Heuristic Evaluation and Usability Testing

Saki Takizawa
Earlham College
Richmond, Indiana, usa
stakiz21@earlham.edu

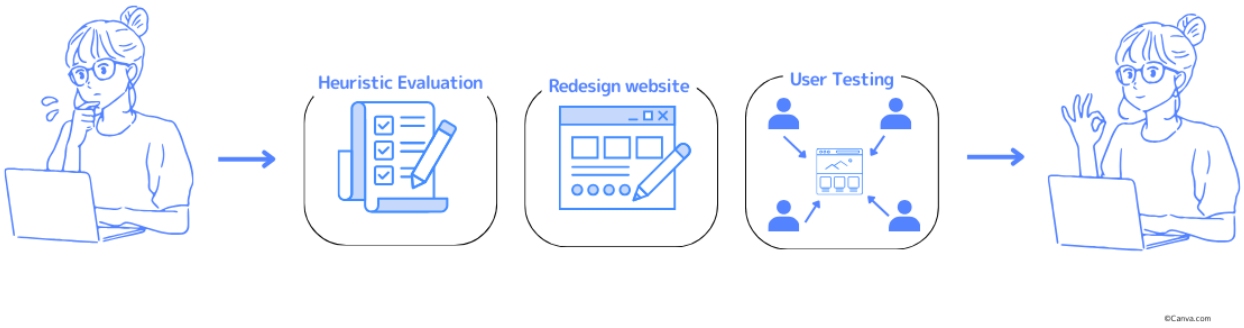


Figure 1: Graphical Abstract

1 ABSTRACT

User experience design (UX) is a critical approach to web design that emphasizes both the aesthetic and functional aspects of a website. However, not all websites possess a well-designed UX. Heuristic evaluation and usability testing serve as two valuable methods to enhance it. The evaluations in this paper aim to identify usability issues on a college website and redesign based on the problems found.

This paper will employ two research methods, heuristic evaluation and user testing, to identify and sort out problems that affect the user experience of a college website. The heuristic evaluation will be conducted referencing the Nielsen Norman Group Heuristic Evaluation Workbook on the tasks of specific user groups: current computer science majors, prospective computer science students [8].

It will uncover points where there might be potential usability issues. Following this step, the college website design will be revised based on the guidelines created by the Research-Based Web Design and Usability Guidelines provided by the US Department of Health and Human Services[6]. When comparing the original and new website, potential users will be asked to perform tasks for user testing. Ultimately, the results and analysis from the user testing will clarify whether usability issues were resolved and specific areas of the website that need improvement, if there are any. The overarching goal is to enhance the website's usability, making it more user-friendly and enjoyable.

2 INTRODUCTION

User Experience (UX) has become a central paradigm within the field of Human-Computer Interaction (HCI). The demand for well-designed digital interfaces has been steadily increasing, necessitating designs that are not just highly functional but also intuitive, enjoyable, and effective in meeting the diverse needs of users. Within this context, the evaluation of user experience is important as it involves understanding the efficiency and user-friendliness of digital interfaces. One of the valuable methodologies for assessing the usability of a website is heuristic evaluation. According to Nielsen, "Heuristic evaluation is a usability inspection method for computer software that helps to identify usability problems in the user interface (UI) design. It involves having a small set of evaluators examine the interface and judge its compliance with recognized usability principles (heuristics) [8]."

In addition to heuristic evaluation, user testing engages actual users in interacting with a system and provides a holistic view of usability. This research examines the intersection of UX design, heuristic evaluation, and user testing, with a focus on the usability of a website. College websites are vital information resources for college students, faculty, prospective students, and even more, so all the information needs to be the latest. Furthermore, the design should be something for the sake of those user groups given that users are a specific group, such as current students and faculty. However, the computer science website at Earlham College is not operated by UX professionals, as is the Earlham College website. Therefore, there is a need for improvement in UX perspectives.

In this research, I will begin by applying heuristic evaluation to the Earlham Computer Science Department's web page, referencing Nielsen's heuristic evaluation principles and the Research-Based

Web Design and Usability Guidelines provided by the US Department of Health and Human Services [6][8]. Subsequently, I will catalog the problems identified in the website and rank them based on the severity of the design issues [3].

