of objective measures of validity can result in invalid or questionable study findings, since the variables may not accurately measure the underlying theoretical construct [124, 125].

Cronbach's alpha estimates of reliability were acceptable or high for most studies, thus indicating that these questionnaires have good internal consistency (assessment of how well individual items correlate between each other in measuring the same construct) [57]. Cronbach's alpha tends to be the most frequently used estimate of reliability mainly because the other estimates require experimental or quasi-experimental designs, which are difficult, time-consuming and expensive [126, 127]. Rather than an insurmountable weakness of the quality of the existing questionnaires, our findings endorse the idea that there is still opportunity to further develop the reliability and validity of these measures. In particular, we think that Heikkinen, Suomi, Jaaskelainen, Kaljonen, Leino-Kilpi and Salantera [92] using a visual analog scale, merits further development. One advantage of this type of measure is the ability to analyze the scores as ease of continuous variables that opens opportunities to conduct more advanced analytic methods when compared to ordered Likert categories that were used in most the questionnaires.

The quality appraisal also revealed a somewhat surprisingly finding: we identified a lack of user centeredness in all but 6 studies [59, 85, 88, 90, 100, 103]. Incorporating users' views to define or modify questionnaire's items is important in any kind of research that involves human perceptions [128]. If the goals of the questionnaire and each of its items are not clear, the answer's given by the users will not reflect what they really think, yielding invalid results.

Feasibility estimates (time, effort, and expenses involved in producing/using the questionnaire) were also lacking in most studies and were insufficiently described in others. Feasibility is an important concern because even highly reliable questionnaires can be too long, causing unwillingness to complete, mistakes or invalid answers [128]. At the same time that respondent burden is a problem, short questionnaires that are too brief (like the ASQ) can lead to insufficient coverage of the attributes intended to be measured [129].

Finally, we acknowledge that both quantitative and qualitative methods play important roles in technology development and improvement. While quantitative methods have the advantages of being generally inexpensive and more suitable for large sample studies, qualitative methods (like think-aloud protocols) are useful to provide details about specific sources of problems that quantitative measures cannot usually match. In addition, qualitative assessments provide information about user behaviors, routines, and a variety of other information that is essential to deliver a product that actually fit into a user's needs or desires [130]. Ideally, both qualitative and quantitative approaches should be applied in the design or improvement of technologies.

4.1. Limitations

Despite our best efforts for rigor in this systematic review, we note a few limitations. First, given the focus of the review on quantitative usability questionnaires, we did not include all methods of usability evaluation. This means that several well important approaches such as heuristic review methods [39], [41], and think aloud [39] were beyond the scope of this review. Second, although we used

specific terms to retrieve measures of usability, we found that several of the questionnaires also included questions about utility in their measures. Although, the concept of utility is related to usability, we do not consider this review to systematically address utility measures. Because of the importance of utility in combination with usability to user's technology acceptance, we think a review of utility measures in e-health is an important direction for future studies. Third, although we combined different terms to increase sensitivity we may have missed some articles with potentially relevant questionnaires because of the search strategy we used. For example, our exclusion criteria may have led to exclusion of research published in other languages than English. Finally, the scoring method selected for quality appraisal of the studies may have underestimated the quality of some studies by favoring certain methodological characteristics (e.g. having more than one type of validity/reliability) over others. However, we were unable to identify a suitable alternative method to appraise the quality of questionnaires.

5. Conclusions

Poor usability of e-health affect the chances of achieving both adoption and positive outcomes. This systematic review provides a synthesis of the quality of questionnaires that are currently available for usability measurement of e-health. We found that usability is often misunderstood, misapplied, and partially assessed, and that many researchers have used usability and utility as interchangeable terms. Although there are weaknesses in the existing questionnaires, efforts to include the strongest and most effective measures of usability in research could be the key to delivering the promise of e-health.

Clinical Relevance Statement

This article provides evidence on the generalizability, attributes coverage, and quality of questionnaires that have been used in research to measure usability of several types of e-health technology. This synthesis can help researchers in choosing the best measures of usability based on their intent and technology purposes. The study also contributes to a better understanding of concepts that are essential for developing and implementing usable and safe technologies in healthcare.

Questions

1. An electronic health record is used to retrieve a patient's laboratory results over time. In this hypothetical system, each laboratory result window only displays the results for one day, so the user must open several windows to access information for multiple dates to make clinical decisions. Which usability attribute is compromised?
A) Efficiency
B) Learnability
C) Memorability
D) Few errors

The correct answer is A: Efficiency. Efficiency is concerned with how quickly users can perform tasks once they learned how to use a system. Thus, a system that requires accessing several windows to perform a task is an inefficient system. The major problem that results from inefficient systems is that users' productivity may be reduced. You can rule out learnability because it is related to how easy it is for users to accomplish tasks, using a system for the first time. Memorability can also be ruled since it is related to the ability of a user to reestablish proficiency after a period not using a system. Errors can be ruled since it is concerned with how many errors users make, error severity or ease of error recovery [16].

2. What is the risk inherent in using a usability questionnaires that has not undergone formal validity testing?
A) Low internal consistency
B) Lack of user-centeredness
C) Inaccurate measurement of usability
D) Inability to compare technologies usability quantitatively

The correct answer is: C inaccurate measurement of usability. Validity testing is a systematic process hypothesis testing methods that concerned with whether a questionnaire measures what it intends to measure [93]. Without testing to determine whether the new questionnaire measures usability, it is possible that the questionnaire items targets a different concept. We found that several of the questionnaires included in this review appear to measure utility not usability. You can rule out internal consistency, often measured by the Cronbach alpha coefficient, because internal consistency refers to whether the items on a scale are homogenous [93]. You can also rule out user centeredness because validity testing does not concern itself with user centeredness. You can rule out option D (Inability to compare technologies usability quantitatively), because it is possible that many researchers may use the same questionnaire that has not been systematically validated allowing them to compare across technologies on the possibly invalid measure.

Conflict of Interest
The authors declare that they have no conflicts of interest in the research.

