Learning Experience Evaluation of Indonesia Online Tutoring Mobile Application using Heuristic for E-learning

Hanifah Muslimah Az-zahra
Department of System Information
Brawijaya University
Malang Indonesia
hanifah.azzahra@ub.ac.id

Retno Indah Rokhmawati
Department of System Information
Brawijaya University
Malang Indonesia
retnoindahr@ub.ac.id

Cheldo Bagus Romansya
Department of System Information
Brawijaya University
Malang Indonesia
romanisme87@gmail.com

ABSTRACT
Ruang Guru has been Indonesia’s most popular online tutoring platform in 2019. This platform is intended for primary to high school students in understanding school curriculums, as well as helping with exams. However, teens may have a low success rate because of their overconfident behavior. Their undeveloped skills require simple and relatable applications. This fact raises the question of whether teenage students are able to learn through this e-learning. This study aims to find user experience problems in this e-learning application that might impact student learning success. The evaluation of the learning experience of e-learning should not only be related to usability but also instructional design aspects. This study evaluates an online tutor mobile application using the heuristic evaluation method by referring to the usability and instructional design heuristics. The results found 28 problems. Some problems were beyond determined heuristics. The mean calculation of severity level indicates minor problems that have a low improvement priority.

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

KEYWORDS
Heuristic evaluation, e-learning, learning experience, usability, user experience, instructional design.

1 Introduction
In early 2019, Ruang Guru was recorded to be the most downloaded online tutoring application on Google Play. This platform is intended for elementary to high school students in understanding the school curriculum, as well as helping with exams, particularly standardized exams. With 300,000 tutors [18], this platform allows students to find tutors and lessons needed. The hundreds of lessons [18] provided are school-related subjects: math, science, languages, and others.

A research revealed that teenagers, including high school students, facing major challenges while interacting with a system that has dense content and poor navigation; thus they require a simple and relatable system. Teenagers, similar to adults, are also goal-oriented but perform worse than adults [6]. However, teens are overconfident in their abilities and impatient. They don’t like to read a lot; resulting in low reading levels and undeveloped research skills. Those reasons lead to reduce teens’ task success. Systems that targeted the teenager, as well as for any other user group, must focus on the usability aspect. Usability is related to how easy user achieving their goal with a specific system in a specific context of use [5]. This fact raises the question of whether teenage students are able to learn through e-learning.

Moreover, the usability alone does not cover the whole experience of the user interacting with the system, because it is only related to the ease of use. Usability is only part of the user experience (UX). Hassenzahl argues that user experience includes pragmatic (instrumental) and hedonic (non-instrumental) qualities [4]. While the pragmatic quality refers to the perceived usefulness, efficiency, and ease of use (so-called utility and usability aspects); the hedonic quality is related to
“joy of use” and emphasizes the stimulation, identification, and evocation generated by the use of a system or product [3].

Regarding educational areas, the definition of usability and user experience, in general, is not sufficient to cover user experience in e-learning [13]. E-learning is very different from other interactive systems because the purpose of e-learning is not only to make the user successfully do the task joyfully but more importantly is to support them learning while doing the task [17]. The interaction on e-learning goes beyond the paradigm of traditional task-based usability, being directly related to the pedagogical value [16]. Therefore, the term “learning experience” is used in this paper. The learning experience is related to experiences that arise as a result of learning through a specific academic setting or educational interactions [2].

Furthermore, the usability and user experience evaluation methods in general are also not enough to measure the success of e-learning [7]. Pedagogical aspects have to be considered in measuring the success of interactions between learners and e-learning [7]. Several studies have focused on examining methods for usability evaluation in e-learning. Freire et al. [2] reviewed dozens of usability evaluation methods for e-learning. They found three main differences among the focus of these methods, including (i) system performance evaluation, (ii) user performance evaluation, and (iii) user-system dialogues evaluation.

This study was conducted based on the fact that teenage students face problems when interacting with complex systems. So, it is necessary to know whether this online tutoring platform can help students to learn. So this study will use a method that focuses on system performance evaluation: a heuristic evaluation. Reeves et al proposed 15 heuristics principles for e-learning evaluation. These fifteen principles were modified from traditionally Nielsen’s 10 usability heuristics by considering instructional aspects. This paper evaluates the learning experience of an online tutor application that is popular in Indonesia, Ruang Guru, using heuristic evaluation which refers to the set of fifteen e-learning heuristics. The results of this study will show whether the set of fifteen e-learning heuristics cover all the problems related to usability in e-learning.

### 2 E-Learning Heuristic Evaluation

Heuristic evaluation is a usability inspection method involving experts to observe the user interfaces of a system [10]. For the best results, double experts who are usability expert and domain expert is suggested to be involved [8]. Reeves et al proposed the e-learning heuristic evaluation instrument and protocol that are intended for use by instructional designers and other experts engaged in heuristic evaluations of e-learning programs [13].

Heuristic evaluation inspection is carried out using established principles to find usability problems. In general, the evaluation principle refers to Nielsen’s ten heuristics for user interface design. However, it is possible to modify the heuristic principles according to the research needs and system domain [10]. The e-learning heuristic evaluation protocol also allows experts to add new heuristics that are relevant to the types of e-learning programs being evaluated or to the expert’s specific expertise [13].