The next step involves conducting user testing that will be implemented based on the heuristic evaluation results. Following this, user testing will be conducted for specific user groups, current computer science major students, and students who are considering majoring in computer science in this research. Each user will be asked to perform some tasks on one of the two website versions, the original and the redesigned. Finally, the data and its analysis by incorporating Analysis of Variance (ANOVA) will identify if usability issues were resolved and specific website areas require improvement.

This research stands out for its unique combination of several different evaluation methods, a focused analysis specialized on the domain of college websites, and a data-driven redesign strategy. They all aimed at significantly enhancing the user experience for college website users.

3 BACKGROUND

Heuristic evaluation and user testing play key roles in UX design. UX design is an approach that aims to produce a better user experience. It encompasses the entire product creation process, including aspects of branding, design, usability, and function [1]. UX design involves two elements: how a person perceives and how a product is used. Both of these elements depend on each other in UX design.

Perception in UX design encompasses visual aspects, such as layout and the use of colors, emotional responses, such as how users rate their overall experience with the website, and comprehension, such as how easily users can understand the website's layout. Additionally, usage in UX design, focusing on how a product is used, includes navigation, such as how easily users can move from one section to another on the website, functionality, such as the performance of functions like buttons, and task efficiency, assessing how effectively users can find the information they were seeking.

These elements are intricately linked in UX design, highlighting the importance of designing products with UX principles that effectively capture and respond to human emotions and experiences. Adhering to these principles is essential for creating products that meet user's needs and preferences.

3.1 Heuristic Evaluation

Heuristic evaluation is a usability evaluation method that involves a systematic inspection of a user interface by usability experts against a set of predefined usability principles. The evaluators identify usability problems, such as violations of these principles, and provide recommendations for improving the interface's overall usability. The aim of heuristic evaluation is to uncover usability issues quickly and cost-effectively, ensuring that the interface is more user-friendly and efficient in meeting the needs of its intended users. Thus, heuristic evaluation is a way to test a user interface for usability problems by comparing it to a set of best practices.

Heuristic evaluation is a useful method for identifying usability problems, but it is important to be aware of its limitations [5]. Experts may identify problems that users do not encounter, or they may miss problems that users are having. This is because heuristic evaluation is based on experts' knowledge and experience, which may not always be aligned with the needs of the target users. In addition, heuristic evaluations are typically performed in a laboratory setting, which may not be representative of how users interact with the product in the real world. Finally, heuristic evaluations can be subjective, and different experts can identify different problems. To mitigate these limitations, it is important to use heuristic evaluation in conjunction with other usability testing methods, such as user testing.

3.2 Usability Testing

Usability testing, particularly in the realm of UX (User Experience), is a method used to assess the ease of use of a web application. This assessment is carried out by instructing users to perform specific tasks with the application. Usability testing can take various forms, including in-person sessions in which a researcher observes users completing tasks, and remote testing, which offers a more accurate perspective on real-world user experiences [4].

Compared to heuristic evaluation, usability testing tends to find out more significant issues. One distinct advantage of usability testing is its ability to pinpoint problems that could affect actual users of the application, without the need for presorting or filtering these issues based on their perceived impact. The test itself helps in assessing the impact of identified problems. This has been consistently demonstrated in the Jeffries and Desurvire studies[7], where nearly all problems identified through usability testing proved to be of above-average severity. In addition, the problems that usability tests find are usually not found in other methods, such as heuristic evaluation, because the variation of the actual user's actions from users exceed the level of expectation of experts in most of the cases [7].

In the case study at Boğaziçi University, one of the user groups that the researchers carried out the users test on the undergraduates at Boğaziçi University. The ten tasks were like following:

(1) Explore the website of the laboratory related to Flexible Automation. Guideline 10.4: Avoid misleading cues to click [6].
(2) Locate the contact number for the chairman of the Industrial Engineering(IE) department. Guideline 2.5: Design for working memory limitations [6].
(3) Find the webpage for the Office of International Relations to gather information about Erasmus or Exchange programs. Guideline 10.4: Avoid misleading cues to click [6].
(4) Search for the courses taught by IE Professor Prof. Dr. Barbarosoğlu. Guideline 10.4: Avoid misleading cues to click [6].
(5) Navigate to the webpage of the "Quantitative Finance Research Group" within the Boğaziçi University Industrial Engineering Dept. Guideline 10.4: Avoid misleading cues to click [6].
(6) Locate the information page to determine if PSY 101 is available as an HSS elective. Guideline 16.2: Structure each content page to facilitate scanning [6].

