Curr Diab Rep (2014) 14:534
DOI 10.1007/s11892-014-0534-9

PSYCHOSOCIAL ASPECTS (KK HOOD, SECTION EDITOR)

Getting a Technology-Based Diabetes Intervention Ready for Prime Time: a Review of Usability Testing Studies

Courtney R. Lyles · Urmimala Sarkar · Chandra Y. Osborn

Published online: 31 August 2014
© Springer Science+Business Media New York 2014

Abstract Consumer health technologies can educate patients about diabetes and support their self-management, yet usability evidence is rarely published even though it determines patient engagement, optimal benefit of any intervention, and an understanding of generalizability. Therefore, we conducted a narrative review of peer-reviewed articles published from 2009 to 2013 that tested the usability of a web- or mobile-delivered system/application designed to educate and support patients with diabetes. Overall, the 23 papers included in our review used mixed (n=11), descriptive quantitative (n=9), and qualitative methods (n=3) to assess usability, such as documenting which features performed as intended and how patients rated their experiences. More sophisticated usability evaluations combined several complementary approaches to elucidate more aspects of functionality. Future work pertaining to the design and evaluation of technology-delivered diabetes education/support interventions should aim to standardize the usability testing processes and publish usability findings to inform interpretation of why an intervention succeeded or failed and for whom.

Keywords Usability testing · Technology · Diabetes · Review · User-centered design · Think aloud · Cognitive walkthrough · Heuristic evaluation · Intervention

Introduction

In recent years, there has been an explosion of self-management support systems and applications ("apps") developed for patients with diabetes. Widespread consumer adoption and use of computers, the Internet, and mobile devices has created an opportunity to leverage these platforms to augment patient care. Technologies can offer an extended reach to patients by delivering diabetes education and self-care support (e.g., medication adherence reminders, mobile or

This article is part of the Topical Collection on *Psychosocial Aspects*

C. R. Lyles · U. Sarkar
UCSF Center for Vulnerable Populations at San Francisco General Hospital, San Francisco, CA, USA

U. Sarkar
e-mail: usarkar@medsfgh.ucsf.edu

C. Y. Osborn
Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
e-mail: chandra.osborn@vanderbilt.edu

C. Y. Osborn
Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA

C. Y. Osborn
Diabetes Research & Training Center, Center for Diabetes Translational Research, Vanderbilt University Medical Center, Nashville, TN, USA

C. R. Lyles (✉) · U. Sarkar
Division of General Internal Medicine at San Francisco General Hospital, University of California San Francisco, 1001 Potrero Ave., SFGH 10, San Francisco, CA 94143, USA
e-mail: lylesc@medsfgh.ucsf.edu

C. Y. Osborn
Division of General Internal Medicine & Public Health, Vanderbilt University Medical Center, 1215 Twenty-First Ave South, Ste 6000, MCE - North Tower, Nashville, TN 37232-8300, USA

web applications for monitoring and improving diet/physical activity) or by facilitating patients' communication with providers or peers (e.g., patient web portals or asocial networking forums) in their everyday lives. While many healthcare systems and researchers assume technology-driven solutions (e.g., online portals for access to electronic health records, mobile phone applications) provide various benefits, including improved patient outcomes [1] and cost savings [1, 2], the effectiveness of these solutions has been mixed [3]. One explanation for a lack of effectiveness may be incomplete front-end formative research prior to evaluating the solutions' effect on outcomes.

There is a continuum of necessary steps to ensure a technology-based solution is meaningful, useful, and engaging to end users, resulting in maximum benefit from utilization. These steps include conducting user-centered design research during the development phase and employing methods of usability testing prior to evaluating a technology-based intervention's effect on patient outcomes. Researchers have been applying user-centered design principles [4] and to a lesser degree post-production usability testing. Borrowing from the fields of design/industrial engineering and human factors engineering, user-centered design involves understanding the needs, values, and abilities of users to improve the quality of users' interactions with and perceptions of the technology [5, 6]. Once a technology-delivered solution has been developed (usually in an iterative fashion with direct user input), usability testing validates the product to determine how it meets user needs, such as ease of use (e.g., presence of technical glitches), accessibility (e.g., developed with simple language and clear instructions), and functionality (e.g., being used as intended) for the target audience with the ultimate goal of improving the spread and sustainability of use (i.e., continued user engagement across various settings) [5]. It is this second phase of usability testing that we are focused on in this manuscript.

There are several usability testing methods that have been adopted by healthcare and health services researchers. These often include qualitative assessments (i.e., focus groups or individual interviews) to learn about users' experience and likeability of the system/application, system usage data that automatically captures objective use of the system/application, and usability questionnaires that gather users' self-reported use along with ratings of satisfaction, ease of use, and desire to use the system/application and/or recommend it to others [7]. Still used, but to a lesser degree, are think aloud interviews, cognitive walkthroughs, heuristic evaluations, and other methods [7]. The goal of our review was to characterize methods for testing usability in the recent diabetes literature, describe the lessons learned across studies, and draw conclusions about how usability testing assists with user engagement and ultimately allows for testing a technology-delivered intervention as intended.

Methods

Data Sources and Search Strategy

In January 2014, a published literature search was conducted of technology-delivered interventions for patients with either type 1 or type 2 diabetes. We generated three lists of relevant key words that included diabetes-related terms (e.g., diabetes, diabetic), technology-related terms (e.g., technology, Internet, mobile, website, patient portal), and usability testing methods/terms (e.g., usability, think aloud, heuristic evaluation, cognitive walkthrough). The lists were intralinked with "OR" and interlinked with "AND" so that all candidate articles contained at least one diabetes-related, technology-related, and usability testing-related term. Hand searching reference lists of articles yielded additional studies for review.

Our search relied on PubMed-indexed articles published from 2009 to 2013. Authors CYO and CRL performed literature searches and independently reviewed each title and abstract for potential relevance to the research questions. Articles included by either investigator underwent full-text screening. Authors CYO and CRL independently evaluated candidate full-text articles for inclusion and resolved discrepancies through review and discussion. A third, independent reviewer, author US, assisted with reconciling discrepancies as needed.

