Usability Study of a Pre-anesthesia Evaluation App in an University Hospital: Before the Revision of User Interface

Po-Yuan Shih[1,2] and Meng-Cong Zheng[1(✉)]

[1] Department of Industrial Design, National Taipei University of Technology, Taipei, Taiwan
zmcdesign@gmail.com
[2] Department of Anesthesiology, National Taiwan University Hospital, Taipei, Taiwan

Abstract. A pre-anesthesia evaluation app named EVAN was developed to facilitate anesthesiologists in their daily practice. EVAN contained a well-organized interface for medical records and an evaluation entry page. The purpose of this study is to assess the usability of EVAN and to find the weakness in the user interface before the revision. Twenty participants were recruited in this study with three components: user task analysis, questionnaires, and semi-structured interview. During the user task analysis, participants were asked to complete 16 tasks under monitoring. The second component of the study is questionnaires, consisted of SUS and AttrakDiff-Short. According to the results, seven participants completed all the tasks successfully. Two tasks had the most failures (8 and 7 failures, respectively). Participants were unable to find a specific function due to inadequate clues on the interactive buttons. Results of task analysis also revealed that the current logic of data storage which aimed to prevent accidental deletion, was not intuitive for most participants. Despite the drawbacks, the gradings of SUS questionnaire was 80.6 ± 14.4, which is an above average result. Scores from the AttrakDiff-Short were all positive, for both pragmatic quality, hedonic quality, and attractiveness. The results of this study are summarized and integrated for the revision of EVAN.

Keywords: User interface · Usability · Anesthesia · Electronic medical record

1 Introduction

A successful surgery requires a high-quality anesthesia. A good pre-anesthesia evaluation is a crucial step for high-quality anesthesia. To complete the evaluation, anesthesiologists must check the medical records, including laboratory examinations and image studies, perform the physical examination, and fill in the evaluation document. Previously in National Taiwan University Hospital (NTUH), a top-notch medical center in Taiwan, anesthesiologists have to write down the evaluation results in paper. To improve the evaluation efficiency of the above meticulous process, an intuitive electronic medical record viewing system and a convenient data entry page are needed. In 2018, a pre-anesthesia evaluation information system was launched. It consisted of a

web page and a mobile application(app). The web page was only for filling in the electronic version of the evaluation form. The app was named EVAN, as "EValuation for ANesthesiologist". It was an iOS-running app on iPad mini, and its user interface (UI) was designed by the Design Psychology Lab of National Taipei University of Technology. EVAN featured an electronic medical record system and an evaluation entry interface. This empirical study is held to evaluate the usability and the reference of future updates of UI design.

2   Methods

2.1   Design Concepts

Two main methods were used to clarify the user needs. Clinical shadowing, where the designers followed and observed the anesthesiologists undergoing the pre-anesthesia evaluation, was used to understand the usage scenarios. After clinical shadowing, the anesthesiologists were interviewed. With these methods, designers could understand the details of pre-anesthesia evaluation and expectations from the users.

The users, in this case anesthesiologists, used EVAN for two major purpose. First, EVAN was used to check the patients' medical records, laboratory examinations, and image studies. Thus, EVAN should contain a well-organized interface for all types of medical data. (Figure 1) The categories were organized according to the users' preference. Second, EVAN was used to fill the pre-anesthesia evaluation form. A questionnaire-like interface with clear hierarchy is essential. (Figure 2) When anesthesiologists filled the form in front of the patient, they were easily interrupted. To prevent the unsaved evaluation from accidental deletion, we designed a "Confirm" button for every category of the questionnaire. In addition, the questionnaire was saved automatically when the users exited the questionnaire page with the top right button.

2.2   Participants

Twenty participants were recruited in this study. All the participants were the attending anesthesiologists or anesthesia residents of NTUH who were authorized to perform the pre-anesthesia evaluation.

