Abstract — The interface of the automatic answer assessment system is plagued by several issues, including an unfamiliar layout, unresponsive design, inconsistency in elements, and a lack of clarity in presenting information. These problems significantly hinder the overall user experience. As a result, this study aimed to address these challenges by designing a user-centered experience for the automatic answer assessment system, using a high-fidelity prototype tailored to meet user needs. The user-centered design method involved four crucial stages: Specify The Context Of Use, Specify Requirements, Create Designs, and Evaluate Designs. Through rigorous usability testing with teachers, the design achieved an impressive effectiveness rating of 90%, firmly establishing it as a "very effective" solution. Additionally, it demonstrated high efficiency with a value of 0.01307 goals/sec, and teachers expressed positive feedback, confirming the satisfaction and usability of the new interface. Similarly, students' usability testing yielded noteworthy results, with a 90% effectiveness rating, also classified as "very effective." The interface showcased a high level of efficiency, with a value of 0.0849 goals/sec. While the satisfaction value fell below the PSSUQ norm, students still found the interface to be user-friendly and satisfactory. Furthermore, the user experience testing, utilizing the UEQ-S, provided valuable insights. For teachers, the pragmatic aspect scored 1.85, the hedonic aspect scored 2.33, and the overall aspect received a commendable score of 2.09, all of which fell within the excellent category on benchmarks. Similarly, students' ratings were highly positive, with scores of 2.14 for both pragmatic and hedonic aspects, and an overall score of 2.14, signifying an excellent user experience. The retrospective think-aloud validation test reaffirmed the positive response from prospective users. Overall, this research, employing a user-centered design approach, successfully delivered a highly satisfactory and effective user experience for both teachers and students using the automatic answer assessment system.

Keywords: User Experience, User Centered Design, Usability Testing, UEQ-S, Retrospective Think Aloud.

Intisari - Antarmuka sistem penilaian jawaban otomatis terganggu oleh beberapa masalah, termasuk tata letak yang tidak biasa, desain yang tidak responsif, elemen yang tidak konsisten, dan kurangnya kejelasan dalam menyajikan informasi. Masalah-masalah ini secara signifikan menghambat pengalaman pengguna secara keseluruhan. Akibatnya, penelitian ini bertujuan untuk mengatasi tantangan tersebut dengan merancang pengalaman yang berpusat pada pengguna untuk sistem penilaian jawaban otomatis, menggunakan prototipe fidelitas tinggi yang disesuaikan untuk memenuhi kebutuhan pengguna. Metode desain yang berpusat pada pengguna melibatkan empat tahap penting: Tentukan Konteks Penggunaan, Tentukan Persyaratan, Buat Desain, dan Evaluasi Desain. Melalui pengujian kegunaan yang ketat dengan para guru, desain tersebut mencapai peringkat keefektifan yang mengesankan sebesar 90%, menjadikannya sebagai solusi yang "sangat efektif". Selain itu, ini menunjukkan efisiensi tinggi dengan nilai 0,01307 gol/detik, dan guru menyatakan umpan balik positif, mengonfirmasi kepuasan dan kegunaan antarmuka baru. Demikian pula, pengujian kegunaan siswa menghasilkan hasil yang patut diperhatikan, dengan peringkat efektivitas 90%, juga diklasifikasikan sebagai "sangat efektif". Antarmuka menampilkan tingkat efisiensi yang tinggi, dengan nilai 0,0849 gol/detik. Sementara nilai kepuasan berada di bawah norma PSSUQ, siswa masih menemukan antarmuka yang ramah pengguna dan memuaskan. Selain itu, pengujian pengalaman pengguna, dengan memanfaatkan UEQ-S, memberikan wawasan yang berharga. Untuk guru, aspek pragmatis mendapat skor 1,85, aspek hedonik skor 2,33, dan aspek keseluruhan mendapat skor terpuji 2,09, yang semuanya masuk dalam kategori sangat baik pada tolak ukur. Demikian pula, peringkat siswa sangat positif, dengan skor 2,14 untuk aspek pragmatis dan hedonis, dan skor keseluruhan 2,14, menandakan pengalaman pengguna yang sangat baik. Tes validasi pemikiran keras retrospektif menegaskan kembali tanggapan positif dari calon
INTRODUCTION

The development of technology encourages various software products to emerge because it can help human activities. However, software products also often fail because they do not match the needs of the user [1] and do not provide comfort to the user when interacting [2]. The process of meeting the needs and comfort of the software cannot be separated from the role of a system interface [3]. The system interface is a part of a computer that acts as the gate of communication between humans and software products [4]. The system interface can make the system more interactive, informative and increase the effectiveness and efficiency of human work [5]. However, the interface made often does not match the needs and does not pay attention to the user experience of the user. Likewise with the interface on the automatic answer assessment system [6].