Study Eligibility

Studies were included in the present review if they satisfied all of the following criteria: (1) included a technology-supported intervention platform with a patient interface (e.g., Internet, cell phone, mobile device); (2) tested a patient-facing diabetes self-care monitoring, support, or educational or behavioral solution; (3) employed a usability testing method; (4) were published in a peer-reviewed journal from 2009 to 2013; and (5) were published in English. Articles were excluded if they focused only on the design phase of the intervention/technology (i.e., because we consider user-centered design to be a distinct and separate entity); were based on secondary analyses; were a meta-analysis, systematic review, or editorial/commentary; targeted nonpatients (e.g., physicians or other medical staff); or tested a medical device (e.g., new insulin pen) instead of a technology-based educational, self-management solution or tool for monitoring patients' behavior for self-care decision-making.

We initially identified 135 studies. After removing duplicate studies and conducting an initial title and abstract review to apply our inclusion and exclusion criteria outlined above, we identified 26 articles for full-text review. Additional hand review of these articles produced two more articles for consideration, for a total of 28 articles reviewed. Authors CYO and CRL excluded one of these articles because the technology-based solution was not designed specifically for people with diabetes, per se, and was therefore not tested with that user group [8]. These authors then excluded three more articles because the technology-based solution collected and transmitted patient data to providers (i.e., telemonitoring) without offering either a patient-facing component that allowed patients to track self-care or a close-looped system in which patient data was then used to educate or provide patients with self-care support [9–11]. Author US reconciled discrepancies, resulting in the removal of one additional article that tested the usability of an artificial pancreas built atop the IOS operating system that offered minimal self-care monitoring, education, or support [12]. Thus, a total of 23 articles were included in our narrative synthesis.

Narrative Synthesis

In an effort to bring together the broadest knowledge from a variety of usability testing study designs and methodologies, we analyzed studies using a narrative synthesis approach [13]. This approach is not to be confused with the narrative descriptions that accompany many reviews. Rather, a narrative synthesis is an attempt to systematize the process of analysis when a meta-analysis or a systematic review may not be appropriate because of the range of methodologies in the studies reviewed. Consistent with the narrative synthesis steps proposed by Popay et al. [13], we conducted (1) a preliminary analysis of the included studies, (2) an exploration of relationships between included studies, and (3) an assessment of the robustness of the synthesis. Due to the exploratory nature of our review, we omitted theory development from our synthesis process [13]. Preliminary synthesis consisted of extracting the descriptive characteristics of the studies in a table and producing a textual summary of the studies employing a specific usability testing design/methodology. When appropriate, a thematic analysis was used to extract themes from the cluster of studies within a given methodological domain. The subsequent narrative results represent the main areas of knowledge available about the types of usability testing methods employed with technology-based diabetes education and behavioral interventions from 2009 to 2013. We conclude with a reflection on our synthesis process and implications for technology-based intervention research in diabetes.

Results

Of the 23 articles included in the final sample (Table 1), over two thirds were published in the past 2 years (i.e., 2012 or 2013)—indicating an increase in this type of research. Eight of these articles tested web-delivered solutions, 12 tested mobile phone-delivered solutions, and 3 leveraged personal digital assistants (PDAs). Usability testing studies were largely conducted in the USA (n=11), followed by Norway (n=5), the Netherlands (n=3), and Canada (n=2). The remaining usability studies were conducted in Japan, Spain, and the UK. Thirteen studies included adult users with diabetes; six included children/adolescents with diabetes and/or their parents; two included expert reviewers/testers; and two included volunteer testers. Finally, of the studies including users with diabetes (n=19), 10 reported patient characteristics or inclusion criteria beyond age, gender, and diabetes type, duration, and/severity (e.g., HbA1c).

The included articles covered all usability methodologies mentioned above; 16 reported results from usability questionnaires, 13 captured system usage data, 12 used qualitative methods in the form of focus groups or individual interviews, 3 used think aloud, 1 used a cognitive walkthrough, and 1 used a heuristic evaluation. These categories were not mutually exclusive as many studies used a combination of different methodological approaches.

Usability Questionnaire

Over two thirds of the studies included in our review relied on user self-report to assess the usability of a technology-based diabetes self-management solution. Survey- or questionnaire-based scoring of usability solicits users' subjective opinion, perceived ease of use and helpfulness, and/or experiences with using a technology-based intervention or application. While studies utilizing this approach differed in terms of what they asked users, common user outcomes were an overall usability score of the system being studied, and satisfaction with and general ease of use of the system/application and/or its feature(s). Most of the 16 usability studies that included surveys relied on homegrown items that mapped onto the solution being tested (n=12) rather than reliable and validate instruments of usability (n=4). Many studies provided only a general discussion of their usability scoring, such as the specific questions asked with the sample mean score for each question. While several studies assessed usability only as users' experience (such as satisfaction or willingness to recommend the technology to someone else), others scored a wider variety of concepts related to functionality and users' acceptability of the system. For example, Britto et al. [14] reported that a patient portal with a diabetes chronic condition-specific page was rated highest for interface pleasantness and likeability, but rated lowest for error messages and clarity of information.

