A Comparative Analysis of Usability Evaluation Methods of Academic Mobile Application: Are Four Methods Better?

Sabda Norman Hayat  
Department of Information System  
Brawijaya University, Malang, Indonesia  
sabdanorman@student.ub.ac.id

Fatwa Ramdani  
Department of Information System  
Brawijaya University, Malang, Indonesia  
fatwaramdani@ub.ac.id

ABSTRACT
Usability evaluation is an important element which has used to add on insights related to the usability problem of an application. This study aims to identify application’s usability problems and to compare the effectiveness of four evaluation methods which has used on an academic portal mobile application: usability testing, interviews, surveys, and heuristic evaluation. The data has collected from users who are in the student category and expert evaluator. The number detail of respondents used are usability testing (N = 10), interviews (N = 10), surveys (N = 110), and heuristic evaluation (N = 3). The four methods identified a total of 44 usability problems: 45% using heuristic evaluation, 24% using surveys, 17% using interviews and 14% using usability testing, resulting into a few similar findings. Then, The problems are categorized using Usability Taxonomy Problem (UPT) which has divided into 5 categories with details of 17 categories of visualness, 6 categories of language, 3 categories of manipulation, 11 categories of task-mapping and 7 others including the category of task-facilitation. The results of this study are capable to prove that the four methods are complementary, each method provides a unique insight to improve the usability of the application user interface. Both researchers recommend using a multi-method approach when evaluating the usability of an application due to it could provide a more comprehensive representation of usability issues.

KEYWORDS
Usability Evaluation, Comparative Analysis, Multi-method Evaluation, Usability Problem Taxonomy

CCS CONCEPTS
• Human-centered computing → Human computer interaction (HCI) → HCI design and evaluation methods

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

1 INTRODUCTION
Gapura UB is a mobile application-based innovation that has been applied to the academics of Brawijaya University in order to realize a digital campus. The application is an integrated service information system that aims to provide convenience related to Brawijaya university information system [1].

In order to explore issues related to the application, both researchers conducted interviews with one of the developers of the application. The interview result is stated that the application has never been evaluated usability by testing the usability of its users or inspection by an expert. Even though, the usability evaluation is an important element which allows identifying the performance of a system or application [2].

Usability testing to users is needed to assess and to ensure that the usability of the application is effective, efficient, and satisfying. Testing needs to be done because the quality and user acceptance of an application largely depend on the characteristics of ease of use [3]. In addition, according to Jabar et al. [4] inspection of usability by an expert is also needed. The inspection by an expert is able to identify problems and improve the usability of the user interface design by inspecting it using existing standards.

Several methods that can be used to evaluate the usability of an application, namely: usability testing, interviews, surveys and heuristic evaluation [5]. Usability testing, interviews and surveys are methods that use users as a respondent. While, heuristic evaluation is a method that uses experts as an evaluator [6].

According to Nielsen [7], usability testing is a simple method that could be used to view what happens when testing. Another advantages is easy to get real usability knowledge and cheap.
Hence, only requires a small number of users to do the test. Interview is one of the efficient methods for finding usability problems, such as when problems occur that can’t be found by observing user behavior by using the usability testing method [8]. Meanwhile, survey is a method used to understand broader user base preferences related to usability issues, and can be used instead of usability testing [9].

Next, heuristic evaluation is a usability evaluation method that uses evaluators to find usability problems or errors that might adversely affect the user’s ability to interact with the system [10]. Nielsen and Molich [11] described the heuristic evaluation as a method that is cheap, easy to do, can be done from the beginning of development and can be used together with other usability evaluation methods to evaluate user interfaces.

These four methods are the most widely used methods for evaluating usability [6]. Previous studies conducted by Walji et al. [12] compared three user-based methods, namely user-testing, interviews, and surveys to evaluate an electronic health records system. This study adds the heuristic evaluation method that was not used in previous studies as an expert-based method to compare how big the role of the expert is in finding usability problems compared to other methods.

This study aims to identify the usability problems of the application, to categorize the problems found and to compare the effectiveness of the four evaluation methods used in this study: usability testing, interviews, surveys, and heuristic evaluation. Both researchers hope to help further research to be able to determine more effective method used to evaluate usability for an application.