(7) Find the Alumni list of the IE Department to connect with fellow graduates after completing your studies. Guideline 10.4: Avoid misleading cues to click [6].
(8) Find the contact number for Instructor Dr. Yasemin Aksoy to inquire about a specific elective course. Guideline 2.5: Design for working memory limitations [6].
(9) Check if Dr. Suat Genç offers any undergraduate courses as elective options due to his reputed success as an instructor. Guideline 16.4: Group related elements [6].
(10) Discover the list and descriptions of IE-Elective courses available for undergraduate education. Guideline 16.2: Structure each content page to facilitate scanning [6].

Addition to the tasks above, users would be asked about demographics and the frequency of internet usage to see if there is any significant difference in those categories as the case study at Boğaziçi University did.

3.3 Analysis of Variance

After collecting the data from usability testing, Analysis of Variance (ANOVA) will be introduced to investigate the results. ANOVA is used to determine if there is a statistically significant difference between user groups.

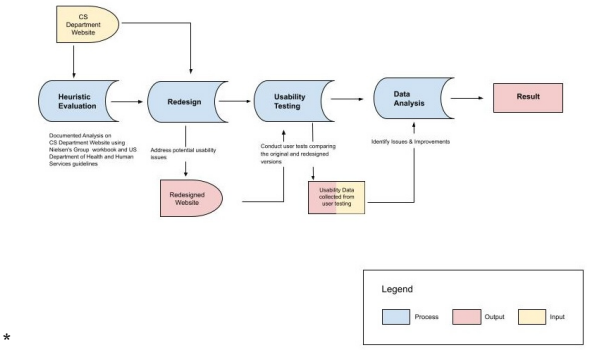
In simpler terms, ANOVA is a statistical test that can be used to compare the means of three or more groups. It is a powerful tool, but it is important to note that it has three key assumptions: the samples must be independent, the variances of the groups must be equal, and the data must be normally distributed.

In the case study, Mahmut Ekşioğlu and colleagues conducted additional comparison tests alongside ANOVA to ensure that their results were reliable [3]. This is a good practice to follow, as it helps to minimize the risk of making false conclusions.

This research used ANOVA, a statistical tool, to dig deeper into how different groups of users interact with the college website. It aims to compare how well different groups perform specific tasks and assess their overall experience. For example, we'll see if there are significant differences in how well different groups, like prospective students, faculty, and current students, complete tasks. Finding these differences can help us identify areas of the website that need to be improved for specific user groups. For instance, if prospective students have much lower success rates in navigating or finding information, it suggests that the website might not be meeting their needs. By using ANOVA to pinpoint these differences, we can make targeted improvements to better address the needs of specific user groups and improve the overall user experience of the college website.

The value of F is a test statistic used in Analysis of Variance (ANOVA) to compare the variance between the mean of the groups to the variance within each group. A higher F-value indicates that the group means are significantly different, suggesting that the observed differences are not due to random chance. The p-value is the probability of observing results as extreme as the data, assuming the null hypothesis (which posits that all group means are equal) is true. A p-value less than 0.05 typically means that the observed differences are statistically significant and the null hypothesis can be rejected. In contrast, a p-value greater than 0.05 suggests that the observed differences are likely due to random chance, meaning that there is no statistically significant difference between the group means. The threshold of 0.05 is commonly used to balance the risk of Type I error (incorrectly rejecting the null hypothesis) at 5%, which is widely accepted in scientific research [10].

4 METHODOLOGY



*

Figure 2: My Research Plan

This research involves four main steps. First, a guideline was created to evaluate the Earlham College Computer Science Department website. This guideline is based on Nielsen's heuristic evaluation principles [9] and Research-Based Web Design and Usability Guidelines. To maintain the website's integrity during the research, a clone of the current version was used. These references identified potential usability issues from a human perspective. Using these guidelines, areas of the website that require improvement or potential redesigns in terms of usability will be identified.