Reeves et al modified Nielsen’s usability evaluation protocol and adjusted it to evaluate e-learning programs. Modifications made are developing Nielsen’s original 10 heuristics into 15 heuristics aimed specifically at e-learning. The added heuristic principle is related to the instructional design aspects that are very important in learning [1]. Table 1 details the fifteen usability and instructional design heuristics for evaluation of e-learning programs.

### 3 Methodology

Nielsen suggests involving three to five evaluators to find more than 60% of usability problems [10]. Nielsen also propounds three types of expert knowledge: novice, regular, and double expert [8]. The evaluators involved were four experts who are educational practitioners; some have knowledge related to education and some knowledge about system and software development.

The evaluation was conducted by referring to the usability and instructional design heuristics for evaluation of e-learning programs. Before the main evaluation is carried out, experts were asked to understand the set of e-learning heuristics. We also give a picture of the user background, viz., high school students; and research results related to adolescent behavior in interacting with interactive systems. Test scenarios had also been determined to help evaluators in exploring the system and direct the focus of evaluation [13]. The scenario includes several tasks to explore application functions, i.e.: register, login, study, read articles, and log out. The evaluators were asked to access the material and exercises through Ruang Guru mobile application. The specified material and practice are mathematics for 12th grade of senior high school; chapter: (i) two-dimensional shapes, (ii) three-dimensional shapes, and (iii) descriptive statistics.

**Table 1.** The Usability and Instructional Design Heuristics for Evaluation of E-Learning Programs

<table>
<thead>
<tr>
<th>Heuristics</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visibility of System Status</td>
<td>H1</td>
</tr>
<tr>
<td>Match Between System and The Real Word</td>
<td>H2</td>
</tr>
<tr>
<td>Error Recovery and Exiting</td>
<td>H3</td>
</tr>
<tr>
<td>Consistency and Standards</td>
<td>H4</td>
</tr>
<tr>
<td>Error Prevention</td>
<td>H5</td>
</tr>
<tr>
<td>Navigation Support</td>
<td>H6</td>
</tr>
<tr>
<td>Aesthetic</td>
<td>H7</td>
</tr>
<tr>
<td>Help and Documentation</td>
<td>H8</td>
</tr>
<tr>
<td>Interactivity</td>
<td>H9</td>
</tr>
<tr>
<td>Message Design</td>
<td>H10</td>
</tr>
<tr>
<td>Learning Design</td>
<td>H11</td>
</tr>
<tr>
<td>Media Integration</td>
<td>H12</td>
</tr>
<tr>
<td>Instructional Assessment</td>
<td>H13</td>
</tr>
<tr>
<td>Resource</td>
<td>H14</td>
</tr>
<tr>
<td>Feedback</td>
<td>H15</td>
</tr>
</tbody>
</table>
Referring to Nielsen’s heuristic evaluation procedure, two stages of inspection were conducted. The first inspection was an examination of the user interface by each evaluator, individually. This stage is important to ensure an independent and impartial evaluation of each evaluator. During the session, the evaluator was guided and was explained to the system mechanism and operation. This first stage lasts about one to two hours for each evaluator. Based on the inspection, the evaluators were asked to submit or write down the problem found and then assessed the severity rating of the problem. Table 2 is a reference for scoring the severity ratings.

The second inspection was a consolidation of individual inspection results aimed to discuss differences between evaluators. There might be the same problems found or the same severity rating score given, but differences or even conflict with each other are very possible. Each evaluator was asked for their opinion regarding the problems and the severity rating given by the other evaluators. The severity rating scoring will be determined by calculating the average scores given by each evaluator. Besides, evaluators were also asked for their opinions on potential solutions.

The consolidation generates a list of problems. Each problem will be categorized into e-learning heuristics. The list of problems will be sorted by the average of the severity rating score, from the highest to the lowest score. The higher the score, the higher the priority of the problem that must be fixed immediately.

4 Results
28 problems were found on the results of expert consolidation from testing with predetermined scenarios. At the first inspection, a total of 39 problems were gathered from all evaluators. Of the 39 problems, there were 11 similar problems which were found by more than one expert. Table 3 lists the problems found severity rating score and heuristic classification.

The next paragraph will discuss the details of each problem per heuristic category.

Problems related to the first heuristic, visibility of system status, were related to the absence or unclear status when the system is loading (problem code: P2) and when the user successfully logged in (problem code: P18). The system must always provide information to users about what is happening in the system.

The second heuristic, the match between the system and the real world, the problem was the writing of decimal numbers in statistical data (problem code: P23). In Indonesia, writing decimal numbers usually separated by commas. The decimal writing needs to be considered well because it might confuse the user.

The third heuristic, error recovery and exiting, related to issues when a user fails to log in. The system did not clearly show the reason why the user failed to login (problem code: P14). When the login fails, the system only informed that the
email and password do not match. The information should be more specific whether it is the wrong email or the wrong password. As a result, the user tends to choose the “forget password” option and the recovery email will be sent even though the email has not been registered. (problem code: P6).

