Avances en Interacción Humano-Computadora, 2024, Volumen 9, No. 1, pp. 109-116
DOI: 10.47756/aihc.y9i1.155

RESEARCH PAPER

From Transcription to Empathy: Employing Artificial Intelligence Tools in User-Centered Design for an Online Assessment Platform

Miguel Angel Rodríguez-Ortiz, Diana Guadalupe López Avalos, Karina Monserrat Ayala Quiles, Laura S. Gaytan-Lugo

Published: 30 November 2024

Abstract

This study examines the potential of Artificial Intelligence (AI) tools to facilitate User-Centered Design (UCD) in the development of an online assessment platform at a public university in Mexico. Employing the Design Thinking methodology, which incorporates surveys and usability tests during the empathy stage, the study identifies and addresses key usability issues affecting educators' interactions with the platform. These methods yielded comprehensive data on user experiences and challenges, facilitating a profound understanding of user needs and pain points. Subsequently, AI tools, including ChatGPT, were used to analyze this feedback and generate detailed Empathy and User Journey Maps, which revealed significant issues. The findings underscore the necessity for a more intuitive user interface and a reduction in cognitive load. In light of these findings, the study recommends a redesign of the navigation schema and interface with the objective of simplifying tasks and improving efficiency. The combination of Design Thinking with advanced AI technologies enables the effective tackling of complex usability challenges, creating a more efficient and inclusive digital assessment environment. This integration not only enhances the UCD process but also offers deeper insights into user experiences, which ultimately increases the likelihood of adoption and enhances educator satisfaction.

Keywords:

AI-Driven User Experience (UX) Design; Design Thinking; Educational Technology; Empathy Maps; Online Assessment; Usability Testing.

Introduction

EvPraxis, an Online Assessment Platform (OAP) utilized at a public university in Mexico to streamline exam creation and administration, is employed by 15.1% of the professors (340 out of 2250), becoming an integral and significant tool for achieving digital education. In the year 2023 alone, the platform facilitated the delivery of 86,409 examinations [4, 6]. Nevertheless, despite its pivotal role, the OAP continues to encounter recurring usability challenges that impede its full potential.

Previous research by Fajardo-Flores et al. [7] has identified accessibility barriers that diminish user satisfaction and platform efficiency, such as the lack of alternative text for images and the absence of keyboard navigability. In order to address these particular challenges, this research employs the Design Thinking methodology, which is a problem-solving approach that places an emphasis on human-centered solutions through the use of iterative design processes [1].

The primary contribution of this study is to demonstrate how integrating advanced Artificial Intelligence (AI) tools within a User-Centered Design (UCD) framework, guided by Design Thinking, can address the complex usability challenges affecting educators' interactions with the OAP. By systematically identifying particular usability issues and employing AI-driven techniques, such as ChatGPT, to refine user feedback analysis this research not only enhances the platform's interface and functionality but also provides a replicable framework for similar educational technologies, deepening our understanding of user needs and experiences.

Improving the usability of the OAP is not just a matter of technical advancement; it is a pivotal step towards more efficient educational processes that could transform the learning environment. According to Adikari et al. [1], an understanding of user needs and context is crucial for User Experience (UX) Design, which encompasses more than mere usability—it also includes considerations of user satisfaction and emotional engagement.

Therefore, a more intuitive and user-friendly platform will result in increased adoption rates, a reduction in the learning curve for newcomers, and an enhancement in overall user satisfaction.

This introduction outlines the planned exploration into how Design Thinking, coupled with the latest AI tools, can be used to systematically address and rectify the existing limitations of the OAP, creating a more effective and inclusive digital assessment environment.

Background

User-Centered Design (UCD) is a framework that emphasizes on the creation of products usable and accessible, by focusing the design process on the needs, wants, and limitations of the end users.