4.1 Heuristic Evaluation Principles

This research carefully selects key principles and guidelines to evaluate the website's usability. These come from two well-respected sources: Nielsen's Heuristic Evaluation Principles and the Research-Based Web Design and Usability Guidelines. To pick the most important ones, we looked for elements marked with stars in these resources. These starred points are considered especially important because they are based on fundamental principles of user-centered design and have a big impact on how people experience the website. By focusing on these crucial points, our evaluation will primarily identify the areas with the biggest potential for improvement. The evaluation guideline will look like the following.

4.2 Selected Heuristic Evaluation Guidelines

The following evaluation guidelines were most relevant based on the major usability issues identified.

(1) **Understand and Establish User's Expectation**
    - Ensure the website includes content that meets user requirements.
(2) **Avoid using color alone to convey call-to-action**
    - Do not rely solely on color to indicate critical actions, ensuring accessibility for all users.
(3) **Cluttered layout is unwanted**

- Create pages that are not visually cramped or overwhelming.

(4) **Clear category labels**
- Ensure all category labels and links accurately reflect their contents.

(5) **Text color and Background**
- Maintain high contrast between text and background for readability; black text on a white background is ideal.

(6) **Speak the users' language**
- Use clear, familiar words, phrases, and concepts suited to the target audience.

(7) **Minimize the user's memory load**
- Provide simple, easily retrievable information so users do not need to memorize steps.

(8) **Consistency**
- Maintain consistency across different pages, actions, and design elements.

(9) **Provide clear exit pathways**
- Allow users to easily backtrack or exit an unintended action or state.

(10) **Provide shortcuts**
- Offer navigation shortcuts to allow users to find information efficiently without excessive steps.

Using the list above, an evaluation will be conducted on the Earlham College CS Department Website [2]. For instance, the homepage contains a section where the text color and background color is not sufficiently contrasted. An item from the list, Text color and Background [6], can be applied to this case.

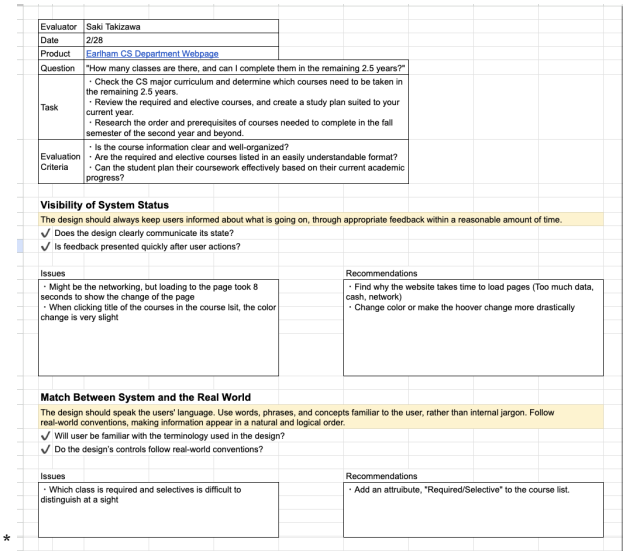
5 RESULT AND ANALYSIS

The heuristic evaluation was conducted on the specific tasks: "How many classes are there, and can I complete them in the remaining 2.5 years?", "What extracurricular activities are available beyond classes?", "What kinds of career paths are available to CS major graduates?", and overall page performance. The process discovered 52 potential usability issues in 10 different sections of the Earlham College computer science department website.

5.1 Heuristic Evaluation Summary

The heuristic evaluation identified 52 usability issues across 10 heuristic categories. Below is a summary of representative issues per category that were prioritized for redesign.

- **1. Visibility of System Status:**
  - Course title color changes are too subtle, making interaction unclear.
  - Iceland Field Study video appears abruptly with no explanation or context.
- **2. Match Between System and the Real World:**
  - It is difficult to distinguish required vs. elective courses at a glance.
  - Career support links redirect externally without explanation or context.
- **3. User Control and Freedom:**
  - No clear way to return to a previous page or exit a process flow.



