

Usability Testing of an Interactive Dashboard for Surgical Quality Improvement in a Large Congenital Heart Center

Danny T. Y. Wu^{1,2} Scott Vennemeyer¹ Kelly Brown³ Jason Revalee⁴ Paul Murdock¹
Sarah Salomone¹ Ashton France³ Katherine Clarke-Myers³ Samuel P. Hanke^{2,3}

¹ Department of Biomedical Informatics, University of Cincinnati, Cincinnati, Ohio, United States

² Department of Pediatrics, University of Cincinnati, Cincinnati, Ohio, United States

³ Heart Institute, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, United States

⁴ DAAP School of Design, University of Cincinnati, Cincinnati, Ohio, United States

Address for correspondence Danny T. Y. Wu, PhD, MSI, Department of Biomedical Informatics, University of Cincinnati, 231 Albert Sabin Way, Cincinnati, Ohio 45229, United States (e-mail: tzuyu@umich.edu).

Appl Clin Inform 2019;10:859–869.

Abstract

Background Interactive data visualization and dashboards can be an effective way to explore meaningful patterns in large clinical data sets and to inform quality improvement initiatives. However, these interactive dashboards may have usability issues that undermine their effectiveness. These usability issues can be attributed to mismatched mental models between the designers and the users. Unfortunately, very few evaluation studies in visual analytics have specifically examined such mismatches between these two groups.

Objectives We aimed to evaluate the usability of an interactive surgical dashboard and to seek opportunities for improvement. We also aimed to provide empirical evidence to demonstrate the mismatched mental models between the designers and the users of the dashboard.

Methods An interactive dashboard was developed in a large congenital heart center. This dashboard provides real-time, interactive access to clinical outcomes data for the surgical program. A mixed-method, two-phase study was conducted to collect user feedback. A group of designers ($N = 3$) and a purposeful sample of users ($N = 12$) were recruited. The qualitative data were analyzed thematically. The dashboards were compared using the System Usability Scale (SUS) and qualitative data.

Results The participating users gave an average SUS score of 82.9 on the new dashboard and 63.5 on the existing dashboard ($p = 0.006$). The participants achieved high task accuracy when using the new dashboard. The qualitative analysis revealed three opportunities for improvement. The data analysis and triangulation provided empirical evidence to the mismatched mental models.

Conclusion We conducted a mixed-method usability study on an interactive surgical dashboard and identified areas of improvements. Our study design can be an effective and efficient way to evaluate visual analytics systems in health care. We encourage researchers and practitioners to conduct user-centered evaluation and implement education plans to mitigate potential usability challenges and increase user satisfaction and adoption.

Keywords

- ▶ data visualization
- ▶ interface and usability
- ▶ data quality
- ▶ data analysis
- ▶ dashboard
- ▶ testing and evaluation
- ▶ human–computer interaction
- ▶ mixed (methodologies)

received
April 19, 2019
accepted after revision
September 2, 2019

© 2019 Georg Thieme Verlag KG
Stuttgart · New York

DOI <https://doi.org/10.1055/s-0039-1698466>.
ISSN 1869-0327.

Introduction

The widespread adoption of electronic health record (EHR) systems in the United States, driven by the HITECH Act, has led to a vast and diverse range of clinical data being captured and stored.^{1,2} Effective utilization of these clinical data, however, requires appropriate informatics solutions to address the high volume and complexity of the data,^{3,4} as well as to overcome potential barriers rising from data incompleteness and inaccuracy.^{5,6} Visual analytics can be an effective solution for identifying meaningful patterns in large clinical data sets.^{7,8} The evaluation of visualization dashboards has been a popular research topic in the health care domain in recent years.^{9–11} The U.S. Office of the National Coordinator for Health Information Technology has also developed a set of interactive dashboards that provide an overview of facts and statistics about Health IT topics, such as adoption rates and levels of interoperability.¹² Many studies have designed and evaluated visual analytics dashboards, especially those with high interactive and real-time response rates, for how they improve clinical operations and care quality.^{13–15}

Interactive dashboards are especially powerful since they allow users to manipulate the data along various dimensions and to determine the analytics focus. Several recent studies have developed visual analytics solutions that show timely and relevant clinical data, as well as the patterns found within them, which can be used for targeting patient and/or quality improvement interventions.^{16–18} These visual analytics solutions often require a significant amount of effort to be put toward data integration and information management,¹⁹ and their effectiveness on improving patient outcomes is unknown in the early stages of deployment and use.¹⁷ More importantly, these visual analytics solutions are subject to potential human-computer interaction (HCI) and usability issues resulting from the complicated interplay between human cognition and system functionality.²⁰ These HCI and usability issues can be detected through user-centered evaluations. Evaluation methods for visual analytics can vary depending on the analytics goals and the context. A recent work done by Wu et al systematically surveys the literature and summarizes evaluation methods of visual analytics in health information technology applications.²¹ This work offers four considerations for new studies wishing to evaluate visual analytics techniques, including (1) using commonly reported metrics, (2) varying experimental design by context, (3) increasing focus on interaction and workflow, and (4) adopting a phased evaluation strategy. Although these are important considerations for evaluation of usability, there was no explicit examination toward the root cause of usability issues in interactive dashboards, although some studies have suggested that many of the identified problems may be attributed to the difference between the mental models of the designers and the users.^{22,23}

A mental model represents how a designer or user perceives that a system should be used. The designers' and the users' mental models differ because these two groups of people usually do not communicate during the system development phase and only have contact through the system space, that is, through use of the system. This gap

can be bridged through user-centered design approaches, where designers and users are brought together in the early system development phase to communicate directly.^{24–26} User-centered design, however, can be time and resource consuming and is not always feasible for the development of interactive dashboards. This leads to inevitable differences in mental models that create usability issues and the need for comparing the mental models when examining underlying reasons for usability issues.

In this study, we had two objectives. Our first objective was to conduct usability testing of an interactive surgical dashboard that was designed to improve the visualization and interpretation of congenital heart surgical data compared with an existing static, spreadsheet-based dashboard. Our second aim was to provide empirical evidence of the potential mismatched mental models between the designers and the users of the dashboards and share lessons learned.