Human Subjects Protections
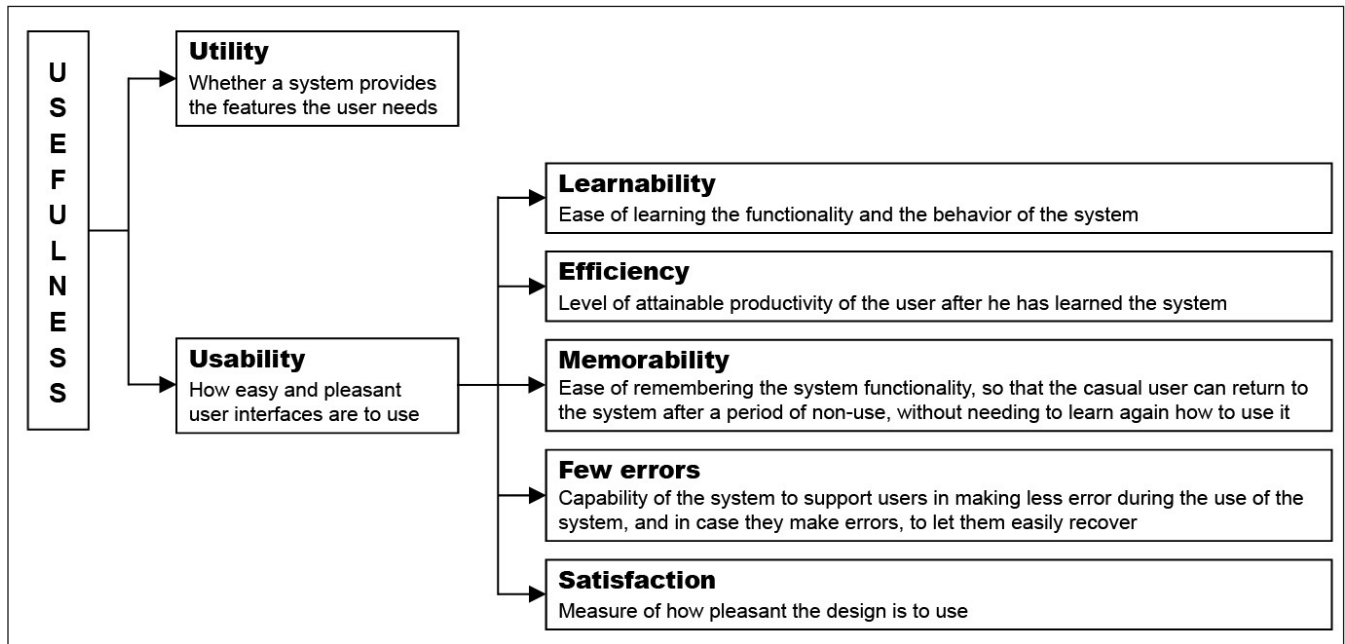Human subjects were not included in the project.

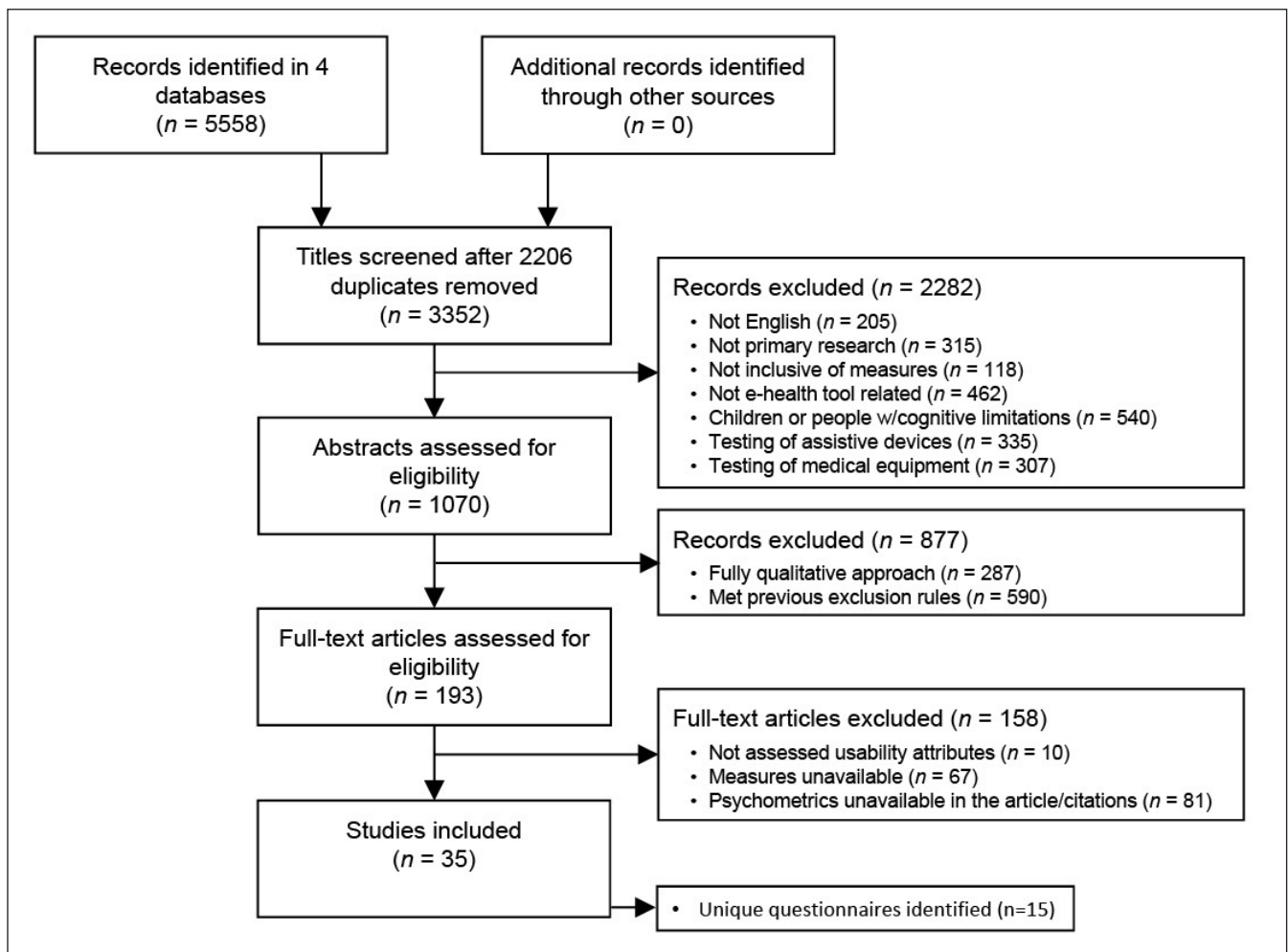**Fig. 1** Nielsen's attributes and definitions [18]