Evaluator	Saki Takizawa
Date	2/28
Product	Earlham CS Department Webpage
Question	"How many classes are there, and can I complete them in the remaining 2.5 years?"
Task	・Check the CS major curriculum and determine which courses need to be taken in the remaining 2.5 years. ・Review the required and elective courses, and create a study plan suited to your current year. ・Research the order and prerequisites of courses needed to complete in the fall semester of the second year and beyond.
Evaluation Criteria	・Is the course information clear and well-organized? ・Are the required and elective courses listed in an easily understandable format? ・Can the student plan their coursework effectively based on their current academic progress?

Visibility of System Status

The design should always keep users informed about what is going on, through appropriate feedback within a reasonable amount of time.

✓ Does the design clearly communicate its state?
✓ Is feedback presented quickly after user actions?

Issues	Recommendations
・Might be the networking, but loading to the page took 8 seconds to show the change of the page ・When clicking title of the courses in the course lsit, the color change is very slight	・Find why the website takes time to load pages (Too much data, cash, network) ・Change color or make the hoover change more drastically

Match Between System and the Real World

The design should speak the users' language. Use words, phrases, and concepts familiar to the user, rather than internal jargon. Follow real-world conventions, making information appear in a natural and logical order.

✓ Will user be familiar with the terminology used in the design?
✓ Do the design's controls follow real-world conventions?

Issues	Recommendations
・Which class is required and selectives is difficult to distinguish at a sight	・Add an attruibute, "Required/Selective" to the course list.

Figure 3: Heuristic Evaluation Management Excel Sheet

  - Student group contact details are missing, limiting user action.
- **4. Consistency and Standards:**
  - Design varies significantly between pages (e.g., color, layout, font sizes).
  - Buttons have inconsistent hover effects and alignment.
- **5. Error Prevention:**
  - Users are redirected to external pages (e.g., CodePath) without notice.
  - Internal links (e.g., "Our Story") are sometimes broken or incomplete.
- **6. Recognition Rather Than Recall:**
  - There is no FAQ section to help users find common answers.
  - Course requirements are repeated in different places with inconsistent wording.
- **7. Flexibility and Efficiency of Use:**
  - Course list appears as a large unstructured text block without headings.
  - There are no filtering tools for viewing by concentration or requirement.
- **8. Aesthetic and Minimalist Design:**
  - The "Our Story" section is visually cluttered and lacks visual hierarchy.
  - Mismatched images reduce the perceived credibility of the content.
- **9. Help Users Recognize, Diagnose, and Recover from Errors:**
  - Confusing video placement and lack of contact info hinder troubleshooting.
- **10. Help and Documentation:**
  - There is no dedicated FAQ or help page for finding support information.

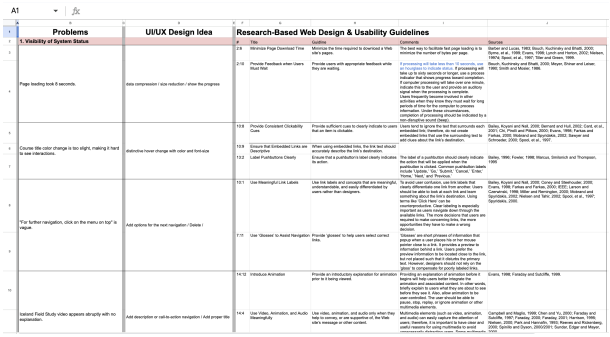


Figure 4: Redesign Guideline

Subsequently, guidelines for redesigning were created from usability issues found in the heuristic evaluation. Redesign ideas for each issue was suggested based on Research-Based Web Design and Usability Guidelines provided by the US Department of Health and Human Services[6].

Research-Based Web Design and Usability Guidelines are ranked by relative importance. Using the ranking, usability issues and redesign ideas were sorted.

5.2 Redesigned Website

Based on the usability issues identified in the heuristic evaluation, two major improvements were made to the website: (1) restructuring the course list and (2) adding a dedicated FAQ page. These changes address specific issues related to visibility, navigation efficiency, content clarity, and support documentation.

5.2.1 Course List Redesign. The original website's course list was presented as an unstructured text block, lacking clear categorization or labels to distinguish between required and elective courses. This created confusion for both current and prospective students trying to plan their academic paths.