Qualitative Assessment

A qualitative assessment of usability identifies barriers to using a technology-based solution and reasons for nonuse or discontinued use more comprehensively than usability questionnaires. A qualitative assessment can detect unanticipated

Table 1 Usability studies of web- and mobile-based systems/applications to support diabetes self-management published from 2009 to 2013

Authors [reference]	Study population	Technology/platform	Usability testing method	Usability outcomes	Findings
Web-based solutions					
Britto et al. [14]	16 parents of children with diabetes, cystic fibrosis or juvenile idiopathic arthritis (mean age 39)—USA[a]	MyCare Connection: a patient portal with the ability to view personal health information and securely message providers with features customized for the chronic condition it serves in a disease-specific tab	• Iterative think aloud • System usage data • Usability questionnaire (CUSQ)	Task completion times; usability satisfaction	• Think aloud identified users' barriers to understanding medical jargon and the need for easier navigation • System data captured tasks requiring the most amount of time requiring the interpretation of medical information or performing several steps • Satisfaction was greatest for interface pleasantness and likeability and lowest for error messages and clarity of information
Gude et al. [37]	4 researchers—Netherlands	Patient Assisting Net-Based Diabetes Insulin Titration system (PANDIT): a web-based insulin self-titration support system for patients and caregivers	• Cognitive walkthrough	Expert assessments of user scenarios and usability problems	• Experts identified 81 unique usability problems: 37 with the patient interface, 43 with the caregiver interface, and 1 with both • 4 of these problems could impact safety
Nijland et al. [23••]	50 adults with type 2 diabetes—Netherlands[a]	Web-based system for online tracking, diabetes education, and secure emailing with nurses	• Qualitative assessment • System usage data • Think aloud • Usability questionnaire	Engagement with the technology and reasons for discontinued use	• 20 users identified 166 unique usability problems and 20 out of 50 users discontinued use by 24 months due to suboptimal user-friendliness, navigation, and interactivity that interfered with users' awareness of the system's utility
Ossebaard et al. [27••]	21 adults with a chronic health condition (7 with diabetes, aged 23–67)—Norway[a]	Patient portal with information for patient decision-making on a variety of chronic health topics	• Qualitative assessment • Think aloud • System usage data • Usability questionnaire	Interview themes; completion of tasks; self-reported judgments of usability	• 62 % of experiences in interviews were positive • 42 % of the think aloud expressions were positive • However, most users did not complete the assigned patient portal tasks in the allotted time • Major usability issues pertained to suboptimal navigation and content (i.e., users wanted personalized information)
Powell et al. [38]	17 adults with diabetes using insulin pumps (aged 22–70)—UK[a]	Virtual clinic with access to information about diabetes, access to health professionals, and the ability to interact with peers	• System usage data • Usability questionnaire	Virtual clinic page views; usability ratings	• Usage was highest in the first 2 months and declined over 6 months • Users rated the ability to interact with peers as the most important feature and gave favorable ratings of design, ease of use, and feature content
Simon et al. [18]	4 adults with type 2 diabetes who used daily insulin and 4 diabetes nurses—Netherlands	PANDIT: a web-based insulin self-titration support system for patients and caregivers	• System usage data • Think aloud	Number of usability barriers when completing 8 typical usage scenarios (Nielsen's severity rating and impact on safety [39])	• Users identified 17 unique usability problems, but none had a high risk of compromising patient safety • The most severe problems had to do with patients being unable to enter their data (e.g., fasting plasma glucose, hypoglycemic events)

Table 1 (continued)

Authors [reference]	Study population	Technology/platform	Usability testing method	Usability outcomes	Findings
Urowitz et al. [15]	17 adults with diabetes and 64 healthcare providers—Canada	Diabetes self-management portal	• Qualitative assessment	Themes on barriers and facilitators of portal use	• For patients, the portal was easy to navigate and user-friendly, but slow dial-up Internet and problems with data entry were barriers to use • Providers were generally dissatisfied with the portal's usability and discoverability (ease with which they could find features within the portal) • Suggested improvements were enhanced technical support and a tutorial to learn about the portal's available features and how to use them
Zayas-Caban et al. [40]	3 student researchers serving as pseudopatients (n=2) and a pseudonurse (n=1)—USA	Blood pressure and glucometer readings connected to Microsoft HealthVault for provider monitoring and advice	• Qualitative assessment (derived from scenario-based user testing)	Time spent on tasks and challenges encountered while completing tasks	• Users identified 57 unique usability problems (both cognitive and physical), with 49 being patient-related problems and 8 being nurse-related problems
Mobile-delivered intervention					
Arsand et al. [41]	12 adults with diabetes (aged 44–70)—Norway	Few Touch: a diabetes self-management mobile application (app)	• Qualitative assessment • System usage data • Usability questionnaire (SUS)	Qualitative themes; number of uploads; System Usability Scale	• Qualitative feedback identified technical problems • Glucose uploads over 6 months ranged from 23 to 564 (higher uploads at the beginning) • Users scored the app high on usability (84 out of 100) • Users liked seeing their glucose values as a historical trend and used the diet and activity features and reported them as useful
Cafazzo et al. [42]	20 adolescents (aged 12–16) in pilot evaluation—Canada	bant: a diabetes mobile app on the IOS platform with a wireless glucometer, rewards system for 3 or more blood glucose tests per day, social networking, and personal health record link	• Qualitative assessment • System usage data	Feedback from exit interviews; number of glucose uploads	• Satisfaction was high, with 88 % reporting they would continue to use the system poststudy • Among users with sufficient system data, glucose uploads increased from 2.4 to 3.6 per day over the 12-week study, and rewards were significantly related to blood glucose testing, and thus user engagement
Carroll et al. [43]	40 adolescents with type 1 diabetes (aged 13–18)—USA	Glucophone: an integrated cell phone and glucometer coupled with the ability to view logged data via the HealthPia website (for patients) and specially designed computer software program (for nurse practitioners providing regimen adjustments)	• Usability questionnaire	Ratings of usability and satisfaction with the system	• Patients reported positive feelings about the system even though 74 % sent their phone back to the company for servicing and 50 % reported the HealthPia website was not user-friendly
Chomutare et al. [44]	7 adults with diabetes (aged 46 to 70)—Norway[a]	Few Touch: a diabetes self-management mobile app coupled with a semi-integrated social forum	• Qualitative assessment	Qualitative feedback; System Usability Scale	• Over 12 weeks, only 3 users initiated social conversation via the app (i.e., social media use was low)

Table 1 (continued)