Fig. 2 PRISMA flow diagram for article inclusion

Fig. 3   Attributes covered by each questionnaire

Questionnaire name	Usability				
	Learnability	**Efficiency**	**Memorability**	**Errors**	**Satisfaction**
After-Scenario Questionnaire (ASQ)*	▪	▪			▪
Computer System Usability Questionnaire (CSUQ)*	▪	▪		▪	▪
Post-Study System Usability Questionnaire (PSSUQ)*	▪	▪		▪	▪
Questionnaire for User Interaction Satisfaction (QUIS)*	▪	▪		▪	▪
System Usability Scale (SUS)*	▪	▪			▪
Albu, Atack, and Srivastava (2015), Not named	▪	▪			
Fritz et al. (2012), Not named	▪	▪			▪
Hao et al. (2013), Not named	▪	▪			▪
Heikkinen et al. (2010), Not named	▪	▪			▪
Huang and Lee (2011), Not named	▪	▪			▪
Lee et al. (2008), Not named	▪	▪		▪	
Oztekin, Kong, and Uysal (2010), Not named	▪	▪	▪	▪	
Peikari et al. (2015), Not named	▪	▪		▪	
Wilkinson et al. (2004), Not named	▪	▪			▪
Yui et al. (2012), Not named	▪	▪			▪
Total	15	12	1	6	12

*All items generalizable across e-health

Table 1   Characteristics of the questionnaires.

Questionnaire name	Number of questions and scoring	Subscales	Item generation
After-Scenario Questionnaire (ASQ) [104]	3; 7-point Likert Scale ('strongly agree' to 'strongly disagree') and N/A	-	Empirical study
Computer System Usability Questionnaire (CSUQ) [105]	19; 7-point Likert Scale ('strongly agree' to 'strongly disagree') and N/A	System usefulness (8); Information quality (7); Interface quality (3)*	Empirical study
Post-Study System Usability Questionnaire (PSSUQ) [99]	19; 7-point Likert Scale ('strongly agree' to 'strongly disagree') and N/A	System usefulness (7); Information quality (6); Interface quality (3)†	Empirical study
Questionnaire for User Interaction Satisfaction (QUIS) [100]	27; 10-point Likert Scale (several adjectives positioned from negative to positive) and N/A	Overall reaction to the software (6); Screen (4); Terminology and system information (6); Learning (6); System capabilities (5)	Empirical study
System Usability Scale (SUS) [103]	10; 5-point Likert Scale ('strongly disagree' to 'strongly agree'	-	Empirical study
Albu, Atack, and Srivastava (2015), Not named [96]	12; 5-point Likert Scale ('strongly disagree' to 'strongly agree')	Ease of use (8); Usefulness (4)	Theory/Model
Fritz and colleagues (2012), Not named [78]	17; 5-point Likert Scale ('strongly disagree' to 'strongly agree')	-	Literature
Hao and colleagues (2013), Not named [85]	23; 5-point Likert Scale ('very satisfied' to 'very dissatisfied')	System Operation (5); System Function (4); Decision Support (5); System Efficiency (5); Overall Performance (4)	Empirical study
Heikkinen and colleagues (2010), Not named [92]	12; Visual analogic scale (100 mm)	-	Literature
Huang and Lee (2011), Not named [86]	30; 4-point Likert Scale ('no idea or disagreement' to 'absolute understanding or agreement')	Program design (8); Function (7); Efficiency (5); General satisfaction (10)	Literature
Lee and colleagues (2008), Not named [88]	30; 4-point Likert Scale ('strongly disagree' to 'strongly agree')	Patient care (6); Nursing efficiency (6); Education/training (6); Usability (6); Usage benefit (6)	Literature

Table 1   Continued

Questionnaire name	Number of questions and scoring	Subscales	Item generation
Oztekin, Kong, and Uysal (2010), Not named [82]	36; 5-point Likert Scale ('strongly disagree' to 'strongly agree')	Error prevention (3); Visibility (3); Flexibility (2); Course management (4); Interactivity, feedback and help (3); Accessibility (3); Consistency and functionality (3); Assessment strategy (3); Memorability (4); Completeness (3); Aesthetics (2); Reducing redundancy (3)	LiteratureEmpirical study
Peikari and colleagues (2015), Not named [90]	17; 5-point Likert Scale ('strongly disagree' to 'strongly agree')	Consistency (4); Ease of use (3); Error prevention (3); Information quality (3); Formative items (4)	LiteratureEmpirical study
Wilkinson and colleagues (2004), Not named [101]	38; 5-point Likert Scale ('strongly disagree' to 'strongly agree')	Computer use (7); Computer learning (5); Distance learning (4); Overall course evaluation (7); Fulfilment of learning outcomes (1); Course support (7); Utility of the course material (7)	NR‡
Yui and colleagues (2012), Not named [59]	28; 4-point Likert Scale ('strongly disagree' to 'strongly agree')	Interface design (6); Operation functions (11); Effectiveness (5); Satisfaction (6)	LiteratureEmpirical study

*Factor analysis of CSUQ showed that Item 19 loaded in two factor, thus this item was not included in any subscale. †The original version of PSSUQ did not contain Item 8. Factor analysis of PSSUQ showed that Items 15 and 19 loaded in two factor, thus they are not part of any subscale. ‡NR: Not Reported

Table 2   Quality appraisal of the questionnaires.