Rodríguez-Ortiz M. A.[1,2]; López Ávalos D. G.[1]; Ayala Quiles K. M.[1,2]; Gaytán-Lugo L. S.[3]
[1]Dirección General de Recursos Educativos Digitales, [2]Facultad de Telemática, [3]Facultad de Ingeniería Mecánica y Eléctrica
Universidad de Colima
Colima, México
Email: {maro, dglopez, kayala2, laura}@ucol.mx

It involves extensive user research, iterative testing, and feedback loops to ensure that the final product aligns with user expectations. User Experience (UX) Design is a more expansive process that encompasses the creation of products that provide meaningful and relevant experiences to users. It is a comprehensive process undertaken by design teams with the objective of developing products that deliver meaningful and relevant experiences.

UX Design is a comprehensive process undertaken by design teams to create products that deliver meaningful and relevant experiences. This process encompasses the entire spectrum of acquiring and integrating a product, which includes branding, design, usability, and functionality. UX Design extends beyond merely simplifying software usage; it involves crafting holistic product-related experiences such as marketing campaigns, packaging, and after-sales support. Ultimately, UX Design aims to enhance the overall customer experience by addressing specific user issues and needs [11].

Design Thinking, as defined in Adikari et al. [1], is a holistic and integral approach that employs cognitive processes inherent to designing to tackle complex and socially ambiguous problems. This methodology merges the methods and sensibilities of designers to forge human-centered solutions that synchronize the needs of people with technological capabilities and viable business opportunities.

By embedding empathy and creativity at its core, Design Thinking transforms abstract ideas into practical solutions that add significant value. It has proven especially effective in refining product design and user interactions through its iterative and User-Centered approach, making it indispensable for innovation and the discovery of new design opportunities. This approach has been increasingly applied to enhance user experience, as it integrates an understanding of user needs with creative problem-solving techniques, providing a solid theoretical foundation for employing both Design Thinking and AI tools in optimizing platforms like this OAP.

Hillner and Lim [11] explore the adoption of Design Thinking in diverse settings, including retail, system, and social design; their research highlights the flexibility and broad applicability of Design Thinking, emphasizing its significance for the OAP, which serves educators with varying needs and technological proficiencies. This adaptability is crucial, as it ensures that Design Thinking can be effectively implemented across different domains, facilitating User-Centered solutions and enhancing overall user experience.

AI Tools in Enhancing UX Design

Mortazavi [14] explores AI tools in User Experience (UX) Design workflows and highlights their ability to enhance efficiency and creativity. Mortazavi's study demonstrates that AI tools assist in data collection, user feedback analysis, and creating detailed user personas and Journey Maps. This directly applies to the current research, where AI tools, like ChatGPT, analyze user feedback and help to generate Empathy Maps, enriching the Design Thinking process and providing deeper insights into user experiences.

For its part, Asadi [2] provides an autoethnographic exploration of using Large Language Models (LLMs), like ChatGPT, in Design Thinking and underscores their transformative potential in facilitating ideation, prototyping, and User-Centered Design (UCD).

Fischer et al. [9] examine the practical applications and limitations of using ChatGPT in Design Thinking workshops, underlining the importance of human intervention and the complementary role of AI. This perspective is integral to the current study, which aims to leverage AI tools to support, rather than replace, human creativity and empathy in improving UCD and UX Design.

Materials and Methods

This study employed Design Thinking principles to systematically enhance the user experience of the OAP. Its iterative nature and emphasis on User-Centered solutions made it ideal for addressing the intricate usability challenges identified in the preliminary analyses.

To gather preliminary data regarding the usage and challenges associated with the OAP, a comprehensive survey was conducted among faculty members of the same public university one year prior to the commencement of this study.

The survey employed a mixed-methods approach, combining 23 open-ended and closed-ended questions to obtain a holistic understanding of the faculty members' experiences with the OAP. The survey included questions pertaining to demographic information, digital skills, and teaching experience, and specific inquiries regarding the use of digital educational tools, such as Learning Management Systems (LMS) and OAPs. The participants were asked about their experience with EvPraxis, including the frequency of use, their perception of the usability of the OAP, and the specific challenges encountered.

The open-ended questions encouraged respondents to describe typical scenarios of platform use, their device preferences, and their needs and expectations for online assessment tools. A total of 197 professors out of 2,211 (8.9% of the total number of respondents) provided valuable insights into their interactions with educational technology.