Authors [reference]	Study population	Technology/platform	Usability testing method	Usability outcomes	Findings
			• Usability questionnaire (SUS)		• However, users scored the app high on usability (84.6 out of 100)
Demidowich et al. [22]	2 reviewers evaluated 6 app features—USA	Evaluated 42 free and paid diabetes self-management apps available on the Android platform as of April 2011	• Heuristic evaluation	Score presence and usability of 6 features (i.e., blood glucose monitoring, medication tracking, insulin calculation, data graphing, data sharing, and other data tracking)	• The average usability score for an app was low (11.3 out of 30), and the mean usability score for a feature was average (3 out of 5)
DeShazo et al. [45]	8 adults with diabetes participated in usability testing (mean age 38 years)—USA	3 mobile phone games to increase nutrition estimation and food comparison skills	• System usage data • Usability questionnaire	Actual usage of games; user ratings of use, utility, and enjoyment	• User input captured debugging of the user interface and the need to improve the system's overall game play, scoring mechanism, and difficulty
Dick et al. [16]	18 adults with type 2 diabetes (mean age 55)—USA[a]	SMS system to support self-care	• Qualitative assessment • System usage data • Usability questionnaire	Engagement; satisfaction and usability score	• Users were not concerned about privacy, but wanted more control over the phrasing of the messages, and when and how often messages were sent • Users responded to 80 % of the 78 text messages requiring responses • 94 % of users reported high satisfaction with the SMS system and that it was easy to use
Froisland et al. [24•]	12 adolescents with type 1 diabetes (mean age 16.2)—Norway[a]	Diamob: a mobile phone-based diabetes diary to photo-document food choices and physical activity, and wirelessly upload blood glucose values Diabetes Messaging System: an SMS platform to educate patients and facilitate their communication with providers	• Qualitative assessment • Usability questionnaire (SUS)	Themes from interviews; system usability score	• Mean usability score was 73 out of 100 • Interviews identified technical problems and requests for additional functionality, but users liked seeing visual summaries of their blood glucose values, diet, physical activity, and insulin doses, and having increased access to their provider
Garcia-Saez et al. [46]	10 adults with type 1 diabetes and an insulin pump (aged 21–62)—Spain	PDA with wireless uploading of blood glucose, a logbook for entering diet and other information, and displaying of graphical/trend data; new medication alert	• System usage data • Usability questionnaire	Usage data (number of sessions, viewing of therapy changes, access to logbook, entering/monitoring values, downloading data); usability and utility scores; usage data	• On average, insulin downloads were 6.32/week and logbook views were 7.54/week • Users gave favorable ratings of usability (i.e., all agreed that the system was easy to learn how to use and become familiar with)
Mulvaney et al. [47]	23 adolescents with type 1 diabetes (aged 13–17)—USA[a]	SuperEgo: SMS system that sends messages to address barriers to self-care adherence and remind users to perform self-care	• Usability questionnaire	Ratings of usability and experiences, and engagement with the platform	• Users were satisfied with the system, rated it favorably for comprehension, and would recommend it to others
Osborn and Mulvaney [48]	20 adults with type 2 diabetes (mean age 52)—USA[a]	SuperEgo: SMS system that sends a daily 1-way text messages to address medication adherence barriers and a daily 2-way text message asking whether adherence was achieved coupled with an IVR system that delivers weekly automated calls to provide medication adherence feedback and ask	• Qualitative assessment • System usage data • Usability questionnaire	User engagement with the system, ratings of usefulness and helpfulness of the text messages and calls	• On average, users responded to 81 % of the text messages they received, rated the average helpfulness of the automated call high (8.3 out of 10), and reported the text messages addressing barriers to adherence were very helpful (8.3 out of 10)

Table 1 (continued)

Authors [reference]	Study population	Technology/platform	Usability testing method	Usability outcomes	Findings
		questions about successes and failures in the past week			
Padman et al. [49]	8 children with type 1 diabetes (aged 10–18) and their parents—USA	Personal Health Record for Quality of Life (PHRQL): mobile app on the Android platform that allows tracking of blood glucose values, food intake, physical activity, and insulin activity coupled with game like social networking	• System usage data • Usability questionnaire	Patterns of use, connectivity between users, and review of user comments within the app	• There was high variability in users' engagement with the app, but game mechanics and social media activity determined user engagement
Tani et al. [50]	20 volunteers (aged 20–69)—Japan[a]	PDA with input for blood glucose, BP, diet, exercise coupled with email feedback from providers	• Usability questionnaire	Usability score (5-grade evaluation: excellent, good, fair, poor)	• Users gave positive ratings for understandability and convenience
Tatara et al. [51]	11 adults with type 2 diabetes (mean age 57)—Norway	Few Touch: a diabetes self-management mobile app coupled with a physical activity component	• Qualitative assessment • System usage data • Usability questionnaire	Usability problems; user engagement with the app; ratings of usability (usefulness of the app, ease of recording information, and understanding of feedback screens)	• Focus group feedback revealed usability problems • 3 user groups identified: frequent use with positive experience, moderate use with neutral experience, and low use with negative experience—suggesting a correlation between perceived usefulness and actual usage of the app • Questionnaire ratings also revealed usability problems
Vuong et al. [52]	180 adults with type 2 diabetes in RCT (mean age in intervention group was 58)—USA[a]	PDA to monitor blood glucose, blood pressure, medication use, physical activity, and diet	• Qualitative assessment	Barriers to using the PDA program derived from an open-ended questionnaire	• Users reported discontinued PDA use because of frustration with the device and/or program due to data entry burden, data inaccuracies, and potential data loss due to a low battery

CUSQ Computer Usability Satisfaction Questionnaire, *SUS* System Usability Scale, *SMS* short message service, *IVR* interactive voice response, *PDA* personal data assistant

[a] Detailed information about the sample provided (beyond age, gender, and diabetes type/duration/severity)

usability challenges that may not be included in a quantitative usability measure. Nearly half (12 of 23, 52 %) of the included studies used qualitative methods, specifically focus groups and/or individual interviews. Data gathered pertained to users' perceptions of and/or personal experiences with an application's ease of use and accessibility, and their ideas for how it might be improved. For instance, Urowitz et al. [15] conducted a qualitative usability evaluation of a web-based diabetes portal. Major themes included users' suggested improvements by differentiating between the most used and well-liked features (e.g., the graphical display of the patient self-monitoring data and the ability to email with providers) and the least used features (e.g., the online library). Moreover, users generated new ideas about additions to the portal, such as an online tutorial to help future users become oriented with the portal's available features and how to use them.