Questionnaire name	Validity	Reliability	User centeredness	Sample size*	Feasibility	Quality score
After-Scenario Questionnaire (ASQ) [104]	*Face/Content validity* Established by experts *Construct validity* PCA†: 8 factors accounted for 94% of total variance *Criterion validity* Correlation between ASQ and scenario success: -.40 (p<.01)	*Cronbach's* α α > 0.9	NR†	Sufficient	*Time* Participants took less time to complete ASQ than PSSUQ (amount of time not reported)	06
Computer System Usability Questionnaire (CSUQ) [105]	*Face/Content validity* Established by experts *Construct validity* PCA: 3 factors accounted for 98.6% of the total variance	*Cronbach's* α α= 0.95	NR	Sufficient	NR‡	05
Post-Study System Usability Questionnaire (PSSUQ) [99]	*Face/Content validity* Established by experts *Construct validity* PCA: 3 factors accounted for 87% of the total variance	*Cronbach's* α α = 0.97	NR	Low	*Time* Participants needed about 10 min to complete PSSUQ	05
Questionnaire for User Interaction Satisfaction (QUIS) [100]	*Face/Content validity* Established by experts *Construct validity* PCA: 4 latent factors resulted from the factor analysis	*Cronbach's* α α= 0.94	Included user's feedback	Sufficient	Perceived difficulty	05
System Usability Scale (SUS) [109]	*Face/Content validity* Established by people with different occupations *Construct validity* PCA§: 2 factors accounted for 56–58% of the total variance	*Cronbach's* α α = 0.91§ Inter-item correlation 0.34–0.69	Included user's feedback	Sufficient	Perceived difficulty	07
Albu, Atack, and Srivastava (2015), Not named [96]	NR	*Cronbach's* α α = 0.86	NR	-	NR	01
Fritz and colleagues (2012), Not named [78]	*Face/Content validity* Established by experts	*Cronbach's* α α = 0.84	NR	-	NR	03
Hao and colleagues (2013), Not named [85]	*Face/Content validity* Established by experts	*Cronbach's* α α = 0.91	Included user's feedback	-	Perceived difficulty Training needs	04
Heikkinen and colleagues (2010), Not named [92]	*Face/Content validity* Established by experts	*Cronbach's* α α = 0.84 to 0.87	NR	-	NR	03
Huang and Lee (2011), Not named [86]	*Face/Content validity* Established by experts *Construct validity* PCA: 3 factors accounted for 98.6% of the total variance	*Cronbach's* α α = 0.80	NR	Low	NR	03

Table 2 Continued

Questionnaire name	Validity	Reliability	User centeredness	Sample size*	Feasibility	Quality score
Lee and colleagues (2008), Not named [88]	*Face/Content validity* Established by experts	*Cronbach's α* α = 0.83 to 0.87	Included user's feedback	-	Perceived difficulty	06
Oztekin, Kong, and Uysal (2010), Not named [82]	*Face/Content validity* Established by the study authors *Construct validity* PCA: 12 factors accounted for 65.63% of the total variance CFAt: Composite Reliability: 0.7; Average Variance: 0.5	*Composite reliability* CR = 0.83 (Error prevention); 0.81 (Visibility); 0.73 (Flexibility); 0.89 (Management); 0.74 (Interactivity); 0.79 (Accessibility); 0.69 (Consistency)	NR	Low	NR	03
Peikari and colleagues (2015), Not named [90]	*Face/Content validity* Established by experts *Construct validity* CFA: Factor loadings: 0.80 to 0.87 (p<0.001, t=15.38); Average Variance: >0.5	*Cronbach's α* α = 0.79 (Information quality); 0.82 (Ease of use); 0.86 (Consistency); 0.78 (Error prevention)	Included user's feedback	Sufficient	Perceived difficulty	07
Wilkinson and colleagues (2004), Not named [101]	*Face/Content validity* Established by experts	*Cronbach's α* α = 0.84 (Computer use); 0.69 (Computer learning); 0.69 (Distance learning); 0.76 (Course evaluation); 0.91 to 0.94 (Learning outcomes); 0.87 (Course support); 0.75 (Utility) *Test–retest reliability* r=0.81	NR	-	NR	03
Yui and colleagues (2012), Not named [30]	*Face/Content validity* Established by experts	*Cronbach's α* α = 0.93	Included user's feedback	-	Perceived difficulty	06

*The sample size quality criterion was assessed only for studies who included factor analysis. †PCA: Principal Component Analysis; CFA: Confirmatory Factor Analysis. ‡NR: Not reported. §For the System Usability Scale, Cronbach's α value was extracted from Bangor, Kortum, and Miller (2008) using 2,324 cases, and PCA was extracted from Lewis and Sauro (2009) using Bangor, Kortum, and Miller (2008) data + 324 cases.

References

1. US Department of Health and Human Services. Expanding the reach and impact of consumer e-health tools. Washington, DC: US Department of Health and Human Services; 2006; Available from: https://health.gov/communication/ehealth/ehealthtools/default.htm.

2. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med 2015; 7(283): 283rv3.

3. Chen J, Bauman A, Allman-Farinelli M. A Study to Determine the Most Popular Lifestyle Smartphone Applications and Willingness of the Public to Share Their Personal Data for Health Research. Telemed J E Health 2016; 22(8): 655-665.

4. West JH, Hall PC, Hanson CL, Barnes MD, Giraud-Carrier C, Barrett J. There's an app for that: content analysis of paid health and fitness apps. J Med Internet Res 2012; 14(3): e72.

5. Silva BM, Rodrigues JJ, de la Torre Diez I, Lopez-Coronado M, Saleem K. Mobile-health: A review of current state in 2015. J Biomed Inform 2015; 56: 265-272.

6. Benhamou PY, Melki V, Boizel R, Perreal F, Quesada JL, Bessieres-Lacombe S, Bosson JL, Halimi S, Hanaire H. One-year efficacy and safety of Web-based follow-up using cellular phone in type 1 diabetic patients under insulin pump therapy: the PumpNet study. Diabetes Metab 2007; 33(3): 220-226.

7. Ostojic V, Cvoriscec B, Ostojic SB, Reznikoff D, Stipic-Markovic A, Tudjman Z. Improving asthma control through telemedicine: a study of short-message service. Telemed J E Health 2005; 11(1): 28–35.

8. DeVito Dabbs A, Dew MA, Myers B, Begey A, Hawkins R, Ren D, Dunbar-Jacob J, Oconnell E, McCurry KR. Evaluation of a hand-held, computer-based intervention to promote early self-care behaviors after lung transplant. Clin Transplant 2009; 23(4): 537-545.

9. Vidrine DJ, Marks RM, Arduino RC, Gritz ER. Efficacy of cell phone-delivered smoking cessation counseling for persons living with HIV/AIDS: 3-month outcomes. Nicotine Tob Res 2012; 14(1): 106-110.