Based on the results of the pre research survey to the instructor and students it is known that there are several interface problems they face in the interface on the automatic answer assessment system such as, layout or layout interface that are not familiar [7]. This causes users to feel foreign to the system display so that it takes more time to learn it. Another problem faced is the lack of information and instructions listed on the system interface [8]. Lack of information and instructions results in users feel confused and difficulty in understanding the scheme or system workflow [9]. In addition, users also find interface designs that are made unresponsive when accessed through a smartphone device. This makes it difficult for users to operate it because of differences in display size. In addition, the lack of interactive interface often makes users feel that the system becomes unattractive so it feels faster to feel bored [10]. The other problems faced by users are unclear and inconsistent designs such as the use of the color and size of the text that changes. Some of these problems greatly affect the user’s user experience level on an automatic answer assessment system [11].

Based on these conditions it is necessary to do a process of studying the character and point of view of the user to produce a system interface that suits the needs and pay attention to the user's user experience. The process is called the design of the user experience. The design of the user experience is done by extracting problems, emotions, needs and desires of users of the system that has not been clearly defined [12]. Design of User Experience for the Automatic Answer Assessment System is carried out with the aim of understanding the user, so that it requires an approach that focuses on the needs and problems of the user [13].

There are several approaches that can be used to design user experiences such as Activity Centered Design (ACD), Goal Directed Design (GDD), and User Centered Design (UCD). However, in this study researchers used the UCD method because the UCD method involved suggestions and input from users from the initial stages to the end. This is not done in the Activity Centered Design (ACD) method which only focuses on the activities to be carried out by the user [14] and the Goal Directed Design (GDD) method which only focuses on the purpose of making the product [15][16]. In addition, the process of design user experience also through the evaluation stage. This is needed to know the level of user experience obtained by the system interface. [17] The User Experience evaluation process at the interface can be done using several ways such as FGD, eyetracking, interviews to the dissemination of the questionnaire [18]. In this study the authors used questionnaire distribution techniques because they were quantitative and could be measured statistically. This cannot be done if using FGD techniques, eyetracking and interviews that are descriptive qualitative. In evaluating the design of the user experience with the questionnaire dissemination technique, the authors use the user experience quisionare-snort (UEQ-S) method [19]. The reason for using the UEQ-S method is because this method is more concise than the UEQ method. This is because the UEQS method measures 2 large parameters of user experience to achieve the objectives of the use of the product and the emotional experience he feels.

LITERATURE REVIEW

In conducting the research, the author refers to several related studies to be used as references, learning materials, and comparisons.

In 2018, Irma Rofni Wulandari and Lilis Dwi Farida conducted a study on User Experience Measurement in E-Learning in the University Environment Using User Experience Questionnaire.
carried out to see research on user experience design that had existed before to be used as a reference and reference in this study. The literature study material used in this study is in the form of 6 previous research journals which are summarized.

Next, the user experience design is carried out using the user centered design method. When the design of the user experience using the UCD method is in accordance with the requirements, a discussion will be carried out related to the results obtained. Finally, conclusions will be drawn on the results and discussion that have been obtained to find out the main results of the research that has been carried out Figure 1.

User-centered design revolves around meeting the needs, preferences, and behaviors of users. By prioritizing these aspects, products and services are tailored to cater to their specific requirements, resulting in a delightful and gratifying user experience. This enhanced experience, in turn, leads to higher user engagement and fosters a loyal user base. The user-centric approach creates a positive cycle of improvement, with satisfied users becoming advocates and drawing more users to the user-friendly and personalized product or service [20]. Furthermore, Dzakkilah Salma Wachid, Satrio Hadi Wijoyo, and Andi Reza Perdanakusuma conducted a study on the User Experience Design of E-Learning on the SMAN 13 Surabaya website using the Human-Centered Design (HCD) approach in the past year [9]. This research involved designing an e-learning prototype and conducting testing using the UEQ method. The researchers analyzed user and system requirements, created user journey maps, information architecture, wireframes, and prototypes to design the system interaction. The results showed that all measurement scales were categorized as very good, except for two scales: clarity, which was above average, and novelty, which was categorized as good.

**MATERIALS AND METHODS**

This research is a research that uses qualitative and quantitative approaches. This research begins with the process of identifying learning site interface problems in the network which are the background to the problems in this research. Then a literature study process was carried out to see research on user experience design that had existed before to be used as a reference and reference in this study. The literature study material used in this study is in the form of 6 previous research journals which are summarized.

Figure 1 Methods Research

Specify the context of use in user centered design is done to identify the problem and study the point of view and character of potential users. This stage needs to be done to find out the problems that can be solved as well as the goals and targets that the user wants to achieve while using the system being developed. Empathy maps and user personas as output from the process of studying users [21].