System Usage Data

Usability data captured from an application or what is often referred to as system usage data can complement users' subjective quantitative or qualitative self-reports. By design, applications gather objective data on how often users log in, upload information (e.g., blood glucose values or dietary information), use/view specific features, or are responsive (e.g., replying to a text message). For example, Dick et al. [16] leveraged system-captured usability data to understand how low-income patients with diabetes used a text messaging program. Usability outcomes included the average length of time it took users to respond to each system-generated message, the average number of text messages users sent over the course of the study, and the content of those messages (e.g., medication adherence vs. blood glucose monitoring). The authors also provided a detailed account of who was approached to participate in the study (i.e., characteristics of potential participants, including their prior use of text messaging) and who ultimately participated in the study, allowing for inferences to be made about the generalizability of findings to other, comparable patient populations. To our knowledge, no studies cited agreed-upon standards for reporting system usage data.

Think Aloud (Cognitive Task Analysis)

A much smaller proportion of included studies (n=3) used the think aloud method to assess system/application usability. In a think aloud usability test, participants are asked to use a system/application while continuously thinking out loud—that is, simply verbalizing their thoughts as they move through the user interface at will, or when asked to complete a specific task [17]. This method facilitates understanding how users perceive each feature by observing how and in what order they interact with a system/application. For example, Simon et al. [18] used a think aloud approach to identify patients' barriers to using a web-based patient-guided insulin titration application. They observed patients interacting with the website in a clinic setting or in a more "experimental" location as well as in patients' homes and in doing so identified 17 unique usability concerns that needed to be addressed prior to evaluating the application's effect on patient outcomes.

Cognitive Walkthrough and Heuristic Evaluation

The final two usability testing methods typically involve expert-based as opposed to user-based system/application inspections [7]. A cognitive walkthrough involves evaluating first-time use (without formal training) of the system/application, in which an expert (1) sets a goal to be accomplished, (2) searches for available actions, (3) selects an action or task that seems most relevant, and (4) performs the action and evaluates the entire system/application for making progress toward the original goal [19]. Alternatively, a heuristic evaluation is a more task-free method in which pre-established "rules of thumb" are applied to using a system/application to uncover general issues unrelated to completing a specific action or task [20]. Examples of such heuristics include, but are not limited to, communicating in user's language, minimizing user memory load, consistency, providing user feedback, clearly marked exits and shortcuts, relevant error messages, and achieving expected criteria [21]. A specific example of the latter is from Demidowich et al. [22] who had two experts review 42 diabetes self-management mobile applications (apps) publicly available on the Android platform as of April 2011. Reviewers graded every app on whether it supported self-monitoring of blood glucose, diabetes medication tracking, prandial insulin calculation, data graphing, data sharing and other data tracking, and the usability of each of these features and found that less than 10 % of Android apps comprehensively supported diabetes self-management [22].

Combining Methodologies

Several included studies combined different usability testing methods to inform fixes and enhancements to the technology-based solution. They are highlighted here to gain a deeper understanding of how usability testing can provide insight into why and how patients choose to engage with a technology-delivered support system/application as part of their everyday lives.

The first example is from Nijland et al. [23••] who conducted a comprehensive usability testing study of their web-based application for diabetes management in the Netherlands. First, this study outlined the recruitment process for their application very clearly, including the number of patients who refused to participate (and their reasons for nonparticipation) as well as detailed information about the final sample who agreed to pilot the application (such as the sample's self-reported health status and level of education). When assessing participants' usage of the application over 2 years, the authors also provided specifics about user engagement, identifying subgroups of continuous users (with both high and low levels of activity) vs. those who discontinued use and specifying the use of each application feature (e.g., glucose uploads, emailing with providers) over time. Finally, the authors assessed qualitative data to understand barriers to both initial adoption and sustained use, such as the lack of interest in engaging with diabetes self-management and the desire for more interactivity within the application.

In a similar way, Froisland et al. [24•] combined methodologies to provide a more comprehensive view of the usability of their diabetes mobile phone application for adolescents with diabetes. For their qualitative assessments, they conducted in-depth interviews with participants about their experiences using the application. These interviews uncovered important themes about the graphical displays of personalized health information, the sense of increased access to providers, and technical glitches with the software. To complement these findings, they also collected and reported quantitative data on system usability. They used the system usability scale (SUS), which is a validated scale of 10 statements (e.g., "I found this system easy to use" and "I feel very confident in using the system") scored from strongly agree to strongly disagree, resulting in a total sum out of 100 [25, 26]. Froisland and colleagues reported a mean SUS of 73 out of 100 among their adolescent participants.

Finally, Ossebaard et al. [27••] combined qualitative assessments (both interviews and focus groups), think aloud, system usage data, and a usability questionnaire to gain a more complete picture of both the functionality of their technology as well as the patients' perceptions of its usability (i.e., observing them interacting with the system followed by in-depth reflections about their experiences). The authors report detailed scenarios of how users engaged with a diabetes web portal account, such as patterns for how users completed an assigned task and how long it took users to complete it. They also specified how different tasks led to different rates of completion and what proportion of the content verbalized during the think aloud was positive, neutral, and negative across categories of navigation, content, layout, acceptance, satisfaction, and empowerment. In the follow-up interviews and focus groups, the authors were then able to dive deeper into each of these categories to similarly assess the proportion of participants' positive and negative sentiments about using the portal.