The problem related to the fourth heuristics, consistency, and standards, was the multiple login and authentication required that might make the user annoyed. The user must enter an email when writing comments (problem code: P7) and the user must match complex captcha multiple times (problem code: P8). If the user has logged in at the beginning of entering the application, it should not require the user to re-login when writing comments. The captcha should also be raised only when specific or important authentication is needed.

Problems of the sixth heuristic, navigation support, related to navigation problems. There was no search column so it is difficult to find a chapter or a content (problem code: P9). When choosing an article, navigation did not match the content chosen (problem code: P16). The page navigation also did not appear when selected with an arrow (problem code: P19). Also, in sorting the list, no information appeared regarding what was sorted (problem code: P20).

The seventh heuristic issue, aesthetics, is the problem of visual design and layout. There was too much space in search results (problem code: P21) which might create an inconsistent and confusing impression. The use of thumbnail images is also considered inappropriate (problem code: P27) because it only used one type of image for all uploaded material.

Problems in the ninth heuristic, interactivity, related to the content of the material. There was a problem related to the absence of the explanation of the variable (problem code: P4), which should be present in each material, not just at the beginning of the material. The provision of variable information could minimize the user’s cognitive load in remembering information. Problem also found in the “reading” feature, the content was dominated by text (problem code: P24). Experts believe that users might expect more interactive content. Many writing errors were also found in the article (problem code: P17). There is should be a writing check before publishing the article.

The problems with the tenth heuristic, message design, was related to the layout of the content. There was an incomplete and truncated material title (problem code: P3). The title must be clear and stand out. It also should be grammatically correct. The table on the article was not legible because it is in the form of an image, so the user needs to zoom in on the table, but when enlarged the image is broken (problem code: P12). Another problem was when the user finished uploading the note. The success alert was not visible so the user might not be aware. Besides, when viewing the note, the image had a different position and cannot be rotated (problem code: P28). The image should be automatically set to the right orientation.

The thirteenth heuristic issue, instructional assessment, is related to the type of assessment. Quiz questions only stimulate the user to calculate numbers (problem code: P10). Whereas, experts think it is better if the questions are a story quiz. The expert also suggested that quiz answers can be copied to other applications (problem code: P22).

Problems related to the fifteenth heuristic, feedback, related to the actions the user must take after completing the assessment. The user had to choose a specific number to see the answer (problem code: P5). Besides, there were no markings on the number of questions that had been answered (problem code: P13) and articles that had not been commented (problem code: P25), so users may not know which questions have not been answered and which articles have not been commented. These problems required more effort and time for the user to get to be used to this mechanism. Interactions should be made more intuitive and simple.

A problem was not quite right when classified as e-learning heuristics. There is a problem with the subject sorting feature (problem code: P26). The list of subjects should be sorted alphabetically or can be customized according to the user’s desire. Users may not access all subjects, so users should be able to arrange the order of subjects based on frequency. This problem is arguably more likely to violate one of Nielsen’s heuristics: user control and freedom.

Other problems found was not to be classified as usability or instructional design problem. The problem with the highest average severity rating score (score 3) was a problem related to system security. The experts mentioned that in the account registration process, there needed to be authentication to ensure the authenticity of the account owner’s data (problem code: P1). Another problem was when the Facebook application is logged in, the Ruang Guru application will also log in with the same account that is currently logged in on Facebook (problem code: P15). The system should provide a choice of which Facebook account will log into the Ruang Guru mobile application.

The problems found were very diverse in severity rating. The severity rating score is given range from 1 to 3. A score of 3 means a major problem that is should be a high priority to be fixed immediately, and a score of 1 means only a cosmetic problem that does not need immediate improvement. The calculation of the average severity rating of all problems results in a score of 2.21. So it can be concluded that the average severity rating of all the problems is a minor problem, which means the problem was given a low priority of improvement.

5 Discussion
The fifteen sets of usability and instructional design heuristics for e-learning are enough to cover the scope of the problem findings in this evaluation. The heuristic evaluation method indeed is not limiting the scope of the findings. The protocol allows for formulating sets of heuristics according to the needs and objectives of the evaluation [10]. Of the 28 problem findings, only three problems are beyond determined heuristics. The scope of the problem finding seems to be influenced by the experience and expertise area of the expert.

The average severity rating of all problems was a minor problem that has a low improvement priority. Only minor problems found because this application has been released since
Learning Experience Evaluation of Indonesia Online Tutoring Mobile Application using Heuristic for E-learning

2016 and went through a lot of improvements. Heuristic evaluation is ideally conducted during the system development process before the system is released [14]. Heuristic evaluation conducted on the development phase is expected to find more major problems.

Heuristic evaluation of this online tutor platform had been conducted on limited scenarios. If other evaluations are carried out with different scenarios, the results are very likely to find very different problems.

Heuristic evaluation is a complement method [15]. The heuristic evaluation alone is not sufficient to evaluate e-learning [13]. The problems found in the heuristic evaluation may not be the real problems faced by users [12]. It is required to conduct a user-based evaluation, by involving real users to evaluate user performance.

REFERENCES