As a result, it was stablished that the mean age of the faculty is approximately 47 years old, with the majority holding master's degrees (39.1%) or doctorates (37.1%).

The gender distribution was nearly equal, with 52.8% of respondents identifying as male and 47.2% as female. Most faculty members do not have significant disabilities; however, a small percentage (8.6%) have visual impairments.

The respondents demonstrated moderate proficiency in digital skills, with an average of 3.5 years of experience using digital teaching tools and 7.02 years creating digital content.

Moreover, they have considerable teaching experience, with an average of 18.5 years. The academic work of these professionals is distributed across a range of disciplines, including Social Sciences (26.4%), Humanities (23.3%), Engineering and Technological Development (15.5%), and Medicine and Health Sciences (14.5%). This reflects a diverse range of knowledge and expertise available.

Figure 1 depicts the process employed in this study to analyze the data, commencing with the transcription of the interviews and usability tests to culminate in the creation of the Ideal User Journey Map. This process underpinned our analysis, enabling us to identify pain points and propose solutions that were specifically tailored to the actual experiences of faculty members, which was a crucial step in refining the OAP based on real-world data.

For those interested in reviewing the data tables referenced in this study, they can be accessed at the following link: https://ihclab.ucol.mx/paginas/from+transcription+to+empathy.

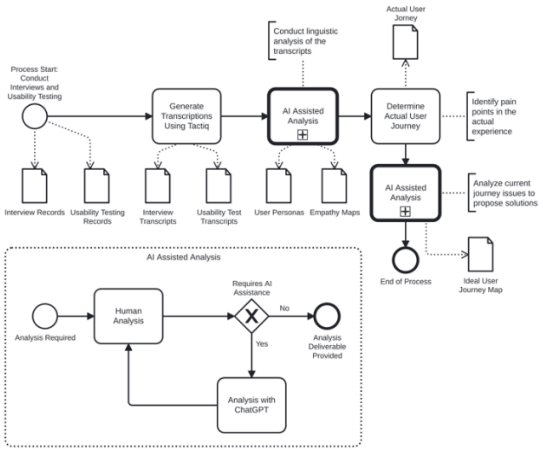

Figure 1. Process of analyzing information, from interview and usability test records to the creation of the Ideal User Journey Map

Participant Selection and Data Collection

The research team selected a diverse group of participants who, a year earlier, had answered a survey aimed at establishing a comprehensive faculty profile. From this survey, 95 professors expressed their willingness to participate in user experience improvement activities for institutional digital platforms; among these, 55.8% were already users of the OAP.

Selection of Participants for Usability Tests and Interviews

Individuals who expressed interest in participating in the "User Experience Improvement Process for the EvPraxis Platform" were invited via institutional email. The selection criteria were based on the participants' prior experience with the platform:

- **Usability tests:** A total of 17 professors who had no prior experience using the OAP responded, indicating their willingness to participate in usability tests. Nevertheless, only 12 professors ultimately participated in the usability tests. This number is consistent with the recommendation by Tullis and Stetson [17], which suggests that a minimum 12-14 participants is necessary to obtain results with a reasonable level of reliability. Moreover, as Faulkner [8] observes, a sample size of 10 users elevates the minimum percentage of identified problems to 82%.
- **Interviews:** Fourteen professors with prior experience using the platform responded and all proceeded to participate in the in-depth interviews. This group was purposely chosen to offer comprehensive insights into their experiences and challenges over time.

Selection of Participants for Usability Tests and Interviews

The usability tests involved professors with no prior experience using EvPraxis, so the tests were designed to simulate typical tasks performed on the platform, allowing researchers to observe real-time user interactions and difficulties.

The demographic data of this group were as follows:

- **Average age:** 48 years.
- **Gender distribution:** 58.3% male, 41.7% female.
- **Educational levels:** 8.3% hold a bachelor's degree, 50% hold a master's degree, and 41.7% hold a doctorate.
- **Teaching experience:** Extensive, an average of 18.25 years.
- **Academic fields:** Physical-Mathematical Sciences and Earth Sciences (14.3%), Biology and Chemistry (7.1%), Humanities (14.3%), Social Sciences (21.4%), and Engineering and Technological Development (42.9%).