Discussion

There is a limited, but growing, body of literature describing usability testing for technology-enabled diabetes self-management education and support. Most studies assessing usability do so with questionnaire items created for the study and to a lesser degree standardized usability instruments. While questionnaires are a potentially useful starting point for understanding users' perceptions of ease of use and satisfaction, users' perceptions may not provide a complete perspective of usability. Specifically, actual user behavior, as measured by system usage, and think aloud, in which users report their understanding/interpretation of each feature/step in use, may provide a more detailed account of users' behavior and understanding, respectively, than usability or satisfaction questionnaires. Qualitative studies provided additional insights by soliciting participant ideas to improve usability and identifying unanticipated usability challenges. Notably, both qualitative and think aloud methods identified more usability problems with a system/application and user concerns than usability questionnaires alone (Table 2).

Our findings are consistent with a recent review of provider-facing usability testing studies conducted by Yen and Bakken [28], such as the call for a more systematic approach to reporting users' experiences and the benefit of combining several complementary usability methodologies. Similarly, our findings are consistent with a broader review article by LeRouge and Wickramasinghe that found that virtually no user-centered design studies of diabetes technologies reported on methods used across the entire development spectrum, from planning/feasibility to postrelease [4]. Taken together, these suggest that an overarching framework for user engagement, from design through development to implementation and evaluation, is needed. Such a framework should include standard and validated metrics for distinct aspects of the user experience.

The research using usability questionnaires suggests a need to develop and implement standardized instruments to measure usability across a range of domains. Studies should report norms such that scores can be interpreted across different types of technology-delivered interventions. In particular, existing usability scales such as the SUS [25, 26] or Computer Usability Satisfaction Questionnaire (CUSQ) [29] may not provide sufficient context for understanding which aspects of a solution are functional and/or accessible and which are less useable. Alternatively, system usage data provides more objective feedback about usability, in that, regardless of perceptions, lack of use likely indicates poor usability. The difficulty lies in determining appropriate thresholds for what would constitute "good" or desired system usage. As an example, if a participant takes 3 days to respond to a text message, but responds to each and every message, it is unclear how to interpret usage vis-à-vis usability. Moreover, long-term usage of a technology—as opposed to a pilot trial that is only a few weeks long—provides more indication of sustained patient engagement; yet, it is difficult to conduct lengthy trials in order to report on ongoing use. Finally, expert

Table 2 Selected problems identified by think aloud/qualitative methods

Usability problem	Study	Description
Web-based interventions		
Medical jargon and terminology	Britto et al. [14]	Participants had trouble understanding terms such as "pathology report" and abbreviations such as "Fe"
Information overload	Britto et al. [14]	Participants reported that information provided on the portal was too detailed and complex
Poor user-friendliness of web application	Nijland et al. [23••]	Participants forgot to use the Web application due to lack of a reminder feature, were unaware of system features, and reported that the application could be more interactive
Redundancies in layout, functionality, and content	Ossebaard et al. [27••]	Participants reported double links and overlapping information
Difficulty entering data	Simon et al. [18]	Participants had trouble figuring out the correct way to enter data, such as not knowing to use a comma instead of a period as a decimal mark
Time consuming to access information	Urowitz et al. [15]	Providers reported that manual data entry made accessing patient information time consuming and expressed concern that using the portal would decrease time spent with patients
Mislabeling of site functions	Zayas-Caban et al. [40]	Users reported that website tab labels did not always correspond to their anticipated functionality
Mobile-delivered interventions		
Burden of data entry	Arsand et al. [41]	Participants found it tiring to record their nutritional intake on a daily basis
Fixed text message content and delivery	Dick et al. [16]	Participants reported wanting more control over the content and frequency of text message delivery
Cumbersome interface	Froisland et al. [24•]	Participants reported it being cumbersome to enter a code and log into an Internet-based SMS system on their mobile phones
Message content error	Osborn and Mulvaney [48]	Participants encountered system problems, including the receipt of text messages indicating to respond with "FalseYes" or "FalseNo"
Inability to edit data	Tatara et al. [51]	Participants reported not being able to edit records that had been saved by mistake
Risk of data loss	Vuong et al. [52]	Participants reported that data was lost when the mobile battery ran low
Openness to data error	Vuong et al. [52]	Due to a lack of methods to prevent erroneous data entry: participants reported that empty and out-of-range glucose values were entered into the interface

evaluation such as cognitive walkthrough and heuristic evaluation detected errors of omission, in that experts will be actively seeking relevant features and content that are needed for diabetes self-management—which patients themselves may not be aware of. Therefore, combining these diverse techniques can leverage the strengths and weaknesses of each method and provide more nuanced insight into why and how patients use a technology-delivered intervention, as evidenced in the highlighted studies in our review.

In almost half of the studies included in this review, the authors did not report on detailed characteristics of their participant population, such as educational attainment, health literacy, or self-reported technical knowledge. This can be problematic for disseminating these solutions to larger patient populations, especially if all or most of the participants involved in usability testing had high levels of technical savviness. It is unclear whether the usability of such technologies would then be generalizable to other patient populations. Furthermore, among the studies that did report some specific information about their participants, there is room for increasing the diversity of the participants involved in usability testing. For example, there is a need to understand how those with limited health literacy or lower technical skills interact with the technology—and methods to evaluate new technologies across health literacy levels exist [30, 31]. This is essential, as there is evidence of a digital divide by health literacy [32, 33] and by socioeconomic status in the USA, particularly among older Americans who are more likely to have type 2 diabetes [34]. There is also a growing literature to suggest that diabetes patients from racial/ethnic minority groups are less likely to engage with health technology, even among those who are already existing computer and Internet users [35, 36]. This may suggest that diverse patients need to be more involved in the early design and testing of technology to ensure that the content and format are relevant to broader populations—an area in which usability testing could be particularly impactful. Purposive recruitment by health literacy level and self-reported technical skills are initial strategies to increase the diverse of usability testing participants.

While we conducted a thorough literature review, we were unable to include papers not written in English. Because methods were dissimilar, we were unable to perform a systematic review of all the included articles.