Methods

Clinical Setting

This study was conducted at a heart institute (HI) at a large, tertiary surgical and medical referral center that treats adults and children with congenital and acquired heart disease. The HI has strong information technology support and contributes its patient data routinely to the Society for Thoracic Surgery Congenital Heart Surgery Database. The medical center has adopted an institution-wide strategy to apply a commercial data analytics and visualization tool (MicroStrategy²⁷) on its clinical and operational data sets. This commercial tool provides a platform that allows users to integrate their own data and build interactive visualizations. The interactive surgical dashboard in this study is the first use case in the HI. This new dashboard was developed using MicroStrategy to replace an existing static, spreadsheet-based dashboard, and is meant to provide actionable information that meets clinical and administrative needs. **→Figs. 1 and 2** show the existing dashboard and the mockup of the new dashboard, respectively.

Note that the existing dashboard was labeled “static” primarily because of its limited ability to allow users to manipulate the data presented on screen. Additionally, the underlying data of the existing dashboard was refreshed less frequently. In other words, the new dashboard was updated with current information automatically each day; the existing dashboard was manually refreshed every quarter due to the labor intensity of the data update process. The new dashboard was developed to support the access and transparency of clinical outcomes data for the surgical program and to inform quality improvement initiatives. While the new dashboard was designed to improve upon the existing dashboard, there was no guarantee that it would meet this goal, especially considering the fact that the design decisions were made by a small number of domain experts. For this reason, we evaluated the usability of the new dashboard in a formative manner with a comparison to the existing dashboard to identify usability issues and seek opportunities for improvement.

Measures		FY16 Final	FY17 Final	FY18 Final	FY19 Q1	FY19 Q2	FY19 Goals/ Benchmarks	Reporting Frequency	Desired Direction
Patient and Employee Safety	Number of serious safety events	x	x	x	x	x	x	Q	↓
	Surgical Site Infection Rate - Class I/II operative procedures	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%	Q	↓
	Number of OSHA Employee Injuries (CTS Specific)	x	x	x	x	x	x	Q	↓
	% of OR "Time-Out"	xx.x%	xx.x%	xx.x%	xx.x%	xx.x%	—	Q	↑
	Total CPB Procedural Volumes	xxx	xxx	xxx	xx	xx	—	Q	N/A
	Total No-CPB Procedural Volumes	xxx	xxx	xxx	xx	xx	—	Q	N/A
	Total Surgical Procedural Volumes (CPB+No-CPB)	xxx	xxx	xxx	xx	xx	—	Q	N/A
	STAT 1 Cases	xx	xx	xx	x	x	—	Q	N/A
	STAT 2 Cases	xx	xx	xx	x	x	—	Q	N/A
	STAT 3 Cases	xx	xx	xx	x	x	—	Q	N/A
Outcomes and Clinical Excellence	STAT 4 Cases	xx	xx	xx	x	x	—	Q	N/A
	STAT 5 Cases	xx	xx	xx	x	x	—	Q	N/A
	MORALITY (Rolling 4 Quarters)								
	STAT 1 Cases	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%	Q	↓
	STAT 2 Cases	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%	Q	↓
	STAT 3 Cases	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%	Q	↓
	STAT 4 Cases	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%	Q	↓
	STAT 5 Cases	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%	Q	↓
	VSD Mortality	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%	Q	↓
	Norwood Mortality (hybrids not included)	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%	Q	↓
	TOF Mortality	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%	Q	↓
	ASO Mortality	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%	Q	↓
	ASO+VSD Mortality	x.x%	N/A	x.x%	x.x%	N/A	x.x%	Q	↓
	Glenn/HemiFontan	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%	Q	↓
	Truncus Mortality	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%	Q	↓
	CAVC Mortality	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%	Q	↓
	Of Bypass Coarctation repair	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%	Q	↓
	Total Surgical Mortality	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%	Q	↓
	Total Surgical Neonatal Mortality (<=30 days)	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%	Q	↓
	TRANSPLANT/HEART FAILURE (Rolling 4 Quarters)								
	Survival to Hospital Discharge s/p VAD/TAH	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%	Q	↑
	Cardiac Transplant Survival to Discharge	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%	Q	↑
	MORBIDITIES								
	% of Post-OP Hemorrhage Requiring Reoperation	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%	Q	↓
	% of Post-Cardiotomy ECMO < 14 days	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%	Q	↓
	% of Neurological Deficit at Discharge	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%	Q	↓
	% Paralyzed Diaphragm	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%	Q	↓
	% of Post-op Chylothorax	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%	Q	↓
	% of Wound Infection (STS definition)	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%	Q	↓
	% Acute Renal Failure	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%	Q	↓
	% of Complete Heart Block requiring permanent pacemaker	x.x%	x.x%	x.x%	x.x%	x.x%	x.x%	Q	↓

Fig. 1 The existing, spreadsheet-based surgical dashboard of heart institute (HI). This dashboard was generated manually and quarterly with color coding to indicate the performance of safety and outcomes: green (at or better than the goal), yellow (within 25% of the goal), red (> 25% from the goal), and white (no data or no benchmark). The actual surgical data were masked and the color coding is randomly assigned in this figure.

Participant Recruitment

The participants of the study fall into two categories: dashboard designers and dashboard users. For the designers, because there were only three people involved in the dashboard design, all three participated in this study. These three designers included an attending pediatric cardiologist, who provided cardiologists' and administrators' viewpoints; a surgical registry manager with a nursing background, who provided surgeons' and the surgical team's viewpoints; and a quality improvement specialist with clinical data analyst skills. The three designers created the prototype of the new dashboard before the research team was formed, and were included in the research team when the user interviews were conducted. For the users, clinicians with the following four clinical roles were selected since they were current or potentially immediate users of the surgical dashboard. These roles were (1) cardiologists, (2) surgeons, (3) physician assistants, and (4) perfusionists. Two to four participants per clinical role were recruited on a voluntary basis, resulting in 12 participating users. The recruitment used convenience sampling through the professional network of the research team. Each participating user was invited to a 30-minute usability testing session due to their limited avail-

ability. We believe the sample participants (users) did not have any bias in the preference of the dashboards since the new dashboard had not been in its official use and the existing dashboard was not frequently used by many of them. We compared the data collected from the designers and the data collected from the users to demonstrate the mismatched mental models between these two groups. Note that the designer interviews ($N = 3$) were separate from the usability testing interviews ($N = 12$). In other words, the designers' opinions were only extracted from the semistructured interviews but their preference between the dashboards was not considered in the usability testing since they were likely to have favorable opinions toward the new dashboard.