10. Lim MS, Hocking JS, Aitken CK, Fairley CK, Jordan L, Lewis JA, Hellard ME. Impact of text and email messaging on the sexual health of young people: a randomised controlled trial. J Epidemiol Community Health 2012; 66(1): 69–74.

11. Kunawararak P, Pongpanich S, Chantawong S, Pokaew P, Traisathit P, Srithanaviboonchai K, Plipat T. Tuberculosis treatment with mobile-phone medication reminders in northern Thailand. Southeast Asian J Trop Med Public Health 2011; 42(6): 1444-1451.

12. Lester RT, Ritvo P, Mills EJ, Kariri A, Karanja S, Chung MH, Jack W, Habyarimana J, Sadatsafavi M, Najafzadeh M, Marra CA, Estambale B, Ngugi E, Ball TB, Thabane L, Gelmon LJ, Kimani J, Ackers M, Plummer FA. Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. Lancet 2010; 376(9755): 1838-1845.

13. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, Patel V, Haines A. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013; 10(1): e1001362.

14. Seidling HM, Stutzle M, Hoppe-Tichy T, Allenet B, Bedouch P, Bonnabry P, Coleman JJ, Fernandez-Llimos F, Lovis C, Rei MJ, Storzinger D, Taylor LA, Pontefract SK, van den Bemt PM, van der Sijs H, Haefeli WE. Best practice strategies to safeguard drug prescribing and drug administration: an anthology of expert views and opinions. Int J Clin Pharm 2016; 38(2): 362-373.

15. Kampmeijer R, Pavlova M, Tambor M, Golinowska S, Groot W. The use of e-health and m-health tools in health promotion and primary prevention among older adults: a systematic literature review. BMC Health Services Research 2016; 16(5): 290.

16. Abran A, Khelifi A, Suryn W, Seffah A. Usability meanings and interpretations in ISO standards. Software Qual J 2003; 11(4): 325-338.

17. International Organization for Standardization/International Electrotechnical Commission. ISO/IEC 9241–14 Ergonomic requirements for office work with visual display terminals (VDT)s – Part 14 Menu dialogues: ISO/IEC 9241–14; 1998.

18. Nielsen J. Usability Engineering. Boston: Academic Press; 1993.

19. Nielsen J. Lifetime Practice Award. Paris, France: SIGCHI; 2013.

20. Gorrell M. The 21st Century Searcher: How the Growth of Search Engines Affected the Redesign of EBSCOhost. Against the Grain 2008; 20(3): 22-26.

21. Sharma SK, Chen R, Zhang J. Examining Usability of E-learning Systems-An Exploratory Study (Research-in-Progress). International Proceedings of Economics Development and Research 2014; 81: 120.

22. Yan J, El Ahmad, A. S. Usability of CAPTCHAs or usability issues in CAPTCHA design. 4th symposium on Usable privacy and security; 2008; Pittsburgh, PA: ACM.

23. Jonsson A. Usability in three generations business support systems: Assessing perceived usability in the banking industry. Norrköping: Linköping University 2013.

24. Aljohani M, Blustein J. Heuristic evaluation of university institutional repositories based on DSpace. International Conference of Design, User Experience, and Usability: Springer International Publishing; 2015; 119-130.
25. Tapani J. Game Usability in North American Video Game Industry. Oulu: University of Oulu; 2016.
26. Lilholt PH, Jensen MH, Hejlesen OK. Heuristic evaluation of a telehealth system from the Danish Tele-Care North Trial. Int J Med Inform 2015; 84(5): 319-326.
27. Lu JM, Hsu YL, Lu CH, Hsu PE, Chen YW, Wang JA. Development of a telepresence robot to rebuild the physical face-to-face interaction in remote interpersonal communication. International Design Alliance (IDA) Congress Education Congress; October, 24–26th; Taipei, Taiwan; 2011.
28. Sparkes J, Valaitis R, McKibbon A. A usability study of patients setting up a cardiac event loop recorder and BlackBerry gateway for remote monitoring at home. Telemedicine and E-health 2012; 18(6): 484-490.
29. Ozok AA, Wu H, Garrido M, Pronovost PJ, Gurses AP. Usability and perceived usefulness of Personal Health Records for preventive health care: a case study focusing on patients' and primary care providers' perspectives. Appl Ergon 2014; 45(3): 613-628.
30. Zaidi ST, Marriott JL. Barriers and Facilitators to Adoption of a Web-based Antibiotic Decision Support System. South Med Rev 2012; 5(2): 42–50.
31. Carayon P, Cartmill R, Blosky MA, Brown R, Hackenberg M, Hoonakker P, Hundt AS, Norfolk E, Wetterneck TB, Walker JM. ICU nurses' acceptance of electronic health records. J Am Med Inform Assoc 2011; 18(6): 812-819.
32. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010; 43(1): 159-172.
33. Sun HM, Li SP, Zhu YQ, Hsiao B. The effect of user's perceived presence and promotion focus on usability for interacting in virtual environments. Appl Ergon 2015; 50: 126-132.
34. Koppel R. Is healthcare information technology based on evidence? Yearb Med Inform 2013; 8(1): 7–12.
35. Koppel R, Metlay JP, Cohen A, Abaluck B, Localio AR, Kimmel SE, Strom BL. Role of computerized physician order entry systems in facilitating medication errors. JAMA 2005; 293(10): 1197-1203.
36. Stead WW, Lin HS. Computational Technology for Effective Health Care: Immediate Steps and Strategic Directions. Washington (DC); 2009.
37. Koch SH, Weir C, Haar M, Staggers N, Agutter J, Gorges M, Westenskow D. Intensive care unit nurses' information needs and recommendations for integrated displays to improve nurses' situation awareness. J Am Med Inform Assoc 2012; 19(4): 583-590.
38. Lurio J, Morrison FP, Pichardo M, Berg R, Buck MD, Wu W, Kitson K, Mostashari F, Calman N. Using electronic health record alerts to provide public health situational awareness to clinicians. J Am Med Inform Assoc 2010; 17(2): 217-219.
39. Jaspers MW. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. Int J Med Inform 2009; 78(5): 340-353.
40. Lopez KD, Febretti A, Stifter J, Johnson A, Wilkie DJ, Keenan G. Toward a More Robust and Efficient Usability Testing Method of Clinical Decision Support for Nurses Derived From Nursing Electronic Health Record Data. Int J Nurs Knowl 2016 Jun 24.
41. Zhang J, Walji MF. TURF: toward a unified framework of EHR usability. J Biomed Inform 2011; 44(6): 1056-1067.
42. Sousa VE, Lopez KD, Febretti A, Stifter J, Yao Y, Johnson A, Wilkie DJ, Keenan GM. Use of Simulation to Study Nurses' Acceptance and Nonacceptance of Clinical Decision Support Suggestions. Comput Inform Nurs 2015; 33(10): 465-472.
43. Christensen BL, Kockrow EO. Adult health nursing. St. Louis: Elsevier Health Sciences; 2014.
44. Albertazzi D, Okimoto ML, Ferreira MG. Developing an usability test to evaluate the use of augmented reality to improve the first interaction with a product. Work 2011; 41: 1160-1163.
45. Freire LL, Arezes PM, Campos JC. A literature review about usability evaluation methods for e-learning platforms. Work-Journal of Prevention Assessment and Rehabilitation 2012; 41: 1038.
46. Yen PY, Bakken S. Review of health information technology usability study methodologies. J Am Med Inform Assoc 2012; 19(3): 413-422.
47. Santos NA, Sousa VEC, Pascoal LM, Lopes MVO, Pagliuca LMF, editors. Concept analysis of usability of healthcare technologies. 17th National Nursing Research Seminar; 2013; Natal, RN, Brazil: Brazilian Nursing Association.
48. Lei J, Xu L, Meng Q, Zhang J, Gong Y. The current status of usability studies of information technologies in China: a systematic study. Biomed Res Int 2014; 2014: 568303.
49. Lyles CR, Sarkar U, Osborn CY. Getting a technology-based diabetes intervention ready for prime time: a review of usability testing studies. Curr Diab Rep 2014; 14(10): 534.