2.3   Procedures

This study composed of three components, including task analysis, questionnaires, and semi-structured interview. All procedures were conducted in an isolated room with a researcher and a research assistant.
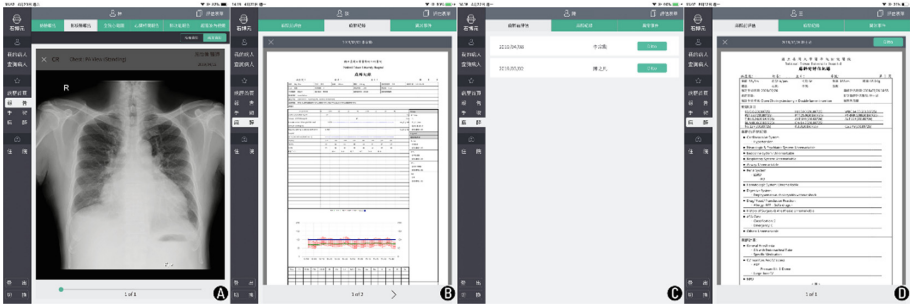
Fig. 1. The interface for different types of medical reports and records. A, medical image; B, previous medical record; C, the list of previous pre-anesthesia evaluation; D, the document of the pre-anesthesia evaluation.
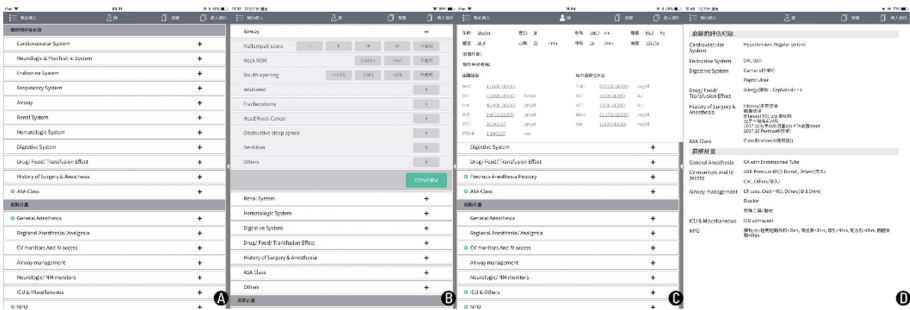
Fig. 2. The evaluation questionnaire entry page. A, the major topics of the evaluation; B, the hierarchy of the evaluation; C, the basic data window by pressing the top middle icon with the patient's name; D, the preview page by pressing the right second icon with the word preview in Traditional Chinese.

2.4   Task Analysis

Participants were asked to use an iPad mini to perform 16 tasks, with instructions projected on screen. The iPad mini was connected to a computer for surveillance and screen recording. The researcher also judged the completeness of every single task via the connected computer. Tasks were marked as failure when participants waived the task or when the completion time for each task was greater than 2 min. The results including failure and task time was recorded by the research assistant. Sixteen tasks consisted of three parts. The detailed tasks were shown in Table 2.

1. Searching the patient with specified conditions. (Task 1) Participants were given a date and a specific operating room number to find the patient.
2. Checking the patient data. (Task 2–8) Participants were asked to check the specified information, including previous pre-anesthesia evaluation, medical records, and laboratory examination results. The tasks were marked as a success only when they were performed in the expected manner.

3. Filling, saving, and previewing the pre-anesthesia evaluation under a paper instruction. (Task 9–16) Participants were asked to perform the tasks three times (first time: task 9–11, second time: task 13–15, and the third time: task 16).

The in-app activity was recorded with the connected computer and with the service from UXCAM.

2.5   Questionnaire

In the first part of the questionnaire, we investigate the participants' experience for EVAN. If the participants had used EVAN during any steps of their daily evaluation work, they were marked as experienced users. We used System Usability Scale (SUS), one of the most popular and standardized questionnaires, to assess the usability of EVAN. SUS consisted of 10 questions with alternating tone. [1]

We used AttrakDiff-Short to assess hedonic quality of user experience, the aspect that could not be evaluated with SUS. AttrakDiff-Short is extracted from the original AttrakDiff. The original AttrakDiff was developed by Hassenzahl and aimed to evaluate the hedonic quality of user interface.[2] Hassenzahl proposed that the perceived attractiveness of a user interface could be separated into two distinct factors: hedonic and pragmatic quality (HQ and PQ). While HQ is related to the quality and aesthetic aspects, PQ can be taken as the perceived usability. The AttrakDiff-Short questionnaire rated 10 pairs of contrasting adjectives in 7 scales. It consisted of 4 questions on PQ, 4 questions on HQ and 2 questions on the general attractiveness.

Table 1. Demographics of participants enrolled in this study

Participant demographics	Number
Total participants enrolled	20
Male	13
Female	7
Attending anesthesiologist	12
Anesthesia resident	8
Experienced user	8
Non-experienced user	12

2.6   Semi-structured Interview

The last component of the study was a semi-structured interview. The participants reviewed previous tasks and were urged to elaborate on their thinking process while performing the tasks. The researcher would make a detailed inquiry when the participants made mistakes or hesitate during the user tasks. After reviewing the tasks, participants were asked to share their opinions on EVAN in all aspects, including the pros and cons.