Selection of Participants for Usability Tests and Interviews

The interviews included professors who were already using EvPraxis. This group provided insights based on their direct experience with the platform. The demographic data were as follows:

- **Average age:** 45 years.
- **Gender distribution:** 57.1% male, 42.9% female.
- **EvPraxis use level:** 57.1% Advanced, 21.4% intermediate and 21.4% expert level.
- **Teaching experience:** Extensive, averaging 14.5 years.
- **Academic fields:** Physical-Mathematical Sciences and Earth Sciences (10.5%), Biology and Chemistry (5.3%), Medicine and Health Sciences (21.1%), Behavioral Sciences and Education (10.5%), Humanities (15.8%), and Engineering and Technological Development (21.1%).

Data Collection Procedures

Both interviews and usability tests were conducted online via Google Meet. Prior to the sessions, participants were informed about the purpose of the study, and their consent was obtained through a detailed consent form that outlined the intended use of recordings and the information generated during their participation. The sessions were recorded with the explicit permission of the participants, with the objective of capturing their interactions and feedback while using the platform. The Think-Aloud protocol was employed to gather participants' verbalized thoughts during tasks, thereby providing real-time insights into their cognitive processes. We ensured that participants' names and personal data would not be disclosed or associated with the recordings or the information used in the platform improvement process.

Tools and Techniques for Data Analysis

To transcribe the audio recordings of the interviews and usability tests, the researchers utilized Tactiq (https://tactiq.io/), an automatic transcription software which significantly expedited the data analysis process. Subsequent linguistic analysis of the transcriptions was conducted to extract patterns of usage and user sentiments, determining what users thought, felt, did, and said (to create the Empathy Maps). Additionally, ChatGPT was employed to assist in creating the detailed user Journey Maps. These tools facilitated the synthesis of large volumes of qualitative data into coherent visual representations, clearly illustrating user needs, frustrations, and interactions.

Usability Evaluation

The usability of the OAP was assessed using the System Usability Scale (SUS), a reliable tool for measuring the usability of various systems and technologies. This scale provided a quantitative measure of user satisfaction, complemented by qualitative insights from interviews and observational data.

The SUS Analysis Toolkit (developed by the Mixality Research Group [3]) was used given this web-based toolkit facilitates the analysis of SUS usability studies, offering insights and contextualization based on scientific literature, since it allows

for easy calculation, comparison, and plotting of SUS scores, as well as the production of scientific figures and calculations for publications and presentations.

Prior to the usability tests, participants underwent a preparatory phase designed to familiarize them with the platform. This phase included:

- **Initial setup and consent:** Participants were briefed on the study's purpose and procedures; then asked to complete an informed consent form and a preliminary questionnaire. This ensured voluntary participation and collected initial demographic and usage data.
- **Scenario 0. Platform exploration:** In this scenario, participants explored the platform without time constraints to become familiar with the interface before starting specific tasks.
- **Scenario-based testing:** The usability testing included three scenarios simulating typical tasks on EvPraxis:
  - **Scenario 1. Creating an exam:** Participants (acting as professors) logged in, created an exam with at least four questions, and saved it to assess the ease of exam creation.
  - **Scenario 2. Scheduling and group management:** Participants created a group, scheduled the exam, and activated it to test functionality for managing groups and scheduling.
  - **Scenario 3. Exam completion as a student:** Participants switched roles to act as students and registered in the created group and completed the exam they had created, evaluating the student experience and overall exam process.

Throughout the usability tests, the Think-Aloud Protocol was employed, wherein participants verbalized their thoughts, decisions, and difficulties while performing tasks. This provided real-time insights into user experiences and identified potential usability issues.

Upon completion of the scenarios, participants were asked to complete the SUS questionnaire to provide a quantitative measure of the platform's usability. Additionally, follow-up interviews were conducted to gather further qualitative feedback, ensuring a comprehensive evaluation of the user experience.