50. Wootten AC, Abbott JAM, Chisholm K, Austin DW, Klein B, McCabe M, Murphy DG, Costello AJ. Development, feasibility and usability of an online psychological intervention for men with prostate cancer: My Road Ahead. Internet Interventions 2014; 1(4): 188-195.

51. Friedman K, Noyes J, Parkin CG. 2-Year follow-up to STeP trial shows sustainability of structured self-monitoring of blood glucose utilization: results from the STeP practice logistics and usability survey (STeP PLUS). Diabetes Technol Ther 2013; 15(4): 344-347.

52. Stafford E, Hides L, Kavanagh DJ. The acceptability, usability and short-term outcomes of Get Real: A web-based program for psychotic-like experiences (PLEs). Internet Interventions 2015; 2: 266-271.

53. Shah N, Jonassaint J, De Castro L. Patients welcome the Sickle Cell Disease Mobile Application to Record Symptoms via Technology (SMART). Hemoglobin 2014; 38(2): 99–103.

54. Shyr C, Kushniruk A, Wasserman WW. Usability study of clinical exome analysis software: top lessons learned and recommendations. J Biomed Inform 2014; 51: 129-136.

55. Anders S, Albert R, Miller A, Weinger MB, Doig AK, Behrens M, Agutter J. Evaluation of an integrated graphical display to promote acute change detection in ICU patients. Int J Med Inform 2012; 81(12): 842-851.

56. Yen PY, Gorman P. Usability testing of digital pen and paper system in nursing documentation. AMIA Annu Symp Proc 2005: 844-848.

57. McDermott RJ, Sarvela PD. Health education evaluation and measurement: A practitioner's perspective. 2 ed. Dubuque, Iowa: William C Brown Pub; 1998.

58. Taveira-Gomes T, Saffarzadeh A, Severo M, Guimaraes MJ, Ferreira MA. A novel collaborative e-learning platform for medical students – ALERT STUDENT. BMC Med Educ 2014; 14: 143.

59. Yui BH, Jim WT, Chen M, Hsu JM, Liu CY, Lee TT. Evaluation of computerized physician order entry system-a satisfaction survey in Taiwan. J Med Syst 2012; 36(6): 3817-3824.

60. Amrein K, Kachel N, Fries H, Hovorka R, Pieber TR, Plank J, Wenger U, Lienhardt B, Maggiorini M. Glucose control in intensive care: usability, efficacy and safety of Space GlucoseControl in two medical European intensive care units. BMC Endocr Disord 2014; 14: 62.

61. Ratwani RM, Hettinger AZ, Fairbanks RJ. Barriers to comparing the usability of electronic health records. J Am Med Inform Assoc 2016; DOI:10.1093/jamia/ocw117

62. 21st Century Cures Act, H.R.6. To accelerate the discovery, development, and delivery of 21st century cures, and for other purposes; 2015.

63. Hesselink G, Kuis E, Pijnenburg M, Wollersheim H. Measuring a caring culture in hospitals: a systematic review of instruments. BMJ Open 2013; 3(9): e003416.

64. Cox CE, Wysham NG, Walton B, Jones D, Cass B, Tobin M, Jonsson M, Kahn JM, White DB, Hough CL, Lewis CL, Carson SS. Development and usability testing of a Web-based decision aid for families of patients receiving prolonged mechanical ventilation. Ann Intensive Care 2015; 5: 6.

65. Im EO, Chee W, editors. Evaluation of the decision support computer program for cancer pain management. Oncology nursing forum 2006; Pittsburgh, PA: Oncology Nursing Society.

66. Kim MS, Shapiro JS, Genes N, Aguilar MV, Mohrer D, Baumlin K, Belden JL. A pilot study on usability analysis of emergency department information system by nurses. Appl Clin Inform 2012; 3(1): 135-153.

67. Kobak KA, Stone WL, Wallace E, Warren Z, Swanson A, Robson K. A web-based tutorial for parents of young children with autism: results from a pilot study. Telemed J E Health 2011; 17(10): 804-808.

68. Levin ME, Pistorello J, Seeley JR, Hayes SC. Feasibility of a prototype web-based acceptance and commitment therapy prevention program for college students. J Am Coll Health 2014; 62(1): 20–30.

69. Lin CA, Neafsey PJ, Strickler Z. Usability testing by older adults of a computer-mediated health communication program. J Health Commun 2009; 14(2): 102-118.

70. Saleem JJ, Haggstrom DA, Militello LG, Flanagan M, Kiess CL, Arbuckle N, Doebbeling BN. Redesign of a computerized clinical reminder for colorectal cancer screening: a human-computer interaction evaluation. BMC Med Inform Decis Mak 2011; 11: 74.

