Usability evaluation of obstetrics and gynecology information system using cognitive walkthrough method

Azadeh Kamel Ghalibaf[1], Majid Jangi[1], Mohammad Reza Mazaheri Habibi[1], Sara Zangouei[2], Reza Khajouei[3,4]

[1] Ph.D. Candidate, Department of Medical Informatics, Faculty of Medicine, Students Research Committee, Mashhad University of Medical Sciences, Mashhad, Iran
[2] M.Sc. of Computer Engineering, Azad University of Mashhad, Mashhad, Iran
[3] Ph.D. in Medical Informatics, Medical Informatics Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran
[4] Ph.D. in Medical Informatics, Department of Health Information Sciences, Faculty of Management and Medical Information Sciences, Kerman University of Medical Sciences, Kerman, Iran

Type of article: Original

Abstract

Background and aim: Obstetrics and gynecology information system is a critical component of the HIS in social security organization health centers. The objective of this study was to evaluate the usability of this system using the cognitive walkthrough method. Also, the present study provided a detailed formal description of how the cognitive usability evaluation can be applied and reported for a health care information system.

Methods: This study was conducted at the Mashhad University of Medical Sciences' usability lab from March 2016 to June 2017. A two-phase approach was used to conduct the cognitive walkthrough evaluation: preparatory and evaluation. The preparation was done in three stages: first, we investigated users' capabilities and background knowledge through a semi-structured interview. Second, the evaluation scenario was developed based on the most common tasks in routine workflow of users. Finally, each task was broken down into sequences of actions. In the evaluation phase, three usability experts independently assessed each action using a four-item checklist. Problems were categorized thematically and were reported from three different perspectives: Question-based, Task-based, and Evaluator-based. The data were then analyzed to understand the contribution of each task, along with its mean severity score.

Results: Evaluators' responses were compared and any conflict was resolved in an expert panel. A total of 116 usability problems were identified based on the consensus of the evaluators. Inadequate system feedback was found to be the main source of 43% of the problems, and resulted in users confusion.

Conclusion: Since the system was evaluated in its pilot implementation phase, there was an opportunity to prevent future potential usability problems. The use of a mixed quantitative and qualitative approach in this usability study provided a more comprehensive perspective of the system problems. This study provided a detailed description of conducting CW usability evaluation which can be used as a practical guide for future studies.

Key words: Health information systems, Usability testing, Cognitive walkthrough, Obstetrics and gynecology

1. Introduction

Prevention of medical errors has become a high priority since the release of the Institute of Medicine (IOM) report in 1999 (1). Health information systems (HISs) were introduced to reduce medical errors; however, there is a growing body of evidence that reveals that the design-related problems in HIS can lead health professionals to make new types of medical errors (2, 3). It seems that it is necessary to evaluate any HIS prior to its public use, in order to ensure its adaptation with users' needs and abilities. Usability testing is an evaluation technique that is concerned generally with human-computer interaction and how easy it is for users to carry out their tasks effectively,

efficiently, and enjoyably (4). There are different usability evaluation methods, this study has focused on Cognitive Walkthrough (CW) which is a task-based, expert-centered, analytical usability evaluation method that tries to identify problems through simulating end-users' cognitive abilities (5). Although CW provides a detailed description of system problems, due to its laborious, tedious form filling and time-consuming execution, researchers barely show any interest in using this method (6). According to the Joint Commission, failures in teamwork and communication are among the leading causes of adverse obstetric events, accounting for over 70% of sentinel events (7). Given the critical situation of this department, the present study tends to apply CW on an obstetrics and gynecology information system to facilitate users with different levels of expertise and experience in doing their tasks correctly. The HIS evaluated in this study belongs to the Social Security Organization (SSO) health centers which is the second most visited health care service provider in Iran, covering over 40 million people (8, 9). These centers were undergoing a switching transition from the traditional MS DOS_ based HIS to a new more comprehensive, integrated and interoperable HIS at the time of this study. The new system was first pilot implemented in five hospitals across the country. One of the first pilot centers was the 150-bed mission, Shahrivar 17th Hospital in Mashhad, Iran, where the present study was conducted. In order to investigate the new system's feasibility and acceptability, and ensure its ability to address the users' expectations, a usability evaluation of the system was considered necessary. Since the majority of the SSOs' visitors are from the working class in the society and the report of several studies that there is an association between low socioeconomic status and high rate of birth (10), the present study focused on the obstetrics and gynecology sub-system as the most crowded and challenging. To the best of our knowledge on the evaluation of obstetrics and gynecology information systems using CW usability method no study has been conducted. The present study also tends to provide a detailed formal description of how to conduct the cognitive usability evaluation in the field of healthcare information systems. The insufficient description of this method in existing literature has made the field vague and challenging.