Task Development

The tasks of the usability testing were designed by domain experts as a set of two clinical scenarios and were presented to the participating users during the usability testing sessions. To create these tasks, the members of the research team discussed and developed the description of each task together. The task descriptions were carefully worded to mimic realistic information needs from the participating

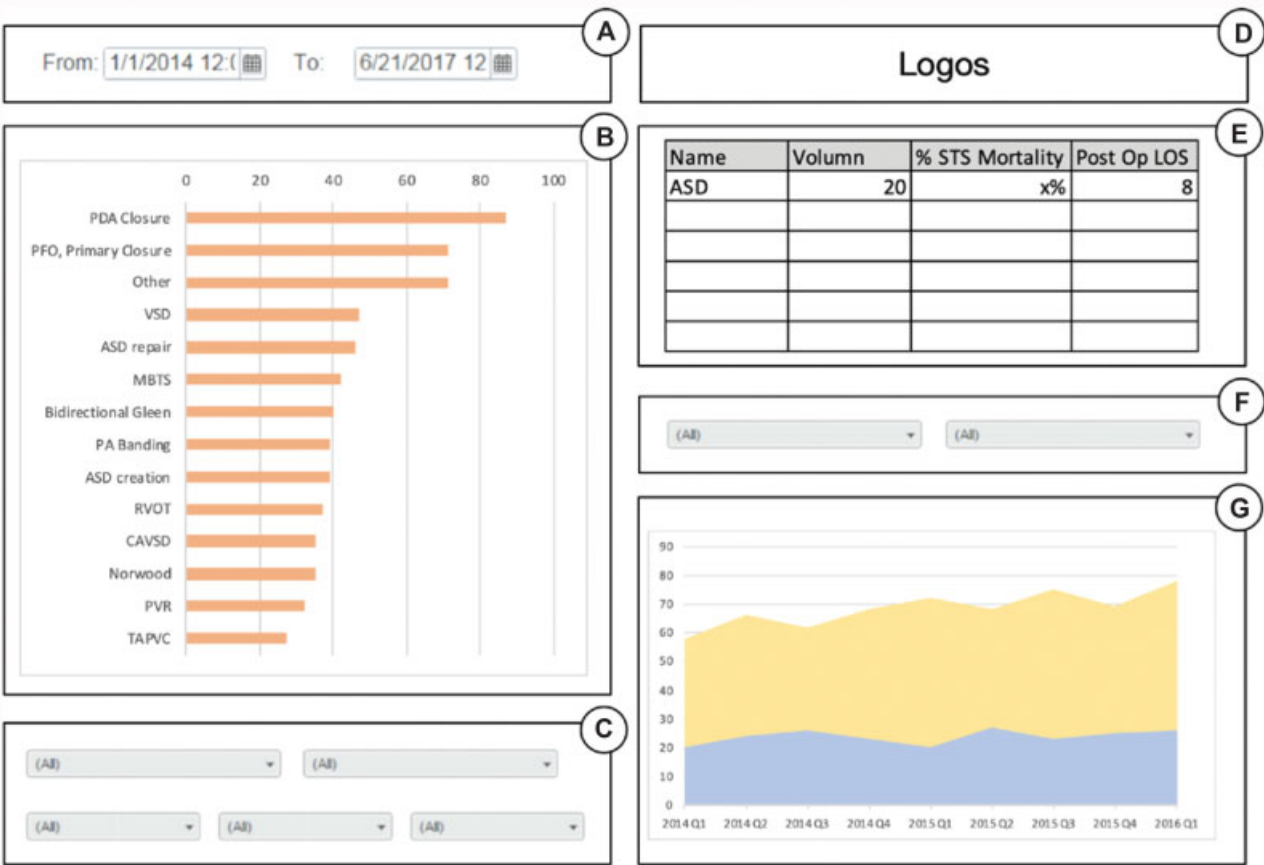


Fig. 2 The mockup of the new interactive surgical dashboard of heart institute (HI): (A) two filters for data ranges; (B) a horizontal bar chart showing procedure volumes by name; (C) a set of dropdowns for drilling down the data, including benchmark and primary procedures, surgeons, STAT scores, and 30-day status; (D) logos of the institution and the Society of Thoracic Surgeons (STS); (E) a table showing the percentage of STS mortality rate and length of stay (LOS) by procedure name; (F) two dropdowns for drilling down the data, including age ranges and operation types; and (G) a line area chart showing procedure volumes by quarters of fiscal years.

users. The first task was designed in favor of the existing static report and the second task was designed in favor of the new interactive dashboard. The descriptions of these two tasks are listed below:

Task 1: “You have been asked to lead a discussion regarding local STAT²⁸ Mortality rates for FY17 (7/1/16–6/30/17). You are in need of these outcomes for each of the five STAT Categories (STAT 1; STAT 2; STAT 3; STAT 4 STAT 5). How would you find these data?”

Task 2: “You are consulting a new mother whose baby was recently diagnosed with Tetralogy of Fallot. She is nervous and wants more information on our center’s experience with this repair. Specifically, she wants to know how many Tetralogy of Fallot repairs (Standard benchmark Tet type) our center has done in the last 2 FY (7/1/15–6/31/17), what the mortality rate is for this procedure and how long they can expect to be in the hospital during this procedure. How would you find this information for this new mom?”

These two tasks covered 80% of the functionality of the existing dashboard since the existing dashboard provided a simple tabular view to navigate the outcome of a measure with color coding to indicate the performance. They covered 40% of the functionality of the new dashboard since only the first two of the five tabs of the new dashboard were tested for

a fair comparison. The other three tabs of the new dashboard provide additional functions that the existing dashboard did not have. Each task was also designed to favor a specific dashboard. This choice was made for two reasons. First, we wanted to compare the dashboards in a fair manner since the existing dashboard is less complex and limited in functionality. Second, we wanted to collect the most useful user feedback, for example, the pros and cons of each dashboard, within a time constraint (a 30-minute session).