Specification requirements in user centered design are carried out to identify system features according to the needs of each prospective user [22]. This stage needs to be done because it acts as a reference in the solution ideation stage. In identifying system features that suit the needs of each user of the automatic answer scoring system, several activities are carried out.

Create design solutions in user centered design aims to design product solutions referring to the idea analysis process that was carried out in the previous stage. The result of this stage will produce a prototype that can be tested at a later stage. In producing a solution design in the form of a design for an automatic answer scoring system.

Evaluate design on user centered design is the final stage carried out in this study. Evaluate design is the testing phase of the high-fidelity prototype design that was made in the previous stage [22]. The evaluate design stage was carried out by teaching respondents and teacher respondents with the process of usability testing and distributing questionnaires. The number of test respondents is not less than 20 people from each prospective user based on usability testing rules [23]. The evaluate design stage in this study was carried out by measuring two aspects, namely usability and the level of user experience.
High fidelity prototype is tested by using usability using usability testing to measure effectiveness, efficiency and satisfaction [24]. The effectiveness aspect is measured using the completion rate metrics approach. The efficiency aspect will be measured using a time-based efficiency approach. The aspect of satisfaction is measured using the post study system usability questionnaire (PSSUQ) approach [25]. Test will be carried out to measure the level of user experience using the user experience questionnaire-short (UEQ-S) method and retrospective think aloud.

The research data used in this study were interview data and observations with 10 teachers and 10 students. The other data is usability test data and user experience level as well as data from literature studies which are used as references and references in writing the final project report.

<table>
<thead>
<tr>
<th>Table 1 Teacher User Interview questions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Point</strong></td>
</tr>
<tr>
<td>User knowledge of the automatic answer scoring system (submitted before observation)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>User experience in using an automated scoring system (submitted after observation)</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

The process of collecting data in this study will be carried out in two ways, namely interviews and observation (Table 1). First of all, an initial interview will be conducted to find out the user's knowledge of the automatic answer scoring system. Interviews are used to find out the user's experience after using the platform and find out the constraints and features needed when using the product.

**RESULTS AND DISCUSSION**

specify the context of use stage, it is known that the prospective users of the automatic answer scoring system in this study are teachers and students. Furthermore, the process of collecting data through interviews and observations is carried out to find out knowledge, experience, problems faced and the needs of prospective users in an automatic answer scoring system for exams. The process of collecting data was carried out together with 10 teaching respondents and 10 student respondents.

The results were carried out with a brainstorming process in identifying features that could be applied to the system interface to answer the problems faced by users. The brainstorming process is carried out by answering the question how might we from user problems.

Create a system interface equipped with:

1. Icons and images complete with writing
2. Additional information as instructions. It can be solved in several ways, such as:
3. Provide class room features that are categorized for each subject.
4. Organized a download feature for the recap of the entire exam in the subject class room.
5. Create a website-based internal automatic answer assessment system that is equipped with a login system for registered students.
6. Holds a description of the remaining time for the exam on the exam execution page.
7. Hold a question work review feature on the exam page to see the number and description of questions that have been done during the exam.

The formulation of solutions and the discovery of design insights that were obtained in the previous stage were then used as a reference in designing the final design of the automatic answer scoring system interface in this study. In order to simplify the design process, first an information architecture diagram is formed which is used to determine the navigation path of the system (Figure 2).

Based on the navigation path in the information architecture diagram of the automatic answer scoring system, a low fidelity design is designed for each page in the diagram. Low fidelity design is a rough design of a website or application that is simple and not detailed. Low fidelity design needs to be made because design changes will be easier and faster to do if it is still in concept form (Figure 3).

The low fidelity design of the automatic answer scoring system above is designed with a high-fidelity prototype. High fidelity is the final solution in the form of a system interface design that is very similar to the appearance of the system to be
developed. High fidelity in this study was formed into an interactive design by turning it into a prototype (Figure 4, Figure 5). The color that used in high fidelity prototype according to the colour of Tut Wuri Handayani (The logo of Ministry of Education, Culture, Research and Technology of the Republic of Indonesia).

The design evaluation process was carried out by testing the high fidelity prototype to each of the 40 teachers and 40 students. Design evaluation testing is carried out using a usability approach to measure the effectiveness, efficiency and user satisfaction in using the system. The evaluation process is carried out with several scenario tasks that must be completed by each user (Table 2).

Usability testing high fidelity prototype interface automatic answer scoring system. First, the user is asked to access the interface design prototype on the maze.design website. The maze.design website will record the direct success of the task and the execution time of the task automatically to be used in calculating the value of effectiveness and efficiency. Furthermore, users are asked to fill out satisfaction questionnaires that have been designed using the PSSUQ method to calculate user satisfaction scores.