2. Materials and Methods

This study was conducted at the Mashhad University of Medical Sciences' usability lab from March 2016 to June 2017. A two-phase approach was used to conduct the CW evaluation: preparatory and execution. In the preparatory phase, evaluators decided upon the evaluation plan components: the scenario, tasks, action sequences, and target users. Thereafter, the plan was executed by performing each action and checking out questions. The collected data were then analyzed to identify possible problems when interacting with the system. In this section, the methodology used in each phase is explained in details.

2.1. Phase I: Preparation

This phase consists of three stages:

2.1.1. Stage 1: Understanding the target users

Users in the obstetrics and gynecology ward have different levels of knowledge and skills working with the system. To understand and analyze users' abilities and skills, a semi-structured interview was conducted with the hospital IT manager who has access to the system log files and a good knowledge of the end-users' abilities. The IT manager is also responsible for planning and conducting training classes for end-user employees and supporting them through their problems working with the system. Thus all of these make the IT manager a credible source of information. We first categorized the users into three groups based on their organizational roles as nurses, secretaries and medical technicians (Table 1). Each user group was associated with a pre-defined set of duties and permissions in the system, requiring a specific level of knowledge and skills. To estimate the users' computer literacy, we examined the data on computer training classes that each user had participated, such as: course's level of advancement, number of training hours, and course outlines. These data were analyzed and the criteria from (11) were used to assign the levels of proficiency to each user group as: beginner, intermediate, and advanced.

Table 1. Users' characteristics

Role	No. users	Education	Experience (years)	Training (hours)
Nurse	6	Bachelor of nursing	1-20	> 60
Secretary	3	High school diploma	1-12	>120
Medical technician	2	Associates in Nursing	5-10	> 60

2.1.2. Stage 2: Defining a Scenario

The second step in implementing CW is to determine a procedure and describe it as a scenario (Box 1). Since there is no guide for selecting the procedure, we decided to consider the most representative and frequently used one. The

scenario was developed by the hospital IT manager due to her knowledge of the workflow and common interactions (Table 2).

Box 1. A routine scenario in the obstetrics and gynecology ward.

"First, a request should be sent to the operation room for a patient who is currently hospitalized in the maternity ward. Time scheduling is done by the users at the operation room. The information required for patient admission is to be filled out. This information includes: the required equipment, type of surgery and its parameters, as well as the type of anesthesia. As soon as the operation is completed, the patient will be discharged from the surgery ward and transferred to the relevant ward."

Table 2. Task list in the scenario

No.	Task Description	No. of actions	Role
1	Retrieve patient's record	4	Ward nurse
2	Send request to operation room	7	Ward nurse
3	Time planning of request	8	O.R[*] secretary
4	Select required equipment for surgery	7	O.R secretary
5	Fill out required data on anesthesia	5	O.R secretary
6	Set the operation parameters	5	O.R secretary
7	Discharge patient from operation room	2	O.R secretary

* O.R= Operating Room

2.1.3. Stage 3: *Determining action sequences*

The proposed scenario was broken down into a series of tasks, indicating steps needed to accomplish the scenario. As shown in Table 2, each task corresponds to a user group, and procedure completion requires information transfer among different units and different users. Each task was also broken down into sequences of actions. Actions are the smallest executable units that lead a task forward to achieve the goal. The granularity of tasks and the level of details to describe them are based on the user's expertise and knowledge. Table 3 shows a list of the actions needed to perform task 1 from Table 2 as an example (actions can be taken by either the user or the system).