2 REVIEW OF RELATED LITERATURE

Research conducted by Georgsson and Staggers [13] has a topic about evaluation of diabetes health system usability. The study used three method approaches to determine usability issues that are usability testing, semi-structured interviews, and questionnaires. The results of this study are the discovery of 117 usability problems which are divided into 19 aspects and have a severity rating of 2.47. Usability testing can find problems as much as 50%, semi-structured interviews as much as 29% and questionnaires as much as 21%.

Next, research conducted by Walji et al. [12] has a topic about usability evaluation on electronic health records systems. The study use three methods. It is user testing, semi-structured interviews and surveys to find problems related to usability. The results of this study prove that user testing can find more problems related to usability than the semi-structured interview and survey methods, with a percentage of 54%, 28% and 18% of the total 187 problems found.

Next, research conducted by Ssemugabi and Villiers [14] has a topic about comparison of usability evaluation methods in web-based learning applications. In this study, the method used is user-based survey and heuristic evaluation. The results of this study indicated that the heuristic evaluation method by experts is more effective to be used in identifying problems related to usability in web-based learning applications. The facts of the study has proved that the evaluation using the heuristic evaluation method is better than the user-based survey method, even though, it’s only done by four experts compared to 61 students. Experts can find 58 problems compared to 55 that can be identified in the student survey. The conclusion of the study is that the heuristic evaluation method by experts corresponds well to the evaluation using the user-based survey method by students.

2.1 Usability Problem Taxonomy (UPT)

Usability Problem Taxonomy (UPT) is a taxonomic model in which is usability problems are detected and found in a graphical user interface. UPT has 5 main categories which are divided into 2 components, namely artifact component and task component. The artifact component focuses on usability issues that relating to artifacts or objects in the application interface. It consists of visualness (the quality, condition, or degree of being visible), language, and manipulation categories. The task component focuses on usability issues related to work on a system that consists of task mapping and task facilitation categories [15].

3 METHODOLOGY

This research methodologies consists of problem identification, literature study, evaluation planning, evaluation implementation, analysis of evaluation results, and conclusions and suggestions.

3.1 Study Sample and Setting

Determination of respondents used purposive sampling technique. Respondents are required to have certain criteria including active students of Brawijaya University which have a Student Identity Card. The total population of users of the Gapura UB application is more than fifteen thousand people, most of whom are students. Respondents used for the usability testing and interview method consist of ten people as suggested by Nielsen [16] to find more problems. While, for the questionnaire survey method the number of respondents is 110 people according to the formula of Slovin [17].

3.2 Data Acquisition

This study uses four evaluation methods, three user-based namely usability testing, interviews, surveys and one expert-based method, heuristic evaluation. Interview method used several questions, the questions used are some open questions related to the Gapura UB application. The question is used to explore and find out more deeply related to usability problems experienced and felt by users [18]. The interview used was a semi-structured interview. Interviews were conducted twice to obtain reliable results.

The questionnaire used in this study was USE Questionnaire [19]. The questionnaire was also added three optional questions regarding satisfaction, dissatisfaction, and suggestions for improvement. Then the questionnaire was tested in two stages, namely the validity test using the 2-sided Bivariate Pearson correlation test with a significance of 0.05 and the reliability test using the Cronbach Alpha formula. Experts involved as an evaluator in this study were experts in the field of user interface.
Comparative Analysis of Usability Evaluation Using User-Based and Expert Methods in the Academic Portal Mobile Application design. The number of experts used in this study was three according to Nielsen's [20] theory. The evaluation activities were carried out in accordance with the nine heuristic principles proposed by Monroy [21]. After each expert conducts an inspection, a consolidation between the experts is carried out.

For usability testing method we used several task scenarios. The task scenario is designed based on several functions that contained in the Gapura UB application. Each task has been assigned to the user to measure the usability of the application [9], a total task scenario used was eight tasks. Prior to being given to respondents, preliminary testing was carried out to ensure that the assigned tasks were appropriate and covered all the functions of the application [16].

### 3.3 Analysis of Evaluation Results

Categorization of usability problems uses the Usability Problem Taxonomy (UPT) proposed by Keenan [16] and compares the percentage of usability problem findings from the four methods used. Furthermore, the percentage of three user-based methods is combined and compared to the percentage of expert-based methods, namely heuristic evaluation.

### 4 DISCUSSION

Each evaluation method found a number of different usability problems was: usability testing found 13 problems, interviews found 14 problems, surveys found 17 problems and heuristic evaluation found 23 problems. There were 13 usability problems that found simultaneously by more than one evaluation method, so it needs to be synchronized to be a unique problem.