Conclusions

We conclude that each usability testing method confers a distinct advantage. While all technology-delivered interventions should investigate and report on usability, it is critical to perform a robust usability evaluation using combination of methodologies in order to achieve a comprehensive assessment, which would ideally follow an iterative user-centered design process. We believe the field of user engagement, across all phases from design through evaluation, could benefit from a standardized approach and validated metrics for prespecified and distinct aspects of the user experience. We advocate for routine inclusion of open-ended inquiry and qualitative analysis to identify unanticipated user concerns.

Acknowledgments Dr. Lyles is supported by a Career Development Award from the Agency for Healthcare Research and Quality (K99 HS022408). Dr. Sarkar is also supported by the Agency for Healthcare Research and Quality (R21 HS021322). Dr. Osborn is supported by a Career Development Award from the National Institute of Diabetes and Digestive and Kidney Diseases (K01 DK087894). The contents of this manuscript are solely the responsibility of the authors and do not necessarily represent these funding entities. We thank Lina Tieu, MPH, for her assistance in reviewing and finalizing the manuscript.

Compliance with Ethics Guidelines

Conflict of Interest Courtney R. Lyles, Urmimala Sarkar, and Chandra Y. Osborn declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:
- • Of importance
- •• Of major importance

1. Bu D, Pan E, Walker J, Adler-Milstein J, Kendrick D, Hook JM, et al. Benefits of information technology-enabled diabetes management. Diabetes Care. 2007;30:1137–42.
2. Archer E, Groessl EJ, Sui X, McClain AC, Wilcox S, Hand GA, et al. An economic analysis of traditional and technology-based approaches to weight loss. Am J Prev Med. 2012;43:176–82.
3. Or CK, Tao D. Does the use of consumer health information technology improve outcomes in the patient self-management of diabetes? A meta-analysis and narrative review of randomized controlled trials. Int J Med Inform 2014.
4. LeRouge C, Wickramasinghe N. A review of user-centered design for diabetes-related consumer health informatics technologies. J Diabetes Sci Technol. 2013;7:1039–56.
5. Corry MD, Frick TW, Hansen L. User-centered design and usability testing of a Web site: an illustrative case study. Etr & D-Educ Tech Res. 1997;45:65–76.
6. Schleyer TK, Thyvalikakath TP, Hong J. What is user-centered design? J Am Dent Assoc. 2007;138:1081–2.
7. Jaspers MW. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. Int J Med Inform. 2009;78:340–53.
8. Napolitano MA, Hayes S, Russo G, Muresu D, Giordano A, Foster GD. Using avatars to model weight loss behaviors: participant attitudes and technology development. J Diabetes Sci Technol. 2013;7:1057–65.
9. Bus SA, Waaijman R, Nollet F. New monitoring technology to objectively assess adherence to prescribed footwear and assistive devices during ambulatory activity. Arch Phys Med Rehabil. 2012;93:2075–9.
10. Hazenberg CE, Bus SA, Kottink AI, Bouwmans CA, Schonbach-Spraul AM, van Baal SG. Telemedical home-monitoring of diabetic foot disease using photographic foot imaging—a feasibility study. J Telemed Telecare. 2012;18:32–6.
11. Morak J, Schwarz M, Hayn D, Schreier G. Feasibility of mHealth and Near Field Communication technology based medication adherence monitoring. Conf Proc IEEE Eng Med Biol Soc. 2012;2012:272–5.
12. Keith-Hynes P, Guerlain S, Mize B, Hughes-Karvetski C, Khan M, McElwee-Malloy M, et al. DiAs user interface: a patient-centric interface for mobile artificial pancreas systems. J Diabetes Sci Technol. 2013;7:1416–26.
13. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, Britten N, Roen K, Duffy S. Guidance on the conduct of narrative synthesis in systematic reviews. A Product from the ESRC Methods Programme 2006:1–92.
14. Britto MT, Jimison HB, Munafo JK, Wissman J, Rogers ML, Hersh W. Usability testing finds problems for novice users of pediatric portals. J Am Med Inform Assoc. 2009;16:660–9.
15. Urowitz S, Wiljer D, Dupak K, Kuehner Z, Leonard K, Lovrics E, et al. Improving diabetes management with a patient portal: a qualitative study of diabetes self-management portal. J Med Internet Res. 2012;14:e158.
16. Dick JJ, Nundy S, Solomon MC, Bishop KN, Chin MH, Peek ME. Feasibility and usability of a text message-based program for diabetes self-management in an urban African-American population. J Diabetes Sci Technol. 2011;5:1246–54.
17. Boren MT, Ramey J. Thinking aloud: reconciling theory and practice. IEEE Trans Prof Commun. 2000;43:261–78.
18. Simon AC, Holleman F, Gude WT, Hoekstra JB, Peute LW, Jaspers MW, et al. Safety and usability evaluation of a web-based insulin self-titration system for patients with type 2 diabetes mellitus. Artif Intell Med. 2013;59:23–31.
19. Wharton C, Rieman J, Lewis C, Polson P. The cognitive walkthrough method: a practitioner's guide. In: Usability Inspection Methods. New York: Wiley; 1994. p. 105–41.
20. Lewis C, Rieman J. Chapter 4. Evaluating the design without users. In Task-centered user interface design: a practical introduction. Online via shareware; 1993, 1994.
21. Nielsen J, Molich R. Heuristic evaluation of user interfaces. In Proc ACM CHI'90 Conf; April 1–5; Seattle, WA. 1990: 249–256.
22. Demidowich AP, Lu K, Tamler R, Bloomgarden Z. An evaluation of diabetes self-management applications for Android smartphones. J Telemed Telecare. 2012;18:235–8.
23.•• Nijland N, van Gemert-Pijnen JE, Kelders SM, Brandenburg BJ, Seydel ER. Factors influencing the use of a Web-based application for supporting the self-care of patients with type 2 diabetes: a longitudinal study. J Med Internet Res. 2011;13:e71. *This study used four usability approaches and collected data on reasons for discontinued use of a website. It only took 20 users to identify 166 unique usability problems, and users attributed a suboptimal user interface and limited interactivity as reasons for discontinued use.*
24.• Froisland DH, Arsand E, Skarderud F. Improving diabetes care for young people with type 1 diabetes through visual learning on mobile phones: mixed-methods study. J Med Internet Res. 2012;14:e111. *This study illustrates how to conduct a comprehensive qualitative and quantitative usability testing approach to identify technical problems, obtain users' ideas for improvements, and, in turn, inform further development and testing.*

25. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale (SUS). Int J Hum Comput Interact. 2008;24:574–94.
26. Brooke J. SUS: a “quick and dirty” usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland AL, editors. Usability evaluation in industry. London: Taylor and Francis; 1996.
27.•• Ossebaard HC, Seydel ER, van Gemert-Pijnen L. Online usability and patients with long-term conditions: a mixed-methods approach. Int J Med Inform. 2012;81:374–87. *This study employed four usability testing approaches, including software that recorded screen action during think aloud sessions. Utilization of a variety of methods uncovered negative user experiences, negative think aloud expressions, incomplete user tasks by most participants, and usability issues pertaining to navigation and non-personalized content that could be improved.*
28. Yen PY, Bakken S. Review of health information technology usability study methodologies. J Am Med Inform Assoc. 2012;19:413–22.
29. Lewis JR. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. In Human Factors Group. Boca Raton, FL: IBM Corporation; 1993.
30. Monkman H, Kushniruk A. A health literacy and usability heuristic evaluation of a mobile consumer health application. Stud Health Technol Inf. 2013;192:724–8.
31. Broderick J, Devine T, Langhans E, Lemerise AJ, Lier S, Harris L. Designing health literate mobile apps. Washington: Institute of Medicine; 2014.
32. Chakkalakal RJ, Kripalani S, Schlundt DG, Elasy TA, Osborn CY. Disparities in using technology to access health information: race versus health literacy. Diabetes Care. 2014;37:e53–4.
33. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, Lopez A, et al. The literacy divide: health literacy and the use of an internet-based patient portal in an integrated health system—results from The Diabetes Study of Northern California (DISTANCE). J Health Commun. 2010;15 Suppl 2:183–96.
34. Smith A. Older adults and technology use. Washington: Pew Research Center; 2014.
35. Lyles CR, Harris LT, Jordan L, Grothaus L, Wehnes L, Reid RJ, et al. Patient race/ethnicity and shared medical record use among diabetes patients. Med Care. 2012;50:434–40.
36. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, Lopez A, et al. Social disparities in internet patient portal use in diabetes: evidence that the digital divide extends beyond access. J Am Med Inform Assoc. 2011;18:318–21.
37. Gude WT, Simon AC, Peute LW, Holleman F, Hoekstra JB, Peek N, et al. Formative usability evaluation of a web-based insulin self-titration system: preliminary results. Stud Health Technol Inf. 2012;180:1209–11.
38. Jennings A, Powell J, Armstrong N, Sturt J, Dale J. A virtual clinic for diabetes self-management: pilot study. J Med Internet Res. 2009;11:e10.
39. Nielsen J, Mack RL, editors. Usability inspection methods. New York: Wiley; 1994.
40. Zayas-Caban T, Marquard JL, Radhakrishnan K, Duffey N, Evernden DL. Scenario-based user testing to guide consumer health informatics design. AMIA Annu Symp Proc. 2009;2009:719–23.
41. Arsand E, Tatara N, Ostengen G, Hartvigsen G. Mobile phone-based self-management tools for type 2 diabetes: the few touch application. J Diabetes Sci Technol. 2010;4:328–36.
42. Cafazzo JA, Casselman M, Hamming N, Katzman DK, Palmert MR. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res. 2012;14:e70.
43. Carroll AE, DiMeglio LA, Stein S, Marrero DG. Using a cell phone-based glucose monitoring system for adolescent diabetes management. Diabetes Educ. 2011;37:59–66.
44. Chomutare T, Tatara N, Arsand E, Hartvigsen G. Designing a diabetes mobile application with social network support. Stud Health Technol Inf. 2013;188:58–64.
45. DeShazo J, Harris L, Turner A, Pratt W. Designing and remotely testing mobile diabetes video games. J Telemed Telecare. 2010;16:378–82.
46. Garcia-Saez G, Hernando ME, Martinez-Sarriegui I, Rigla M, Torralba V, Brugues E, et al. Architecture of a wireless personal assistant for telemedical diabetes care. Int J Med Inform. 2009;78:391–403.
47. Mulvaney SA, Anders S, Smith AK, Pittel EJ, Johnson KB. A pilot test of a tailored mobile and web-based diabetes messaging system for adolescents. J Telemed Telecare. 2012;18:115–8.
48. Osborn CY, Mulvaney SA. Development and feasibility of a text messaging and interactive voice response intervention for low-income, diverse adults with type 2 diabetes mellitus. J Diabetes Sci Technol. 2013;7:612–22.
49. Padman R, Jaladi S, Kim S, Kumar S, Orbeta P, Rudolph K, et al. An evaluation framework and a pilot study of a mobile platform for diabetes self-management: insights from pediatric users. Stud Health Technol Inf. 2013;192:333–7.
50. Tani S, Marukami T, Matsuda A, Shindo A, Takemoto K, Inada H. Development of a health management support system for patients with diabetes mellitus at home. J Med Syst. 2010;34:223–8.
51. Tatara N, Arsand E, Bratteteig T, Hartvigsen G. Usage and perceptions of a mobile self-management application for people with type 2 diabetes: qualitative study of a five-month trial. Stud Health Technol Inf. 2013;192:127–31.
52. Vuong AM, Huber Jr JC, Bolin JN, Ory MG, Moudouni DM, Helduser J, et al. Factors affecting acceptability and usability of technological approaches to diabetes self-management: a case study. Diabetes Technol Ther. 2012;14:1178–82.