Study Conduction

This study employed a mixed-method design to evaluate the usability of the dashboards. The study design included a set of semistructured interviews and a usability testing method using a think-aloud protocol and a questionnaire containing the System Usability Scale (SUS). This study design has been used by multiple usability studies in the areas of clinical and health informatics.^{29–31} The SUS has also been used to evaluate interactive visualizations.^{32,33} Since the study focused on user feedback rather than expert feedback on the dashboards, heuristic evaluation³⁴ was not considered. The study was conducted in two phases. **Fig. 3** illustrates the two-phase study design. In the first phase, the participating designers were interviewed in a semistructured manner to understand

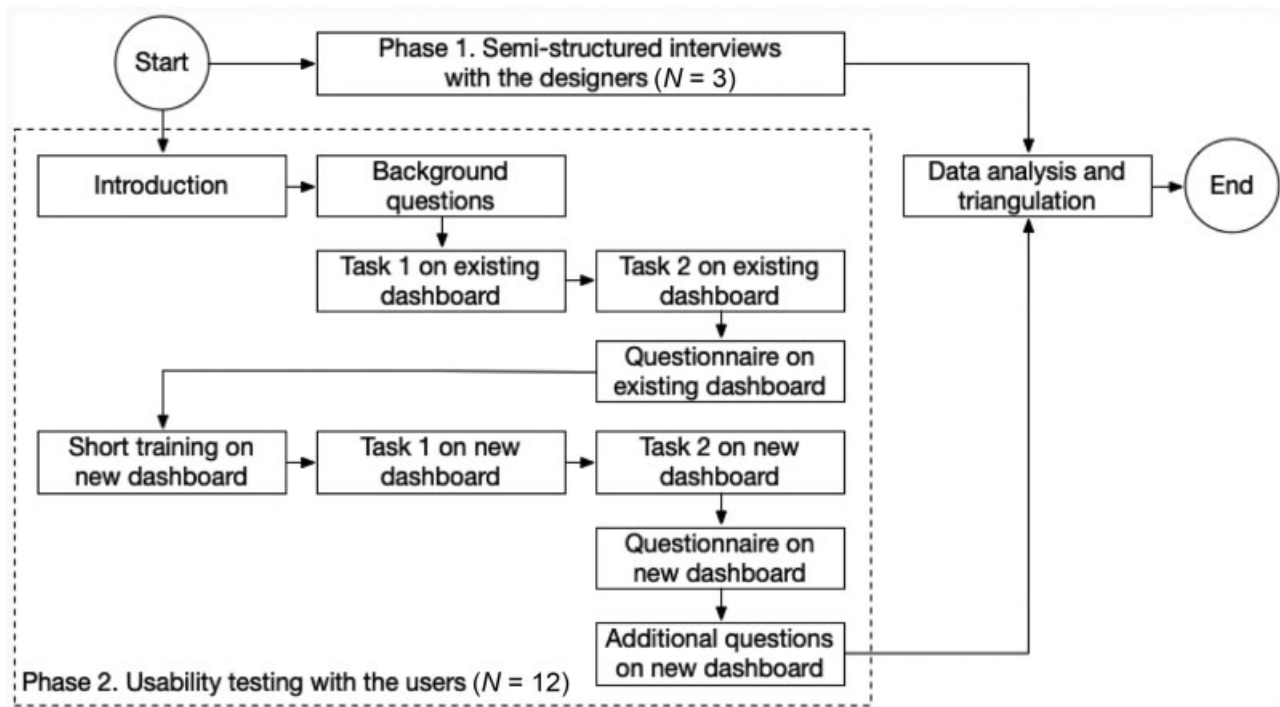


Fig. 3 The two-phase study design with detailed workflow of the usability testing (phase 2). Note that the evaluation order of the dashboards (i.e., existing dashboard first vs. new dashboard first) was switched in half of the users to reduce potential learning effect. This is not reflected in the figure.

the design principles of the new interactive dashboard. Each semistructured interview lasted 30 minutes with predefined questions, such as questions about job responsibility and dashboard design ideas, and follow-up questions based on the response from the designers.

In the second phase, the participating users individually took part in a usability testing session, where they were provided with a laptop with access to the existing (►Fig. 1) and new dashboards (►Fig. 2). The participants were introduced to the functionality of the new dashboard when they first saw it. The participants were also given a paper-based questionnaire to fill out during the session to collect more detailed feedback. Order randomization was used to reduce potential learning effect in the experiment. Half of the participants were asked to use the new dashboard first, while the other half were asked to use the existing dashboard first. After the participating users completed both tasks, each participant was asked to switch to the other dashboard and complete the same tasks again. Upon completing a task, the participant wrote down the responses (i.e., numbers or percentages) to the tasks and received comments from the researchers about the accuracy of their responses.

After completing the two tasks using each dashboard, a structured questionnaire was administered to collect the participant's feedback. This structured questionnaire was adopted and revised from the SUS, which is a standard and validated tool for usability testing of systems.³⁵ As prescribed by the SUS, each participant responded to the same set of 10 questions on a 5-point Likert scale as listed in ►Table 1. Following the scoring procedure, the scores of all questions were turned into a composite score for each participating user.

Table 1 Questionnaire used to assess the usability of the dashboards

Question ID (QID)	Description
1	I think that I would like to use this system frequently
2	I found <i>this</i> system unnecessarily complex
3	I thought this system was easy to use
4	I think that I would need the support of a technical person to be able to use this system
5	I found the various functions in this system were well integrated
6	I thought there was too much inconsistency in this system
7	I would imagine that most people would learn to use this system very quickly
8	I found <i>this</i> system very cumbersome to use
9	I felt very confident using <i>this</i> system
10	I needed to learn a lot of things before I could get going on this system

Note: Answered on a 5-point Likert scale. Slight modifications in questions 2, 8, 9, and 10 (in *italics*). The original questions can be seen at Usability.gov.

This composite score, ranging between 0 and 100, can sometimes be misleading because it is not a percentile score. Based on the literature, any score above 70 is considered "passing" and has "good" usability, and anything below this mark is considered "not passing" and is of "poor usability."^{36–38}

In addition to audio recording during the usability testing sessions, researchers that were present in the sessions documented observation notes, which reflected the user's observed behaviors, experience, and feedback toward the dashboards. After completing the tasks using the new interactive dashboard, each participating user was asked questions for improvement opportunities. These questions were not asked when using the existing static dashboard. The study protocol was reviewed by the University of Cincinnati Institutional Review Board and determined as "non-human subject" research.