3   Results

Table 2. The results of SUS and task analysis

Results			
SUS	80.6 ± 14.4		
Experienced user	85.3 ± 13.5		
Non-experienced user	77.5 ± 12.8		
Task results	Task time (sec)	Number of failures	
Task 1: Find the patient under instruction	53.5 ± 14.8	0	
Task 2: Find specific data in previous pre-anesthesia evaluations	60.6 ± 30.7	7	
Task 3: Find specific data in medical records	30.2 ± 10.9	0	
Task 4: Find specific data in medical records	21.0 ± 5.9	0	
Task 5: Find specific data in medical records	33.4 ± 18.8	0	
Task 6: Find specific data in medical records	23.9 ± 17.2	0	
Task 7: Enter the evaluation entry page	10.7 ± 4.8	0	
Task 8: Find specific data in basic data window	16.5 ± 10.9	8	
Task 9: Fill the evaluation form (first time)	110.7 ± 38.1	2	
Task 10: Save the document (first time)	35.7 ± 33.5	4	
Task 11: Preview the results (first time)	52.1 ± 41.3	2	
Task 12: Find specific data in medical records	18.8 ± 9.4	1	
Task 13: Fill the evaluation form (second time)	115.7 ± 32.4	1	
Task 14: Save the document (second time)	11.2 ± 16.0	3	
Task 15: Preview the results (second time)	20.5 ± 14.0	1	
Task 16: Fill, save, and preview the document (third time)	21.0 ± 11.5	0	
Observational results in repeat tasks:	First time	Second time	Third time
Negligence of "Confirm"	11	3	0
Failed to save the document	4	3	0
Preview the document	2	1	0

3.1   Demographic Results

Twenty participants completed this study including 7 females and 13 males. The demographic data were shown in Table 1. Eight participants were experienced users, while 12 participants did not use EVAN in their daily practice. The mean age of the participants was 38.6 ± 11.0-year-old. The mean working experience was 13.4 ± 10.8 years.

3.2 Usability and User Experience

The average score of SUS was 80.6 ± 14.4. According to the Sauro-Lewis curved grading scale (CGS), EVAN was graded as A-, which was an above-average usability. We performed a subgroup analysis according to the participants' experience of EVAN. The scores of experienced and non-experienced users were 85.3 ± 13.5 and 77.5 ± 12.8, respectively. However, there were no differences in the SUS scores between experienced and non-experienced users. The AttrakDiff results were shown in Fig. 3. All the 10 adjectives in PQ, HQ, and ATT were positive, which showed a favorable attitude from the participants. The mean PQ, HQ, and ATT gradings were 1.2, 1.0, and 1.6 respectively.
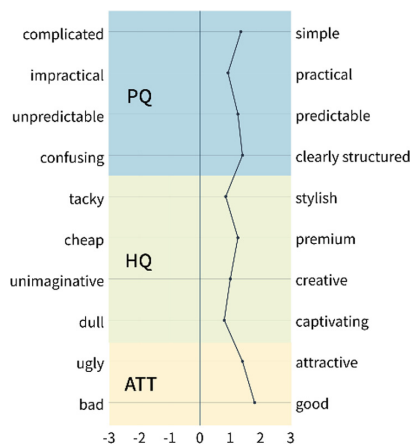

Fig. 3. The results of AttrakDiff Short questionnaire.

3.3 Task Analysis

The task time and the number of tasks failed was shown in Table 2. Eight participants completed all the tasks without failure. Among them, 5 participants were experienced users. Task 8 had the most failures, followed by task 2; 8 and 7 participants failed respectively. Eight participants failed in the task 8, and seven participants failed in task 2. Among eight participants who failed in task 8, three waived the task, while the other five completed the task with incorrect methods. Similar situations were noted in task 2, five out of seven participants completed the task in incorrect ways.

As mentioned previously, we asked the participants to fill, save, and review the evaluation form three times. One behavior noted during filling the evaluation was that the negligence of the "Confirm" button. (Figure 2B) Eleven out of twelve non-experienced users shared the same behavior in their first time. As shown in Table 2, the number of failures decreased as the participants gained experience while performing tasks.