Table 3. Sequence of actions for 'Retrieve patient's record' task

No.	Description	Type
1	Click on the 'hospitalization ward' in tree view	User action
2	Click on the 'maternity unit' under 'hospitalization wards'	User action
3	Double click on the name of target patient in the list	User action
4	Display patient's information form and hospitalization details	System response

2.2. Phase 2: Evaluation

During the evaluation phase, the interaction between the user and the interface is analyzed in depth for each required action identified in the preparation phase.

2.2.1. Measurement criteria

In this phase, the ease with which users can perform a task with little or no formal instruction had been assessed. Each action was performed by the evaluators, and the system's responses were inspected by answering the questions related to the user's cognitive model. These four questions were

1. Will the user try to achieve the right effect?
2. Will the user notice that the correct action is available?
3. Will the user associate the correct action with the effect that user is trying to achieve?
4. If the correct action is performed, will the user see that progress is being made toward solution of the task?

2.2.2. Analysis criteria

The evaluators independently stepped through the interface to carry out tasks. Based on the information obtained from the user's background knowledge and skills, each evaluator examined the appropriateness of design features by answering a checklist consisting of the above-mentioned questions for each action. To answer the questions, evaluators compared the users' conceptual model with that of the designers, while performing actions. Questions that were voted positively (success) by all three evaluators did not require any further consideration. Those that were voted negatively (failure) by at least one evaluator, were aggregated into a data collection form for further analysis. The data collection form consisted of the following items: evaluator's code, task number, action number and question number along with possible causes of failure. In order to organize the data, evaluators' responses were

categorized into four groups based on the aforementioned four questions. The inter-group and intra-group interpretations were provided in discussion. After that, in an expert panel, three evaluators shared their commentaries, resolved disagreements, and concurred on a final list of problems. Subsequently, each evaluator suggested a practical solution to each and every problem and rated the severity of each problem between 1 and 3 based on the level of difficulty for end-users. The mean scores of rating were assigned for the severity of a given problem (the higher the sore, the more severe the problem). In general, each problem was described with seven characteristics: the evaluator's ID, task number, action number, question number, severity of the problem, possible causes of the problem, and suggested solutions. The first five variables are quantitative, while the last two are qualitative. Consequently, the data were analyzed using both quantitative and qualitative approaches. In the rest of this section, the results obtained from each approach are provided, respectively. The relevant findings are presented in the results section.

2.2.3. Evaluators

In this research, the participants included three Ph.D. candidates of Medical Informatics who had passed the Usability Engineering (theoretical and practical) course and had experience with the usability evaluation methods. Two are M.S. holders in Computer Engineering and were familiar with hospital information systems, and the other one is an M.S. holder in medical records with over eight years of experience working as hospital IT manager.

3. Results

In general, 38 actions were identified to accomplish the defined scenario. Each action was analyzed by the evaluators to check if the user would be capable of fulfilling the target task through the current interface design. The final evaluation checklist for each expert ended up with 152 questions (38*4). The final list of usability problems after duplication removal and evaluators' consensus agreement comprised 116 problems.

3.1. Quantitative analysis

Quantitative results were categorized and reported from three different perspectives: Question-based, Task-based, and Evaluator-based.

3.1.1. *Question-based:*

The first row of Table 4 shows the percentages of the whole identified problems with respect to each of the four question types with the number of problems in parentheses. The second row of the table specifies the mean severity score in each category.

3.1.2. Task-based:

The first row of Table 5 shows the percentage of the whole problems shared by each task along with the problem counts presented in the parentheses. The second row of the table shows the mean severity of problems in each category.