Data Analysis

The data collected during the first phase from the semistructured interviews of the designers were transcribed verbatim and analyzed qualitatively using thematic analysis.³⁹ The analysis of data with participating users was handled differently. The background information collected about each user was summarized in a table with participants placed in each row and corresponding background information placed in the columns. Each participant had two SUS scores: one for the existing static report and one for the new interactive dashboard. The difference of the average of the two scores was compared using a two-tailed pairwise *t*-test. Similarly, the average raw score of each SUS question between the existing and the new dashboard were compared. The task accuracy was calculated based on the proportion of participants that answered all questions correctly for each given task. The audio recordings and the observation notes were reviewed and informative sentences were extracted. These feedback data were organized in a spreadsheet to summarize the pros and cons of both the existing and the new dashboard. The findings were triangulated to determine the performance of the new dashboard and to identify areas needing improvement. The data from the designers and the users were also compared to identify gaps in their mental models.

Results

Participant Characteristics

Three designers and 12 users were recruited. The 12 users included 4 cardiologists (C01–04), 2 surgeons (S01–02), 3 perfusionists (P01–03), and 3 physician assistants (A01–03). Most of the participating users are male (75%) with ages ranging between 30 and 40 (50%) and intermediate computer expertise (83%). Detailed background information about each participant is not revealed here due to the small number of surgeons in the study site to protect the identity of the participants. Each participant was assigned a letter and number corresponding to their position and order interviewed in the study.

All participating designers played a significant role during the design of the new interactive dashboard. Specifically, the registry manager had the most frequent access to the existing static dashboard due to the job responsibility in the surgical team. This person, therefore, explained the definition of the data elements in the existing static report to other designers and provided viewpoints of how the new dashboard can be used by surgeons. The clinical data analyst was trained to program the dashboard on the commercial tool

Table 2 Average SUS score by participant

Participant ID	SUS composite scores	
	Existing	New
P01	72.5	90.0
P02	70.0	97.5
P03	77.5	82.5
C01	40.0	92.5
C02	62.5	77.5
C03	55.0	90.0
C04	62.5	75.0
A01	85.0	90.0
A02	32.5	87.5
A03	72.5	82.5
S01	52.5	62.5
S02	80.0	67.5
Average	63.5	82.9 ^a

Abbreviation: SUS, System Usability Scale.

^aA significantly higher average score of the new dashboard ($p = 0.006$).

(MicroStrategy²⁷) and to connect necessary data elements from the HI data repository to the new dashboard. The cardiologist suggested design ideas and provided use cases of the new dashboard based on his knowledge and clinical experience, with the goal of improving care quality, data transparency, and patient satisfaction.

Preference toward the New Dashboard

→ **Table 2** lists the average SUS score of each participant. Compared with a passing score of 70, the new interactive dashboard performed significantly better with a score of 82.9. The new dashboard was scored significantly higher than the static dashboard ($p = 0.006$). The analysis of the user feedback provided several possible reasons why the new interactive dashboard had higher system usability, one of which was easy navigation. For example, A03 stated that they "really like how [the new system] was easy to navigate, visually it makes sense and flows." Another benefit was the user was able to perform drill-down analysis: "I could narrow the results. Everything was there. And what I mean by that is that some of the data that I was asked to look for previously wasn't even available on our current dashboard (P01)." The majority felt there was a learning curve that made it difficult to use initially, but that the new interactive dashboard was "easy to use after a couple of minutes of orientation or self-instruction (A03)."

In terms of task accuracy, the participants performed better using the new dashboard. Specifically, 10 out of 12 participants (83.3%) answered all the questions correctly in Task 1 and all of them (100%) got the right answers in Task 2 using the new dashboard. On the other hand, 8 out of 12 participants (66.7%) answered all the questions correctly in Task 1 and one of them (8.3%) missed the questions in Task 2 for the existing dashboard.

Table 3 Average scores by questions

Question ID (QID)	SUS raw score (Mean)			Pairwise t-test
	Existing	New	Difference	p-Value
1	2.4	4.1	1.7	0.014 ^a
2	2.6	1.2	-1.4	0.007 ^a
3	3.5	4.4	0.9	0.025 ^a
4	1.6	1.8	0.2	1.000
5	2.7	4.2	1.5	0.005 ^a
6	2.3	1.4	-0.9	0.103
7	4.0	4.8	0.8	0.151
8	2.2	1.2	-1.0	0.034 ^a
9	3.7	4.1	0.4	0.236
10	1.4	1.9	0.5	0.391

Abbreviation: SUS, System Usability Scale.

^aA significant difference of raw score ($p < 0.05$).

A further examination of the average raw SUS scores per questions (→ **Table 3**) shows that the new dashboard was not ranked significantly different from the existing dashboard in some questions due to the simple design and limited functionality of the existing dashboard. The participants did not feel the need to have technical support when using both dashboards (QID: 4, existing: 1.6, new: 1.8). C02 said that the existing static dashboard was “easy to use for the data that is included in it” but that it was still lacking in quantity of useful information. The overall consensus in the qualitative data of the usability testing was that although the existing dashboard is “straightforward” (A01) and “not complex” (C03). However, the existing dashboard was also “not detailed, [and] not very helpful (C03)” and the users “would need more data to interpret it (A01).” Both dashboards were considered at the same level in their consistency (QID: 6) and easiness to learn (QID: 7). The participants felt confident to use both dashboard (QID: 9) without too much learning before using them (QID: 10).

Opportunities to Improve the New Dashboard

The analysis of the user feedback identified three main opportunities to improve the new interactive dashboard. First, some of the participants had trouble with the calendar input feature, despite it initially appearing easy to use. Currently, the participants are not able to easily enter the date to retrieve information they are looking for. Instead, they have to page through the calendar input to find the desired date, which is time consuming and prone to error. During the interview of S01, he asked “Is there a faster way to click this? Do you just have to keep going until you get there?”

Second, several participants described the graph output feature as complicated and said that they required additional help to determine the usefulness of the information presented. While the participants mentioned the benefits of the visuals introduced with this new dashboard, they were not able to easily interpret the information presented and draw actionable conclusions accordingly. This was described by

A01 saying “[Using the graph] was pretty easy to do, but you have to kind of get used to the actual interface...It would be hard for someone not educated or familiar with this stuff to figure it out.” Furthermore, some individuals mentioned not understanding the terminology and acronyms presented along with the visuals produced, and even complained about the small font size (S01). Therefore, improving the readability and interpretability of the visuals in the new dashboard is a necessary step to improve the usability of the new dashboard and further increase user adoption.