Results

The application of Design Thinking and advanced technological tools yielded significant insights into the usability of the OAP. This section presents the findings from the usability tests, interviews, and the analytical processes employed during the study.

Usability Test Findings

The usability tests involved a series of tasks that participants were required to complete, which provided valuable data on user interaction with the platform. The average times to complete tasks across different scenarios were notably varied, indicating inconsistency in the platform's ease of use. Key issues identified included difficulties in navigating the platform, confusion over certain interface elements, and obstacles in exam creation and management processes.

Scenario Analysis: Participants took an average of 23.75 minutes (SD = 8.69) to complete the first scenario, which involved creating a digital exam. The subsequent scenarios, which included editing and applying exam settings, took significantly less time (5.75 (SD = 1.93) and 3 (SD = 1.27) minutes, respectively) as we can see on Figure 2, suggesting a learning curve that could be mitigated with better design.

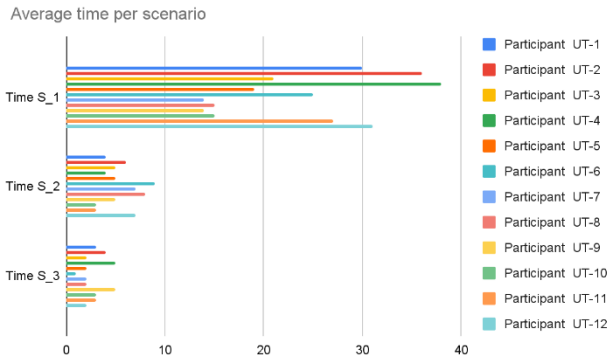

Figure 2. Average time per scenario measured to evaluate participant efficiency.

Interview Insights

The qualitative data from the interviews provided deeper insights into the user experience. Participants frequently expressed frustration with specific features of the OAP, such as the complexity of adding questions to exams and the unintuitive nature of certain functionalities. These issues were particularly evident among less technologically adept users.

Linguistic Analysis

The discourse analysis of the interview transcripts was conducted using the AI tool ChatGPT and mechanical coding. This allowed for the assignment of categories representing predefined or emerging themes, such as: errors made, feelings (frustrations, satisfactions), thoughts, facts, and statements.

In particular, the linguistic analyses employed the technique of content analysis, as defined by Silverman [16], Corbin & Strauss [5], and Krippendorff [13]. This is a systematic method for interpreting text data subjectively by categorizing and identifying themes or patterns; it involves condensing extensive text into fewer categories based on clear coding rules, with the aim of making replicable and valid inferences about the data's context and meaning. This technique may be applied to the analysis of a range of qualitative data, including interviews, documents, and audio and video transcripts, with the aim of extracting meaningful insights.

In this study, the stages of coding, thematization, and quantification were conducted as follows:

- **Identification of relevant data:** The coding of errors made, findings, areas for improvement, feelings, thoughts, facts, and statements of the subjects under study shall be undertaken.
- **Thematic grouping:** The color-coded codes are then thematized to form categories in the Empathy Maps.
- **Quantitative visualization:** The use of quantification serves to highlight the prevalence of certain feelings or thoughts, thereby supporting the process of decision-making in the field of UX design.

Consequently, the linguistic analysis enabled the identification of the terminology and processes that were causing confusion, which facilitated the targeted redesign of the user interface; see an example on Figure 3.

The linguistic analyses, in conjunction with the AI tool ChatGPT, enabled the creation of the **Empathy and User Journey Maps**. These revealed common emotional responses and the mental journey undertaken by users when interacting with the OAP. They also identified areas of user interaction that are particularly

prone to difficulties, particularly in situations that require frequent and complex interactions.