71. Schnall R, Cimino JJ, Bakken S. Development of a prototype continuity of care record with context-specific links to meet the information needs of case managers for persons living with HIV. Int J Med Inform 2012; 81(8): 549-555.

72. Schutte J, Gales S, Filippone A, Saptono A, Parmanto B, McCue M. Evaluation of a telerehabilitation system for community-based rehabilitation. Int J Telerehabil 2012; 4(1): 15–24.

73. Sharma P, Dunn RL, Wei JT, Montie JE, Gilbert SM. Evaluation of point-of-care PRO assessment in clinic settings: integration, parallel-forms reliability, and patient acceptability of electronic QOL measures during clinic visits. Qual Life Res 2016; 25(3): 575-583.

74. Shoup JA, Wagner NM, Kraus CR, Narwaney KJ, Goddard KS, Glanz JM. Development of an Interactive Social Media Tool for Parents With Concerns About Vaccines. Health Educ Behav 2015; 42(3): 302-312.

75. Staggers N, Kobus D, Brown C. Nurses' evaluations of a novel design for an electronic medication administration record. Comput Inform Nurs 2007; 25(2): 67–75.

76. Stellefson M, Chaney B, Chaney D, Paige S, Payne-Purvis C, Tennant B, Walsh-Childers K, Sriram P, Alber J. Engaging community stakeholders to evaluate the design, usability, and acceptability of a chronic obstructive pulmonary disease social media resource center. JMIR Res Protoc 2015; 4(1): e17.

77. Bozkurt S, Zayim N, Gulkesen KH, Samur MK, Karaagaoglu N, Saka O. Usability of a web-based personal nutrition management tool. Inform Health Soc Care 2011; 36(4): 190–205.

78. Fritz F, Balhorn S, Riek M, Breil B, Dugas M. Qualitative and quantitative evaluation of EHR-integrated mobile patient questionnaires regarding usability and cost-efficiency. Int J Med Inform 2012; 81(5): 303-313.

79. Meldrum D, Glennon A, Herdman S, Murray D, McConn-Walsh R. Virtual reality rehabilitation of balance: assessment of the usability of the Nintendo Wii((R)) Fit Plus. Disabil Rehabil Assist Technol 2012; 7(3): 205-210.

80. Memedi M, Westin J, Nyholm D, Dougherty M, Groth T. A web application for follow-up of results from a mobile device test battery for Parkinson's disease patients. Comput Meth Prog Bio 2011; 104(2): 219-226.

81. Ozel D, Bilge U, Zayim N, Cengiz M. A web-based intensive care clinical decision support system: from design to evaluation. Inform Health Soc Care 2013; 38(2): 79–92.

82. Oztekin A, Kong ZJ, Uysal O. UseLearn: A novel checklist and usability evaluation method for eLearning systems by criticality metric analysis. International Journal of Industrial Ergonomics 2010; 40: 455-469.

83. Stjernsward S, Ostman M. Illuminating user experience of a website for the relatives of persons with depression. Int J Soc Psychiatry 2011; 57(4): 375-386.

84. Van der Weegen S, Verwey R, Tange HJ, Spreeuwenberg MD, de Witte LP. Usability testing of a monitoring and feedback tool to stimulate physical activity. Patient Prefer Adherence 2014; 8: 311-322.

85. Hao AT, Wu LP, Kumar A, Jian WS, Huang LF, Kao CC, Hsu CY. Nursing process decision support system for urology ward. Int J Med Inform 2013; 82(7): 604-612.

86. Huang H, Lee TT. Evaluation of ICU nurses' use of the clinical information system in Taiwan. Comput Inform Nurs 2011; 29(4): 221-229.

87. Lacerda TC, von Wangenheim CG, von Wangenheim A, Giuliano I. Does the use of structured reporting improve usability? A comparative evaluation of the usability of two approaches for findings reporting in a large-scale telecardiology context. Journal of Biomedical Informatics 2014; 52: 222-230.

88. Lee TT, Mills ME, Bausell B, Lu MH. Two-stage evaluation of the impact of a nursing information system in Taiwan. Int J Med Inform 2008; 77(10): 698–707.

89. Moattari M, Moosavinasab E, Dabbaghmanesh MH, ZarifSanaiey N. Validating a Web-based Diabetes Education Program in continuing nursing education: knowledge and competency change and user perceptions on usability and quality. J Diabetes Metab Disord 2014; 13: 70.

90. Peikari HR, Shah MH, Zakaria MS, Yasin NM, Elhissi A. The impacts of second generation e-prescribing usability on community pharmacists outcomes. Res Social Adm Pharm 2015; 11(3): 339-351.

91. Sawka AM, Straus S, Gafni A, Meiyappan S, O'Brien MA, Brierley JD, Tsang RW, Rotstein L, Thabane L, Rodin G, George SR, Goldstein DP. A usability study of a computerized decision aid to help patients with, early stage papillary thyroid carcinoma in, decision-making on adjuvant radioactive iodine treatment. Patient Educ Couns 2011; 84(2): e24-e27.

92. Heikkinen K, Suomi R, Jaaskelainen M, Kaljonen A, Leino-Kilpi H, Salantera S. The creation and evaluation of an ambulatory orthopedic surgical patient education web site to support empowerment. Comput Inform Nurs 2010; 28(5): 282-290.

93. Jeon E, Park HA. Development of a smartphone application for clinical-guideline-based obesity management. Healthc Inform Res 2015; 21(1): 10–20.

94. Li LC, Adam PM, Townsend AF, Lacaille D, Yousefi C, Stacey D, Gromala D, Shaw CD, Tugwell P, Backman CL. Usability testing of ANSWER: a web-based methotrexate decision aid for patients with rheumatoid arthritis. BMC Med Inform Decis Mak 2013; 13: 131.

95. Joshi A, Amadi C, Trout K, Obaro S. Evaluation of an interactive surveillance system for monitoring acute bacterial infections in Nigeria. Perspect Health Inf Manag 2014; 11: 1f.

96. Albu M, Atack L, Srivastava I. Simulation and gaming to promote health education: Results of a usability test. Health Education Journal 2015; 74(2).