Third, users had some difficulty in filtering information on the new dashboard to perform drill-down analysis. The filters were in multiple regions of the new dashboard, meaning that identifying a specific filter and its relationship with the visual information that it affects requires some learning and has been shown to be challenging. One participant suggested that combining all of the filters together in a single place (panel) would make working through the dashboard more convenient: “The first thing I did, I went to the overall range and found the set numbers, but at the bottom where you could select individually the benchmark procedure, I didn’t realize that that was going to affect everything in here [in visuals]” (P01). In addition, the new dashboard did not provide feedback about whether or not a filter had been changed. If a user clicks on various filters while working with the dashboard, they may not notice the changes in the visuals and may not be able to reset one filter to its default value. Since drill-down analysis is the key intended benefit of the new dashboard, addressing these filtering difficulties is required to help users navigate through the visuals and identify actionable information.

Gap between the Users and the Designers

Comparing data from the designers and the users helped us identify the gaps between the design of the system and its use.²² Through thematic analysis of the designers’ interview data, we found that the new dashboard was designed to address three major problems of the existing dashboard: (1) the lack of real-time data availability and interactivity, (2) inability to provide external reporting, and (3) disagreements in the time range for fiscal years. The new dashboard, therefore, provided corresponding features that were intended to fix these problems. However, these features also introduced some of the usability issues brought up during the study.

To address the first problem of the existing dashboard, the design team wanted to make sure that the users were able to use “real time data that [they] could use to inform outcomes at time of care (D01)” and to avoid manual calculation that “leads to inaccuracies (D01).” However, several users have reported that the readability of the figures and tables may be improved and the graph output feature was complicated. This complexity of the new dashboard was not reflected in the SUS scores (QID: 2, 4, 7). The filters also added complexity when performing the drill-down analysis and failed to provide system feedback after a user action. This means that users may get lost and be unable to remember which settings they had changed when they tried to reset the dashboard. The filters were indicated as a barrier to the learnability of the new dashboard, which may contribute to

the low level of confidence experiences by some while using the new dashboard (QID: 9).

To address the third problem of the existing dashboard, the new dashboard allowed the users to enter the date range and present the data accordingly since “the fiscal year number might be different than what [the user thinks] the fiscal year number is (D03).” Some users, however, could not find a way to quickly input the desired date using the calendar input component or even could not remember the definition of a fiscal year. Based on the data analysis and triangulation, we found that usability issues correspond closely with and can be attributed to the designed features. The mismatch between the mental models, in this case, was less about the failure of the system design to address known user problems, but more about the inability of the designers to anticipate user behaviors. This created system features that introduced “unexpected and unintended” usability issues.

Discussion

Key Findings

In this study, we conducted usability testing to compare the performance of an existing static and a new interactive dashboard in a large congenital heart center. The results show that the new interactive dashboard was preferred due to the higher usability and the potential to improve care quality from the users' perspective. Specifically, through two tasks designed to mimic clinical usage of these dashboards, the participating users gave significantly higher SUS scores to the new interactive dashboard than the existing static dashboard and indicated that the system's high usability may come from the easy navigation and ability to perform a drill-down analysis. It is worth noting that while the existing static dashboard was scored below the passing mark, some participating users thought the static interface had a simple tabular design that allowed quicker information identification, which can be useful in some circumstances. The new dashboard, on the other hand, had the features that introduced usability issues although these features were designed to fix the user problems identified by the design team in the existing dashboard.

Lessons Learned

Here we share four lessons learned. First, we conducted a two-phase, mixed-method study to assess the usability of an interactive data visualization. The designer feedback was collected through semistructured interviews while the user feedback was collected through a think-aloud protocol and a questionnaire that combined SUS with other structured questions and a few open-ended questions. This study design and its process and techniques have shown to be effective in collecting detailed data and be time-efficient to complete the usability testing in a 30-minute session. Such mixed-method design may be a necessary approach to conduct usability testing with clinicians since their time is prioritized for patient care.

Second, our study shows that the opportunities for improvement of an interactive visualization can be attributed to the original design ideas and principles, which provides empirical

evidence of the mismatched mental models between the designers and the users. Specifically, the designers intended to develop features that provide real-time data exploration, drill-down analysis, and date filtering to address the problems of the existing static report. While these features did provide a solution to the problem, they introduced several usability issues and therefore created opportunities for improvement. This shows the importance of conducting user-centered evaluation that identifies these gaps before they become a problem. These usability issues must be carefully addressed before the new dashboard is launched. We believe well-planned education sessions are necessary, especially when user-centered design is not employed. These well-developed education plans and user training prior to the official launch of a newly developed dashboard could help to address the usability issues by bridging the gaps between the users' and the designers' mental models. Doing so would increase the users' familiarity with the new dashboard, speed up their learning curve, and further increase their adoption of and satisfaction toward the dashboard.

Third, we demonstrated an alternative, yet effective, process to design and develop an interactive visualization in a clinical setting. The new dashboard was designed by a small group of experts that represented the viewpoints from the key stakeholders, including the administrators, the cardiologists, the surgeons, and the data analytics. Then, the design was evaluated through a mixed-method usability testing by a larger group of users including physician assistants and perfusionists in addition to cardiologists and surgeons. The new dashboard will be refined based on the user feedback collected from the present study. User training and education sessions will be conducted to mitigate the mismatched mental models. We believe this design and evaluation approach is more feasible to develop interactive visualizations in a clinical setting since clinicians have limited time and may not be able to support user-centered and codesign activities.

Finally, it is worth noting the cost and benefit of using a commercial tool to develop interactive visualizations. The main benefit is that commercial tools are an easy-to-use solution that gives users control of their own data to increase data transparency and reduce tool development time. On the other hand, the cost in our study includes data analyst's time to identify and pull related data sets, a certified programmer's time to set up and test the interactive visualization on a centralized server, and creating necessary training to introduce the new dashboard to the study participants (users). The license fee was covered by the institution. Moreover, the ability to modify features to address user feedback may be limited to the technical ability of the commercial tool. For example, the bars in a bar chart were colored automatically and each color cannot be assigned; the system feedback of user actions (e.g., popup and color change) may have few options. Note that our initial design choices in the usability testing were not constrained very much by the component of the commercial tool. However, we found that the user feedback collected from the study may not be easily addressable for this reason.