The redesigned course list introduces collapsible sections and uses visual tags (e.g., "Required," "Elective") to enhance clarity. Additionally, filtering by concentration and study plans is supported, improving usability and supporting individual learning paths.

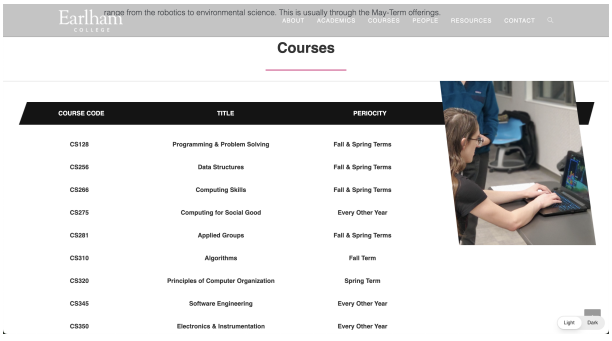


Figure 5: Original Course List Layout

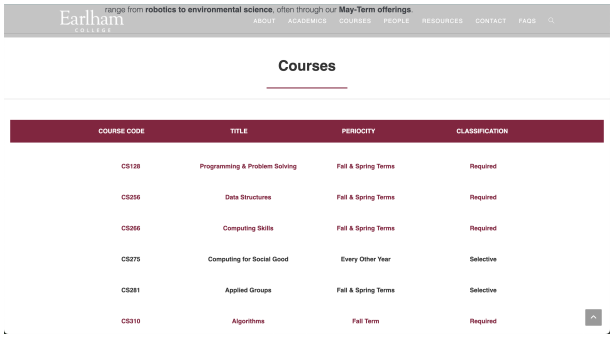


Figure 6: Redesigned Course List Layout

5.2.2 FAQ Page Implementation. One key issue from the heuristic evaluation was the complete absence of an FAQ page, which forced users to navigate multiple sections or rely on external links to find essential information.

In the redesigned website, a comprehensive FAQ section was introduced. This section consolidates commonly asked questions into categorized segments, including course planning, extracurricular activities, contact options, and career support. This reduces cognitive load and provides quicker access to critical information.

Figure 7: No FAQ Page on Navigation Bar on Original Website

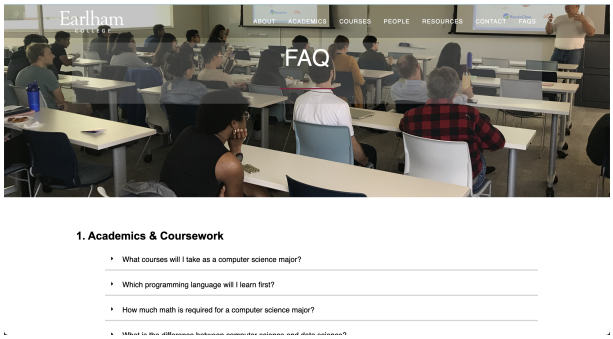


Figure 8: FAQ page on Redesigned Website

These two redesign features are aligned with the most critical usability principles identified, particularly those related to content organization, memory load minimization, and help and documentation support. They were tested during usability testing and showed measurable improvement in task completion time and user satisfaction.

5.3 ANOVA Analysis

The Analysis of Variance (ANOVA) was used to assess whether there was a significant difference in users' ability to find information on the old and new website versions. The results did not show statistically significant differences in mean scores between the two

groups ($p > 0.05$), with an F value of 0.036 and a p value of 0.854. This suggests that while users performed similarly on both versions, there was no clear improvement in task performance between the two versions. However, open-ended responses revealed that users of the redesigned website were more accurate on certain tasks, indicating potential improvements in early-stage task clarity.