97. Daniels JP, King AD, Cochrane DD, Carr R, Shaw NT, Lim J, Ansermino JM. A human factors and survey methodology-based design of a web-based adverse event reporting system for families. Int J Med Inform 2010; 79(5): 339-348.

98. Lewis JR. IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. Int J Hum-Comput Int 1995; 7(1): 57–78.

99. Lewis JR. Psychometric evaluation of the post-study system usability questionnaire: The PSSUQ. Human Factors Society 36th Annual Meeting; Atlanta: Human Factors and Ergonomics Society; 1992; 1259-1263.

100. Chin JP, Diehl VA, Norman KL, editors. Development of an instrument measuring user satisfaction of the human-computer interface. SIGCHI conference on Human factors in computing systems; 1988.

101. Wilkinson A, Forbes A, Bloomfield J, Fincham Gee C. An exploration of four web-based open and flexible learning modules in post-registration nurse education. Int J Nurs Stud 2004; 41(4): 411-424.

102. Lewis JR. Psychometric evaluation of an after-scenario questionnaire for computer usability studies: The ASQ. ACM Sigchi Bulletin 1991; 23(1): 78–81.

103. Brooke J. SUS: A quick and dirty usability scale. Usability evaluation in industry 1996; 189(194): 4–7.

104. Lewis JR, Henry SC, Mack RL. Integrated office software benchmarks: A case study. In: Dapler D, editor. Interact 90 – 3rd IFIP International Conference on Human-Computer Interaction; Cambridge, UK: North-Holland; 1990; 337-243.

105. Lewis JR. Psychometric evaluation of the computer system usability questionnaire: The CSUQ (Tech. Report 54.723). Boca Raton, FL: International Business Machines Corporation; 1992.

106. Frytak J. Measurement. In: Kane R, editor. Understanding health care outcomes research. Gaithersburg, MD: Aspen Publications; 1997; 213-237.

107. Lewis JR, Sauro J. The Factor Structure of the System Usability Scale. In: Kurosu M, editor. Human Centered Design. Heidelberg: Springer; 2009; 94–103.

108. Streiner DL. Starting at the beginning: An introduction to coefficient alpha and internal consistency. J Pers Assess 2003; 80(1): 99–103.

109. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. Journal of Human–Computer Interaction 2008; 24(6): 574-594.

110. Bland JM, Altman DG. Statistics notes: Cronbach's alpha. BMJ 1997; 314: 572.

111. Streiner DL, Norman GR. Health measurement scales 2ed. New York: Oxford University Press; 1995.

112. Kass RA, Tinsley HEA. Factor analysis. J Leisure Res 1979; 11: 120–138.

113. Brooke J. SUS: a retrospective. Journal of usability studies 2013; 8(2): 29–40.

114. Yen PY, Wantland D, Bakken S. Development of a Customizable Health IT Usability Evaluation Scale. AMIA Annu Symp Proc 2010; 2010: 917-921.

115. Yen PY, Sousa KH, Bakken S. Examining construct and predictive validity of the Health-IT Usability Evaluation Scale: confirmatory factor analysis and structural equation modeling results. J Am Med Inform Assoc 2014;21(e2):e241–8.

116. Sousa VE, Matson J, Lopez KD. Questionnaire Adapting Little Changes Mean a Lot. Western J Nurs Res 2016: 0193945916678212.

117. Lin MI, Hong RH, Chang JH, Ke XM. Usage Position and Virtual Keyboard Design Affect Upper-Body Kinematics, Discomfort, and Usability during Prolonged Tablet Typing. PLoS One 2015; 10(12): e0143585.

118. Koppel R, Leonard CE, Localio AR, Cohen A, Auten R, Strom BL. Identifying and quantifying medication errors: evaluation of rapidly discontinued medication orders submitted to a computerized physician order entry system. J Am Med Inform Assoc 2008; 15(4): 461-465.

119. Ahmed A, Chandra S, Herasevich V, Gajic O, Pickering BW. The effect of two different electronic health record user interfaces on intensive care provider task load, errors of cognition, and performance. Crit Care Med 2011; 39(7): 1626-1634.

120. Duong M, Bertin K, Henry R, Singh D, Timmins N, Brooks D, Mathur S, Ellerton C. Developing a physiotherapy-specific preliminary clinical decision-making tool for oxygen titration: a modified delphi study. Physiother Can 2014 ;66(3): 286-295.

121. Vuong AM, Huber JC, Jr., Bolin JN, Ory MG, Moudouni DM, Helduser J, Begaye D, Bonner TJ, Forjuoh SN. Factors affecting acceptability and usability of technological approaches to diabetes self-management: a case study. Diabetes Technol Ther 2012; 14(12): 1178-1182.

122. Jokela T, Iivari N, Matero J, Karukka M, editors. The standard of user-centered design and the standard definition of usability: analyzing ISO 13407 against ISO 9241–11. The Latin American conference on Human-computer interaction; 2003; New York: ACM.

123. Anastasi A. Psychological testing. New York: Macmillan; 1988.

124. Parke CS. Missing data. In: Parke CS, editor. Essential first steps to data analysis: Scenario-based examples using SPSS. Thousand Oaks: Sage Publications; 2012; 179.

125. Gaber J. Face validity. In: Salkind NJ, editor. Encyclopedia of Research Design. Thousand Oaks: Sage Publications; 2010; 472–475.

126. Johnson EK, Nelson CP. Utility and Pitfalls in the Use of Administrative Databases for Outcomes Assessment. The Journal of Urology 2013; 190(1): 17-18.

127. Fratello J, Kapur TD, Chasan A. Measuring success: A guide to becoming an evidence-based practice. New York: Vera Institute of Justice; 2013.

128. Greenhalgh J, Long AF, Brettle AJ, Grant MJ. Reviewing and selecting outcome measures for use in routine practice. J Eval Clin Pract 1998; 4(4): 339-350.
129. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, Bakken S, Kaplan CP, Squiers L, Fabrizio C, Fernandez M. How we design feasibility studies. American Journal of Preventive Medicine 2009; 36(5): 452-457.
130. Madrigal D, McClain B. Strengths and weaknesses of quantitative and qualitative research. UXmatters; 2012; Available from: http://www.uxmatters.com/mt/archives/2012/09/strengths-and-weaknesses-of-quantitative-and-qualitative-research.php.