Limitations

This study has a few limitations. First, the interviews were conducted in a single institution and strictly time-limited (30 minutes) due to the heavy clinical duties and busy schedules of the participants. Short interview time may be unavoidable when conducting user studies in the medical field, making it nearly impossible to test more scenarios for comparison in our study. We addressed this limitation, however, by capturing user feedback in multiple ways (mixed-method) to increase the richness of the data and put significant effort toward data triangulation. Second, since only two scenarios were tested (due to the time constraint mentioned above), we were unable to reflect the full range of users' information needs and to identify potential usability issues when operating the dashboards in response to these needs. We designed the tasks, however, in a realistic manner, carefully wording the task descriptions to narrow the user's behaviors to those of our interest. Third, users may have remembered their previous answers to the tasks which could help them find the answer more quickly when they complete the tasks again using a different dashboard. We addressed this potential bias by creating two subgroups that had opposite testing orders and randomly assigned users based on their clinical roles to these subgroups. In this way, the learning bias can affect equally on the existing and the new dashboard and can therefore be controlled. Finally, the usability testing in the present study is a formative evaluation with the goal of identifying usability issues during the development process, rather than a summative evaluation to benchmark measurable performance of a near-complete product. Such formative evaluations have known methodological limitations in their reliability and validity.^{40,41} We addressed these limitations by employing a mixed-method design, increasing the numbers and the types of users to collect different viewpoints, designing tasks to support fair comparison in a timely manner, and switching the evaluation order of dashboards to reduce potential learning effects in the usability testing.

Future Work

Our future work includes revising the interactive dashboard based on user feedback, working with the quality improvement team in the HI to develop and roll out the education plan, monitoring the user adoption through surveys, and redesigning the interactive dashboard to expand its scope in patient consultation and research study support such as cohort identification. We will continue adopting user-centered evaluation in phases to drive the improvement of the interactive surgical dashboard.

Conclusion

We designed and conducted a mixed-method usability study on an existing and a new interactive dashboard for surgical quality improvement. The findings highlighted the usability issues of the interactive surgical dashboard and suggest well-planned education sessions to guide the users before the new dashboard's official use. The stakeholders in the

development and implementation of a new dashboard should include key and extended users, administrators, and technical support group. We encourage researchers and practitioners to conduct user-centered evaluation with realistic tasks and standardized metrics for meaningful comparison and implement education plans to address the mismatched mental models and increase user adoption, which can unleash the full potential of dashboards and visual analytics tools to better utilize and maximize the value of clinical data.

Clinical Relevance Statement

Interactive data visualization and dashboards can be a powerful tool to explore meaningful patterns in large clinical data sets but are subject to human-computer interaction and usability issues. The findings of the present study highlight the importance of conducting user-centered evaluation on interactive dashboards as well as creating education plans to improve usability and increase user satisfaction and adoption of interactive dashboards.

Multiple Choice Questions

1. When improving the usability of interactive visualization and dashboards, which of the following is not effective:
 - a. User-centered design.
 - b. User-centered evaluation.
 - c. User training and education.
 - d. Design only by domain experts.

Correct Answer: The correct answer is option d. Designing an interactive visualization dashboard only by using domain experts creates mismatched mental models between the designers and the users. This contributes greatly to usability issues. User-centered design can help improve the usability of interactive dashboards by engaging users in the early system development stage to bridge the mismatched mental models. It can also help detecting the usability issues for improvement; user training and education can help improve the usability by providing necessary training to the users prior to the official use to shift their mental models to be closer to the designers' mental model.

2. When conducting usability testing with clinicians, time is often a constraint. Which of the following may not effectively evaluate usability in a shorter usability testing session (e.g., 30 minutes)?
 - a. Using a single method to collect user feedback.
 - b. Using mixed methods to collect user feedback.
 - c. Increasing the number and the types of participants.
 - d. Designing realistic and meaningful tasks.

Correct Answer: The correct answer is option a. Due to the time constraint, using multiple and mixed methods can collect richer user feedback than using a single method and will allow data triangulation. A larger number and various types of participants can provide different

viewpoints and may identify more hidden usability issues. Designing realistic and meaningful tasks as opposed to hypothetical tasks can trigger the actual usage behaviors of the clinicians and may identify more significant usability issues.

Protection of Human and Animal Subjects

The study protocol was reviewed by the University of Cincinnati Institutional Review Board (IRB) and determined as “non-human subject” research. All the research data were deidentified.

Conflict of Interest

None declared.

Acknowledgments

We thank Ms. Keyin Jin for her input to the study design and questionnaire development. We also thank Ms. Shwetha Bindu for commenting on the manuscript.