3.1.3. Evaluator-based:

The percentage of the problems detected by each evaluator with their corresponding counts constitute the first row of the table and mean severity is presented in the second row of Table 6.

Table 4. Categorization of questions based on question type

Problems	Questions			
	Q1	Q2	Q3	Q4
% of problems (No.)	18% (21)	33% (38)	6%(7)	43% (50)
Mean severity	2.5	2	1.4	1.2

Table 5. Categorization of questions based on task list

Problems	Tasks						
	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7
% of problems(No.)	8% (9)	37% (43)	21% (24)	11% (13)	2% (3)	6% (7)	15% (17)
Mean severity	1.3	2.4	1.8	0.7	1.2	0.4	1.7

Table 6. Categorization of problems based on evaluators

Problems	Evaluators		
	Evaluator 1	Evaluator 2	Evaluator 3
% of problems (No.)	33% (38)	25% (29)	42% (49)
Mean severity score	2	1.5	2

3.2. Qualitative analysis

The most severe problems (score>2) for each question type are shown in Table 7. The qualitative findings of the study included a description of each problem and how it can be fixed.

Table 7. Results based on a qualitative evaluation of the usability of obstetrics and gynecology information system

Questions no.	Problem description	Suggested solution
1	No directions to guide user through sequence of tasks.	Hints are to be provided for users in each step of procedures.
2	The physician's name box has no label.	All boxes need to be labeled.
3	Use of incomprehensible English abbreviations for table headers.	Labels should use familiar terms
4	No feedback is provided on *save* button click	User should be informed of what is going on in system.

4. Discussion

4.1. Principal findings

To the best of our knowledge, this is the first study conducted to evaluate an obstetrics and gynecology information system using the CW method. One major challenge in the application of CW in healthcare information systems is that there is no pre-specified guideline for conducting methodology and reporting results (12). Therefore, the present study tried to present a step-by-step systematic procedure for conducting CW usability evaluation that can be used as a practical guide for future researches. In this study, a 7-item data collection form was introduced to facilitate the problem analyzing process. We had also presented the results in an innovative three-dimensional fashion (Question-based, Task-based, and Evaluator-based) which provides a more comprehensive perspective of the system problems. Since the CW method does not deal with the patients' data and merely evaluates the quality of the system interactivity, the results obtained in the present study can be generalized to other hospitals running the same version of the system. We have used the same three-dimension structure in this section to discuss and interpret the obtained results.

4.2. Question-based analysis

In the present study, it was found that the highest percentage of the system problems was related to the fourth question which investigated "whether the user is aware of progressing towards the goal after going through all the actions defined or not". Inadequate system feedback was found to be the main source of such problems, making users confused. Problems related to the 2nd and 3rd questions dealt with the accessibility of features on the interface and whether or not a logical relation exists between the UI design and its goal. These together comprised more than one third of all problems. They are mainly caused by an aberration from standards in icon design, hypertext phrases, and improper or irrelevant choice of labels. Studies (13-15) have also reported a high rate of violating the compatibility principle and disobeying the standards. In the long run, such problems lead to users' dissatisfaction and reluctance to use the system. Therefore, it is crucial for HIS designers to be aware of the standards. Although the problems related to the first question had the lowest percentage (18%), it is perceived as the most determinant by evaluators. In two other usability studies, problems concerning lack of user guidance were rated as serious (15, 16). Lack of guidance, imposes a cognitive load, fatigue and dissatisfaction on users (17).

4.3. Task-based analysis

Here, we examined the contribution of each task to the total problems identified and with the mean severity score for each task. The majority of the problems belong to task 2 (Sending a request to operation room). Crowded interface design with too many controls, ambiguous button labels, and invisible location of frequently used functions are some of the problems found for this task. Another explanation for the high number of errors in this task, can be due to its high number of actions (seven) as compared with other tasks. It is rational that huge tasks with several actions are subjected to more problems. Tasks 3 and 4 comply with this assumption.