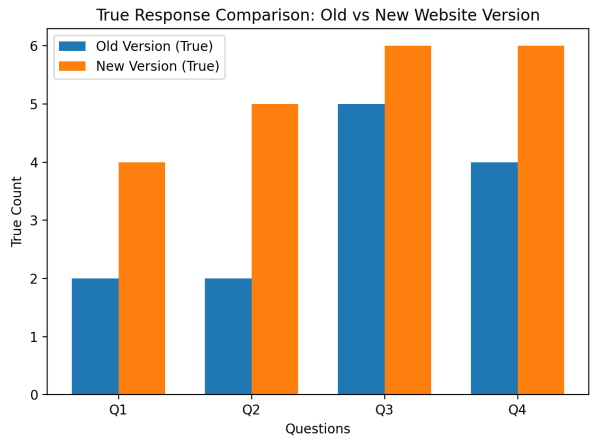


Figure 9: The Response Comparison: Old vs New Version

For open-ended questions, the content of each response was analyzed and categorized as 'TRUE' or 'FALSE' based on whether each response was the same as the predefined correct answer. If the user's response matched the specific criteria for the question, it was classified as 'TRUE'; otherwise, it was categorized as 'FALSE.' The number of correct responses ('TRUE') was then counted and analyzed to assess user accuracy across different tasks and website versions. By quantifying responses in this way, we were able to compare the amount of correct information obtained by users of the old and new versions of the website. The analysis showed a tendency for users of the new version to provide a higher percentage of correct answers for some questions. In particular, the new version outperformed the old version on Questions 1 and 2, demonstrating increased user accuracy in early-stage tasks. Both versions performed equally well on Questions 3 and 4, where most users succeeded. These results suggest that while no statistically significant differences or consistent trends were observed overall, the redesigned interface may have improved task clarity and learnability in the initial stages of interaction.

6 FUTURE WORK

Since most of the participants in this study were computer science majors or students with knowledge in related fields, it is possible that some of the questions were answered based on their original knowledge rather than information obtained from the website. Therefore, future research should further increase the sample size of participants and conduct the evaluation with diverse user groups, including users with backgrounds not related to computer science. By examining the differences in comprehension and operability among different user groups, design improvements are expected to be more clearly identified, leading to a more user-friendly user interface design.

7 ACKNOWLEDGMENT

I would like to express my sincere gratitude to Dr. David Barbella for his invaluable guidance and support during the initial stages of this project, particularly in the preparation and development of the research proposal. I also wish to extend my thanks to Dr. Charlie Peck for his insightful feedback and continuous support, which were instrumental not only in the technical aspects of this report but also in the overall progression of this research. Finally, I am deeply appreciative of Dr. Yunting Yin for her unwavering encouragement and assistance throughout the entire research process, from the setup stages to the completion of this work.

REFERENCES

[1] 2023. User Experience (UX) Design Your constantly-updated definition of User Experience (UX) Design and collection of topical content and literature. https://www.interaction-design.org/literature/topics/ux-design
[2] Earlham CS Department. [n. d.]. Home - Earlham CS. https://cs.earlham.edu/
[3] Mahmut Ekşioğlu, Esin Kiris, Burak Çapar, Murat N. Selçuk, and Selen Ouzeir. 2011. Heuristic Evaluation and Usability Testing: Case Study. *Department of Industrial Engineering, Boğaziçi University, Istanbul-Turkey* (2011).
[4] Interaction Design Foundation. 2016. What is Usability Testing? *Interaction Design Foundation* (2016).
[5] Bendoly Laura Goldman, Kate Haley. 2003. Investigating Heuristic Evaluation: A Case Study. *ERIC* (2003).
[6] Health and Human Services Department. 2006. Research-Based Web Design Usability Guidelines. *Health and Human Services Department* (2006).
[7] Robin Jeffries and Heather Desurvire. 1992. Usability Testing vs. Heuristic Evaluation: Was There a Contest? *SIGCHI Bull.* 24, 4 (oct 1992), 39–41, numpages =. https://doi.org/10.1145/142167.142179
[8] Jakob Nielsen. 1995. How to conduct a heuristic evaluation. *Nielsen Norman Group* 1, 1 (1995), 8.
[9] Molich Rolf and Nielsen Jakob. 1990. Improving a Human-Computer Dialogue. *Commun. ACM* 33, 3 (mar 1990), 338–348. https://doi.org/10.1145/77481.77486
[10] The Pennsylvania State University. [n. d.]. 10: One-Way ANOVA. https://online.stat.psu.edu/stat200/book/export/html/212