#### 4.1 Categorizing the Problem

Unique usability problems found as many as 44 problems. Then, these problems were categorized by UPT, categorization can be seen more detail in Figure 1.

**Figure 1: Categorizing Usability Problems Using UPT**

Based on Figure 1, it can be seen that of the 44 problems found divided into five categories: 17 problems included in the visualness category, 6 problems included in the language category, 3 problems included in the manipulation category, 11 problems included in the task-mapping category and 7 problems included in the task-facilitation category.

#### 4.1.1 Categorizing the Problems for Each Method

Categorizing the problems for each method is used to find out which problems in the category are most frequently found by each method.

**Figure 2: Problem Categories Usability Testing Methods Using UPT**

Based on the Figure 2-5, it can be concluded that each evaluation method had a majority of different problem finding categories. Usability testing method found usability problems that are mostly related to the task-mapping category or technical issues of use. Survey methods found usability issues that were mostly included in the visualness category and were related to artifacts or objects on the interface or can be said graphically related problems.

**Figure 3: Problem Categories Interview Methods Using UPT**

Heuristic evaluation method found usability problems that are more complex. It is a combination of problems related to artifacts
or objects in the interface and problems related to work tasks with finding the same number of problems in each category.

The usability testing method finds many problems related to technical use because the user directly tries the application by doing the given tasks. The survey method finds many problems related to artifacts or visuals due to the user does not know directly what to do so that the assessment is only limited to the appearance of the application interface. While the heuristic evaluation method of problem finding is almost balanced between usage problems and visual problems because the experts can find problems in accordance with the rules of science being studied.

4.2 Mapping and Comparing Usability Evaluation Methods

Usability problem mapping is used to classify usability problem findings according to the method used. Based on Figure 6, it can be concluded that each method succeeded in finding several unique insights regarding usability issues. The unique problems found by the heuristic evaluation method are far more than the problems found using the other three methods. The usability testing method only found 2 unique problems, the interview method found 4 problems, the survey method found 8 problems and the heuristic evaluation method found 17 problems.

4.3 Comparison of User and Expert Based Evaluation Methods

This section explained the comparison of findings the three user-based methods with one expert-based method. The respondents’ details of the three user-based methods were the usability testing method using 10 respondents, the interview method using 10 respondents, and the survey using 110 respondents. The total number of the three user-based methods used 130 different
The expert-based method is heuristic evaluation uses as many as 3 experts in the field of interface. The following is a mapping of problem findings by users compared to experts.

Based on Figure 8, user-based methods can find 21 unique problems and 6 similar problems with experts, while expert-based methods can find 17 unique problems. This proves that the user-based method and the expert-based method can correspond well, the usability problems found are diverse from one another. Users and experts have a tendency to find usability problems with different segments.

Based on Figure 9, the percentage of problem findings from the three user-based methods when combines into 55% with 130 respondents, slightly above than the 3 experts who successfully to find a problem of 45% overall.

5 CONCLUSION
The results of the usability evaluation on the Gapura UB application using four methods that is usability testing, interviews, surveys and heuristic evaluation found 44 unique problems, 13 of the 44 problems are the same problems found by more than one different method. The usability testing method found 13 usability problems. The interview method found 14 problems with semi-structured interviews. The survey method found 17 problems. The heuristic evaluation method found 23 problems.

Categorizing 44 usability problems using UPT divides the problem into 5 categories with the details that 17 problems include the visualness category, 6 problems including the language category, 3 problems including the manipulation category, 11 problems including the task-mapping category and 7 problems including the task-facilitation category. Each evaluation method has a majority of different problem finding categories.

Each method provided several unique insight for increased usability of the application user interface. Calculation of the percentage comparison of findings of usability problems in the Gapura UB application using four methods including, the usability testing method by 14% using 10 respondents, interview method by 17% using 10 respondents, survey method by 24% using 110 respondents and heuristic evaluation method by 45% using 3 experts. Expert-based methods are more effective in evaluating application usability than user-based methods. Expert-based and user-based methods find different problem segmentation hence could be complementing each other.

The researchers recommend for further research to use a multi-method approach when evaluating application or system usability due to it could provide a more comprehensive representation of usability issues.

ACKNOWLEDGMENTS
We are indebted to all respondents who have participated in this study. Hence, this study could be done very well.

REFERENCES