4.4. Evaluation-based analysis

Considering the influence of evaluators' background and expertise on the type, number and severity of problems, it was observed that the 3rd evaluator who held a medical degree and had the experience of working in a hospital IT center, managed to identify more problems than the other two evaluators. This shows that more familiarity with work environment, hospital procedures, and end-users cause a more accurate assessment.

4.5. Strengths and weaknesses
This study is the first step to address the need for formal and standardized reporting of CW usability evaluation, and can guide future studies in this field. Evaluation of the system in its pilot implementation phase, as carried out in the present study, had been emphasized by several software development models (18) and helped the designers, users, and organizations to prevent problems in early stages which can reduce future potential costs. A list of problems detected in this study was provided to the special workgroup in the hospital IT department in order to modify and customize the system to meet the users' needs. It should be noted that the methodology explained in this study is generalizable and can be used in other domains as well. This study has the limitation of employing only three evaluators. However, previous studies showed that only a small sample of five to eight evaluators is needed in usability testing to detect 80-85% of usability problems (19).

4.6. Meaning of the study and directions for future research
Since several problems were found, in this study, to be associated with the obstetrics and gynecology information system, it is suggested that other SSO's sub-systems should be evaluated using the same method. While it is less costly in enhancing an information system before its implementation, it is recommended that usability evaluation starts from the early phases of system development life cycle. Also, future studies should make use of other evaluation methods as a complement for CW, to evaluate the system from different dimensions and also provide an opportunity to compare the problem list detected by the evaluators with the one detected by end-users. A review study conducted by Khajouei et al. (20) on the evaluation methods used in health information systems in Iran, showed that the dominant methods in most usability studies were quantitative methods. Although the statistical analysis of quantitative data is more straightforward, it does not provide explanation on the rationale and reasons behind system problems (20). Many methodologists believe that researches that use predefined quantitative tools and procedures cannot embrace the complexities of issues of usability problems. In order to take advantages of both quantitative and qualitative methodologies, a mixed approach was used in this study.

5. Conclusions
In this study, the obstetrics and gynecology information system was evaluated using the CW method, and found 116 usability problems which were then categorized into four groups. A problem list had been provided to the software development team in order to adapt the system to meet the organizational and clinical purposes. The system development team should not only take into account the functional issues but also attend to the usability aspects in the design and development phases. The other objective of this study was to provide a detailed formal description of how the cognitive usability evaluation can be applied and reported in a health care information system. The detailed, systematic methodology of conducting CW usability evaluation can be used as a practical guide for future studies. Evidently, no single usability evaluation method is superior to the other. There should be a balance in the time spent and the precision of evaluation method. To select the right method, users' goal and usage need to be defined. The use of multiple and mixed methods can be considered in future studies.

Conflict of Interest:
The authors declare no conflict of interest.

Authors' contributions:
R.K. and A.K. contributed to the conception and design of the study, acquisition and interpretation of the data, and drafting the paper. M.J., M.M. and S.Z. contributed to the acquisition, analysis and interpretation of the data. All five authors read and approved the final version of the article submitted.

References:
1) Kohn L, Corrigan J, Donaldson M. To err is human: building a safer health system. National Academy of Science, Institute of Medicine. 2002. PMID: 25077248.
2) Borycki E, Kushniruk A, Carvalho C. A methodology for validating safety heuristics using clinical simulations: identifying and preventing possible technology-induced errors related to using health information systems. Computational and mathematical methods in medicine. 2013; 2013.

3) Rahimi B, Vimarlund V. Methods to evaluate health information systems in healthcare settings: a literature review. J Med Syst. 2007; 31(5): 397-432. doi: 10.1007/s10916-007-9082-z. PMID: 17918694.

4) Bullard M, Holroyd BR, Meurer DP. How usability of a web-based clinical decision support system has the potential to contribute to adverse medical events. human-computer interaction. 2008; 16: 17.