References

- Adler-Milstein J, Jha AK. HITECH Act drove large gains in hospital electronic health record adoption. *Health Aff (Millwood)* 2017;36(08):1416–1422
- Non-federal Acute Care Hospital Health IT Adoption and Use. Available at: <https://dashboard.healthit.gov/apps/hospital-health-it-adoption.php>. Accessed April 15, 2019
- Luo G. MLBCD: a machine learning tool for big clinical data. *Health Inf Sci Syst* 2015;3:3
- Malykh VL, Rudetskiy SV. Approaches to medical decision-making based on big clinical data. *J Healthc Eng* 2018;2018:3917659
- Kahn MG, Eliason BB, Bathurst J. Quantifying clinical data quality using relative gold standards. *AMIA Annu Symp Proc AMIA Symp* 2010;2010:356–360
- Zhang Y, Sun W, Gutchell EM, et al. QAIT: a quality assurance issue tracking tool to facilitate the improvement of clinical data quality. *Comput Methods Programs Biomed* 2013;109(01):86–91
- Caban JJ, Gotz D. Visual analytics in healthcare—opportunities and research challenges. *J Am Med Inform Assoc* 2015;22(02):260–262
- Gotz D, Borland D. Data-driven healthcare: challenges and opportunities for interactive visualization. *IEEE Comput Graph Appl* 2016;36(03):90–96
- Bauer DT, Guerlain S, Brown PJ. The design and evaluation of a graphical display for laboratory data. *J Am Med Inform Assoc* 2010;17(04):416–424
- Plaisant C, Wu J, Hettinger AZ, Powsner S, Shneiderman B. Novel user interface design for medication reconciliation: an evaluation of Twinlist. *J Am Med Inform Assoc* 2015;22(02):340–349
- Mandalika VBH, Chernoglazov AI, Billingham M, et al. A hybrid 2D/3D user interface for radiological diagnosis. *J Digit Imaging* 2018;31(01):56–73
- Quick Stats. These data visualizations of key data and statistics provide quick access to the latest facts and figures about health IT. Published February 6, 2019. Available at: <https://dashboard.healthit.gov/quickstats/quickstats.php>. Accessed April 15, 2019
- Huber TC, Krishnaraj A, Monaghan D, Gaskin CM. Developing an interactive data visualization tool to assess the impact of decision support on clinical operations. *J Digit Imaging* 2018;31(05):640–645
- Romero-Brufau S, Kostandy P, Maass KL, et al. Development of data integration and visualization tools for the Department of Radiology to display operational and strategic metrics. *AMIA Annu Symp Proc AMIA Symp* 2018;2018:942–951
- Fletcher GS, Aaronson BA, White AA, Julka R. Effect of a real-time electronic dashboard on a rapid response system. *J Med Syst* 2017;42(01):5
- Radhakrishnan K, Monsen KA, Bae S-H, Zhang W. Clinical Relevance for Quality Improvement. Visual analytics for pattern discovery in home care. *Appl Clin Inform* 2016;7(03):711–730
- Simpao AF, Ahumada LM, Larru Martinez B, et al. Design and implementation of a visual analytics electronic antibiogram within an electronic health record system at a tertiary pediatric hospital. *Appl Clin Inform* 2018;9(01):37–45
- Hester G, Lang T, Madsen L, Tambyraja R, Zenker P. Timely data for targeted quality improvement interventions: use of a visual analytics dashboard for bronchiolitis. *Appl Clin Inform* 2019;10(01):168–174
- Karami M, Safdari R. From information management to information visualization: development of radiology dashboards. *Appl Clin Inform* 2016;7(02):308–329
- Federici S, Borsci S. Usability evaluation: models, methods, and applications. *International encyclopedia of rehabilitation. Center for International rehabilitation research information and exchange (CIRRIE)*; 2010:1–17
- Wu DTY, Chen AT, Manning JD, et al. Evaluating visual analytics for health informatics applications: a systematic review from the American Medical Informatics Association Visual Analytics Working Group Task Force on Evaluation. *J Am Med Inform Assoc* 2019;26(04):314–323
- Norman DA. *The Design of Everyday Things*. Revised and expanded edition. New York, New York: Basic Books; 2013
- Mental Models. Available at: <https://www.nngroup.com/articles/mental-models/>. Accessed April 15, 2019
- Kinzie MB, Cohn WF, Julian MF, Knaus WA. A user-centered model for web site design: needs assessment, user interface design, and rapid prototyping. *J Am Med Inform Assoc* 2002;9(04):320–330
- Kurzahls K, Hlawatsch M, Seeger C, Weiskopf D. Visual analytics for mobile eye tracking. *IEEE Trans Vis Comput Graph* 2017;23(01):301–310
- Ploderer B, Brown R, Seng LSD, Lazzarini PA, van Netten JJ. Promoting self-care of diabetic foot ulcers through a mobile phone app: user-centered design and evaluation. *JMIR Diabetes* 2018;3(04):e10105
- Anoshin D, Rana H, Ma N. *Mastering Business Intelligence with MicroStrategy Build World-Class Enterprise Business Intelligence Solutions with MicroStrategy 10*. Birmingham, UK: Packt Publishing; 2016
- The Meaning Behind STAT Scores & Categories. Available at: <https://www.luriechildrens.org/en/specialties-conditions/heart-center/volumes-outcomes/the-meaning-behind-stat-scores-categories/>. Accessed August 5, 2019
- van Lieshout R, Pisters MF, Vanwanseele B, de Bie RA, Wouters EJ, Stukstette MJ. Biofeedback in partial weight bearing: usability of two different devices from a patient's and physical therapist's perspective. *PLoS One* 2016;11(10):e0165199
- Hamm J, Money AG, Atwal A, Ghinea G. Mobile three-dimensional visualisation technologies for clinician-led fall prevention assessments. *Health Informatics J* 2019;25(03):788–810
- Nitsch M, Adamcik T, Kuso S, Zeiler M, Waldherr K. Usability and engagement evaluation of an unguided online program for promoting a healthy lifestyle and reducing the risk for eating disorders and obesity in the school setting. *Nutrients* 2019;11(04):E713
- Hirschmann J, Sedlmayr B, Zierk J, et al. Evaluation of an interactive visualization tool for the interpretation of pediatric laboratory test results. *Stud Health Technol Inform* 2017;243:207–211
- He X, Zhang R, Rizvi R, et al. ALOHA: developing an interactive graph-based visualization for dietary supplement knowledge graph through user-centered design. *BMC Med Inform Decis Mak* 2019;19(Suppl 4):150

- 34 Dowding D, Merrill JA. The development of heuristics for evaluation of dashboard visualizations. *Appl Clin Inform* 2018;9(03): 511–518
- 35 Brooke J. System Usability Scale (SUS): A Quick-And-Dirty Method of System Evaluation User Information. Read UK Digit Equip Co Ltd.; 1986:43
- 36 Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. *J Usability Stud* 2009;4(03):114–123
- 37 Coe AM, Ueng W, Vargas JM, et al. Usability testing of a web-based decision aid for breast cancer risk assessment among multi-ethnic women. *AMIA Annu Symp Proc AMIA Symp* 2016;2016: 411–420
- 38 Georgsson M, Staggers N. Quantifying usability: an evaluation of a diabetes mHealth system on effectiveness, efficiency, and satisfaction metrics with associated user characteristics. *J Am Med Inform Assoc* 2016;23(01):5–11
- 39 Maguire M, Delahunt B. Doing a thematic analysis: a practical, step-by-step guide for learning and teaching scholars. *AISHE-J Irel J Teach Learn High Educ* 2017;9(3). Available at: <http://ojs.aishe.org/index.php/aishe-j/article/view/335>. Accessed September 15, 2019
- 40 Wenger MJ, Spyridakis JH. The relevance of reliability and validity to usability testing. *IEEE Trans Prof Commun* 1989;32(04):265–271
- 41 Kessner M, Wood J, Dillon RF, West RL. On the reliability of usability testing. In: *CHI '01 Extended Abstracts on Human Factors in Computing Systems - CHI '01*. Seattle, Washington: ACM Press; 2001:97