Table 2 Usability Testing Scenario Design Prototype

<table>
<thead>
<tr>
<th>Task</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create a new test draft</td>
<td>User is asked to create a new exam draft and enter the exam draft page</td>
</tr>
<tr>
<td>Create exam questions</td>
<td>Users are asked to create exam questions. The process of making exam questions is carried out with the type of essay questions and multiple-choice questions.</td>
</tr>
<tr>
<td>Preview exam questions and publish them</td>
<td>User is asked to preview the exam draft he has made. Then carry out the exam publication process and set the start and end times of the exam</td>
</tr>
</tbody>
</table>

Based on the results of calculating the level of efficiency using the completion rate approach, the percentage of effectiveness for teaching users is 90% and is included in the very effective category according to standard effectiveness measures. The percentage of the same effectiveness value was also obtained by student users, namely 90% in the very effective category.

Based on the results of calculating the efficiency level using the time-based efficiency approach, the efficiency value for teaching users is 0.1307 goals/sec and is included in the very fast category according to interval time behavior standards. The efficiency value for student users is 0.0849 goals/sec with a very fast category according to the standard interval time behavior.

Based on the calculation of the results of the PSSUQ questionnaire, the final value of each PSSUQ variable was obtained for the two potential users. The final value of each of these variables is then compared with the PSSUQ standard norm. As for the comparison graph of the PSSUQ value of the automatic answer scoring system with the PSSUQ norm (Table 3).

Comparison graph of teacher UEQ-S results with UEQ-S benchmarks. Based on the figure, the
UEQS results belong to the excellent category on benchmarks. This shows that the high-fidelity prototype design on the teacher’s side already has a very good level of user experience (Figure 8).

The results are then compared with the UEQ-S benchmarks that have been set. UEQS results belong to the excellent category on benchmarks. This shows that the high-fidelity prototype design on the student side also has a very good level of user experience.

The final score for each PSSUQ variable of the automatic answer scoring system is below the value of each variable at the lower limit, average and upper limit of the PSSUQ standard norm. These results indicate that the high-fidelity prototype of the automatic answer scoring system tested has provided satisfaction to its users (Figure 8).

<table>
<thead>
<tr>
<th>Var</th>
<th>User</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>Student</td>
<td>1.83</td>
</tr>
<tr>
<td>System Usability</td>
<td>Teacher</td>
<td>1.81</td>
</tr>
<tr>
<td>Information Quality</td>
<td>Student</td>
<td>2.00</td>
</tr>
<tr>
<td></td>
<td>Teacher</td>
<td>1.94</td>
</tr>
<tr>
<td>Interface Quality</td>
<td>Teacher</td>
<td>1.95</td>
</tr>
<tr>
<td></td>
<td>Teacher</td>
<td>1.91</td>
</tr>
<tr>
<td></td>
<td>Student</td>
<td>1.38</td>
</tr>
<tr>
<td></td>
<td>Teacher</td>
<td>1.83</td>
</tr>
</tbody>
</table>

Testing and filling out the questionnaire 10 teaching respondents and 10 student respondents were interviewed to find out their experiences after using the prototype system. Based on the results of the interviews conducted using the three questions described in subchapter 3.3.4, all respondents were satisfied with the high-fidelity prototype being tested because it met their needs to conduct or take online exams. In addition, all respondents stated that there were no deficiencies in the form of features or interface deficiencies.

CONCLUSION

The usability testing conducted with both teachers and students yielded excellent results, showing a remarkable 90% effectiveness rating, which falls into the “very effective” category. The efficiency values were also impressive, with 0.1307 goals/sec for teachers and 0.0849 goals/sec for students, both categorized as “very fast.” Although the satisfaction scores measured using the PSSUQ method were below the expected norm for both groups, the high-fidelity prototype design for the teacher side managed to achieve a satisfactory level of satisfaction. It provided a well-designed interface with high-quality information, which positively impacted the user experience. The user experience testing, employing the UEQ-S method, delivered exceptional results. For teachers, the pragmatic scale scored 1.85, and the hedonic scale scored 2.33, resulting in an overall score of 2.09, all of which fall into the Excellent category. For students, the pragmatic scale scored 2.14, and the hedonic scale scored 2.14, resulting in an overall score of 2.14, which also falls into the Excellent category.

In conclusion, the high-fidelity prototype for the teacher side demonstrated a highly positive user experience, despite the slightly lower satisfaction values. The success of the user-centered design approach is evident through the impressive usability and user experience testing results, ensuring an effective and enjoyable experience for both teachers and students.
REFERENCE