5) Wharton C, Rieman J, Lewis C, Polson P. The cognitive walkthrough method: A practitioner's guide. Usability inspection methods. John Wiley & Sons, Inc; 1994.

6) Wechsung I. An evaluation framework for multimodal interaction. T-Labs Series in Telecommunication Services. Springer. 2014; 10: 978-3. doi: 10.1007/978-3-319-03810-0.

7) Guise JM, Segel S. Teamwork in obstetric critical care. Best Pract Res Clin Obstet Gynaecol. 2008; 22(5): 937-51. doi: 10.1016/j.bpobgyn.2008.06.010. PMID: 18701352, PMCID: PMC4987289.

8) Hajialiafzali H, Moss J, Mahmood M. Efficiency measurement for hospitals owned by the Iranian social security organisation. J Med Syst. 2007; 31(3): 166-72. doi: 10.1007/s10916-007-9051-6. PMID: 17622018.

9) Fariba Begdeli SJT, Amir Ashkan Nasiripour , Farhad Ghaffari Assessing Hospital Governance in Iran Social Security Organization Journal of Applied Environmental and Biological Sciences 2015;5(9):290-5.

10) Sinding SW. Population, poverty and economic development. Philos Trans R Soc Lond B Biol Sci. 2009; 364(1532): 3023-30. doi: 10.1098/rstb.2009.0145. PMID: 19770153, PMCID: PMC2781831.

11) Computer Skills: Levels of Proficiency. In: university C, editor. 2011.

12) Ellsworth MA, Dziadzko M, O'Horo JC, Farrell AM, Zhang J, Herasevich V. An appraisal of published usability evaluations of electronic health records via systematic review. J Am Med Inform Assoc. 2017; 24(1): 218-26. doi: 10.1093/jamia/ocw046. PMID: 27107451.

13) Edwards PJ, Moloney KP, Jacko JA, Sainfort F. Evaluating usability of a commercial electronic health record: A case study. International Journal of Human-Computer Studies. 2008; 66(10): 718-28. doi: 10.1016/j.ijhcs.2008.06.002.

14) Choi J, Bakken S. Web-based education for low-literate parents in Neonatal Intensive Care Unit: Development of a website and heuristic evaluation and usability testing. Int J Med Inform. 2010; 79(8): 565-75. doi: 10.1016/j.ijmedinf.2010.05.001. PMID: 20617546, PMCID: PMC2956000.

15) Joshi A, Arora M, Dai L, Price K, Vizer L, Sears A. Usability of a patient education and motivation tool using heuristic evaluation. J Med Internet Res. 2009; 11(4): e47. doi: 10.2196/jmir.1244. PMID: 19897458, PMCID: PMC2802560.

16) Pressler TR, Yen PY, Ding J, Liu J, Embi PJ, Payne PR. Computational challenges and human factors influencing the design and use of clinical research participant eligibility pre-screening tools. BMC Med Inform Decis Mak. 2012; 12(1): 47. doi: 10.1186/1472-6947-12-47. PMID: 22646313, PMCID: PMC3407791.

17) Horsky J, Kaufman DR, Oppenheim MI, Patel VL. A framework for analyzing the cognitive complexity of computer-assisted clinical ordering. Journal of biomedical informatics. 2003; 36(1): 4-22. doi: 10.1016/S1532-0464(03)00062-5.

18) Dybå T, Dingsøyr T. Empirical studies of agile software development: A systematic review. Information and software technology. 2008; 50(9): 833-59. doi: 10.1016/j.infsof.2008.01.006.

19) Nielsen J, Landauer TK. A mathematical model of the finding of usability problems. Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems; 1993: ACM. doi: 10.1145/169059.169166.

20) Khajouei R, Ahmadian L. Methods Used for Evaluation of Health Information Systems in Iran. Journal of Health Administration. 2013; 16(53): 7-21.