4 Discussion

4.1 Perceived Usability

In the questionnaire, we used SUS to assess perceived usability. The average rating of EVAN, a business app, was 80.6 ± 14.4. In a previous study by Kortum et al. [3], the top 10 mobile apps in platforms were rated 77.7 in SUS. It is surprising that a business app was rated higher than that of 10 popular apps. Part of the reason might be that EVAN was inadvertently compared with the hospital information system (HIS) in NTUH, and the UI design of HIS was not user-oriented designed. On the other hand, the high rating could also indicate the improved workflow while undergoing pre-anesthesia evaluation with EVAN. These steps might be cumbersome, and some non-experienced users said that they would incorporate EVAN in their workflow after the study. The AttrakDiff-Short ratings from the participants were positive in all PQ, HQ and ATT. Reasons for these results were similar with that of SUS. Participants pointed out that the most distinguish difference between the two was the well-organized information layouts with distinct hierarchy. In the past, the hedonic quality was more emphasized in web page and entertaining software. A study held by Schrepp et al. showed that PQ and HQ of business software were perceived independently by users and contributed equally to attractiveness. [4] In our study, EVAN was taken as an evaluation tool with high PQ (1.2). However, the high ATT gradings (1.6) could be partially owing to the high HQ (1.0).

4.2 Task Analysis

Many participants failed in two specific tasks. One of them was to find the patient's body weight in the basic data window. (Figure 2C) This information could be found in the "personal basic data page" after tapping button consisted of the patient's name and icon. However, eight participants could not recognize this functional button. During the interview, all the non-experienced users who successfully completed this task said that they found this function by accident. Although similar design was used for the "pre-view page", (Fig. 2D) almost all participants thought that the button with the word "preview" (in Traditional Chinese) made it easier to predict the function of the button. Thus, the number of participants failed to preview was far less than that of participants failed to find patient's body weight. Most of them agreed that it is an easy-to-remember function, but they thought the hint of the functional area should be more recognizable. This is one of the designs to be changed in the future updates.

The other task, in which seven participants failed, was to find a specified information from a list of previously assessed pre-anesthesia evaluation. The standard procedure was to look for the most recent evaluation, tap the list (white rectangle in Fig. 1C) to open the evaluation form and find the specified data in a previous evaluation report. Five of the failed participants did not picture an interactive bar for each evaluation. Instead, they pressed the "Ditto" button (the green button in the white rectangle in Fig. 1D), which function was to copy the content and paste it to a new evaluation form. Although they still have the right data, this task would still be viewed as a failure. Participants later revealed that they neglected the bar when seeing the

"Ditto" button. To improve the performance, the "Ditto" button should only be seen after viewing the content of the previous evaluation.

The most important finding in the screen recording was that the "Confirm" button was neglected initially by eleven participants. To prevent the whole entries being deleted without saving, we designed the "Confirm" button in every major topic. Once the users completed the questions under a major topic, they should click on the "Confirm" button to save the content within this topic. This logic was intended to save the whole entries in small parts instead of the whole document as pre-anesthesia evaluation rarely go uninterrupted. However, neglecting the "Confirm" button meant that all current entries in this topic were futile. During the interview, most participants, even those who completed this task successfully, mentioned that the button was easily neglected and impractical. On the other hand, they appreciate the concept of saving the document in small parts. These evidences gave us clues for future UI updates.

5 Conclusions and Future Works

In the current study, the UI design of EVAN was satisfactory for users in NTUH, but with minor flaws. These empirical results could be concluded into two major directions for future design:

1. Improve the hint of interactive areas: Most failures in the user tasks resulted from inadequate clues of the interactive buttons. One of the solutions is to enhance the metaphor of the interactive areas with icon and its adjacent keywords (as in the case of Fig. 2C). Another way is to increase the contrast between the interactive area and the background (as in the case of Fig. 1C).
2. Adjust the workflow for data storage to be more intuitive: The rationale of the "Confirm" button was to prevent accidental deletion. However, the logic of "saving document in parts" was not intuitive for many users. To correct the design flaw, the "Confirm" button will be cancelled. The mechanism of saving will be shifted to "automatic mode", with every change saved.

References

1. Lewis, J.R.: The system usability scale: past, present, and future. Int. J. Hum. Comput. Interact. **34**(7), 577–590 (2018)
2. Hassenzahl, M.: The interplay of beauty, goodness, and usability in interactive products. Hum. Comput. Interact. **19**(4), 319–349 (2004)
3. Kortum, P., Sorber, M.: Measuring the usability of mobile applications for phones and tablets. Int. J. Hum. Comput. Interact **31**(8), 518–529 (2015)
4. Schrepp, M., Held, T., Laugwitz, B.: The influence of hedonic quality on the attractiveness of user interfaces of business management software. Interact. Comput. **18**(5), 1055–1069 (2006)