6. **Hace:** Supo a dónde dirigirse para crear el examen, pero después ya no supo cómo comenzar el proceso de redacción.
7. **Hace:** No comprendía qué era una "Sección" y sus características; le costó mucho trabajo entender a qué se refería el campo del "Título".
8. **Piensa:** "Pensé: '¡Híjole! Es el examen, ¿y si lo hago mal y no le doy a entender al estudiante lo que yo le quiero preguntar o lo que yo le quiero evaluar?'."
9. **Hace:** Escribió las opciones de respuestas correctas en los campos de retroalimentación general del examen.
10. **Siente:** Dos tipos de preguntas le causan estrés y conflicto al momento de crear los exámenes: las preguntas abiertas y las de relacionar columnas.
11. **Hace:** No supo cómo darle el formato correcto al reactivo de "Completar" (dos guiones bajos + espacio en blanco). [Nota: no existe una manera visible en la plataforma de que el usuario nuevo lo sepa, a menos de que se le indique expresamente o tome una capacitación previa.]
12. **Hace:** Al enfrentarse a problemas con ciertos tipos de reactivos, decidió evitarlos y avanzar con la creación del examen usando otros tipos de preguntas.
13. **Piensa:** "Tal vez, con el apoyo de la inteligencia artificial [se] puedan generar exámenes [...] mejor redactados."
14. **Siente:** "A mí se me ha hecho fácil porque me va guiando [en cuanto a] qué tipo de pregunta quieres generar y luego [agregas] la respuesta correcta."
15. **Dice:** "Yo siento que sería lo ideal: hacer un banco de preguntas."

Figure 3. A detail account of the linguistic analysis derived from interviews conducted for creating Empathy Maps and User Journey Maps.

Interview Insights

The System Usability Scale (SUS) provided a quantitative measure of the platform's usability, with an average score of 77.08.

As seen in figure 4, this score places EvPraxis in the "Good" category of usability yet underscores the potential for significant improvements. The SUS scores correlated with user feedback from interviews and usability tests, validating the concerns and areas for enhancement identified during qualitative assessments.

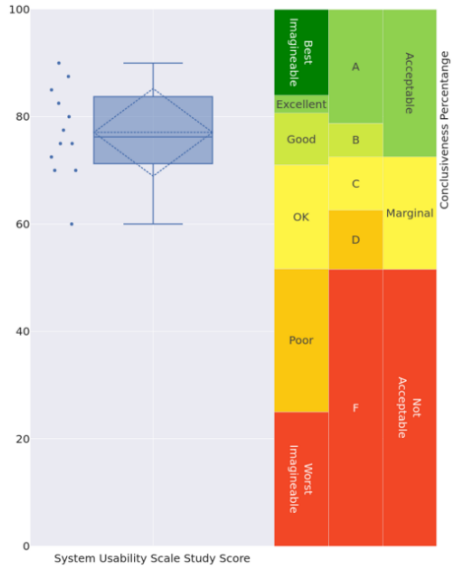

Figure 4. The System Usability Scale (SUS) score obtained using the SUS Analysis Toolkit.

Discussion

The comprehensive analysis conducted in this study has highlighted critical areas needing enhancement within the OAP, providing valuable insights that align with the theoretical framework presented in the background.

The tenets of user-centered design (UCD) prioritize the creation of products that are accessible and usable, with a particular focus on addressing the needs and limitations of the users. This study reinforces the imperative for adherence to these principles, underscoring the existence of significant usability issues that necessitate a more intuitive design.

User-Centered Design (UCD) emphasizes creating products that are accessible and usable by focusing on the needs and limitations of the users. This study underscores the necessity of these principles, revealing significant usability issues that necessitate a more intuitive design.

The application of Design Thinking, as defined by Adikari et al. [1], proved effective human-centered solutions that integrate empathy and creativity; when it is used in conjunction with UCD Design Thinking facilitated the identification of both emotional and practical aspects of the user experience, reinforcing the importance of addressing these elements during the design process. Insights derived from the Empathy and User Journey Maps were crucial in identifying user pain points, confirming the value of addressing both emotional and practical aspects in the design process.

The study also supports Mortazavi's [14] assertion on the role of AI tools in enhancing UCD and UX workflows. The use of ChatGPT for feedback analysis and empathy map creation enriched the Design Thinking process, providing deeper insights into user needs. This also aligns with Asadi's [2] findings on the potential of large language models in facilitating User-Centered Design (UCD).

However, the limitations highlighted by Fischer et al. [9] are acknowledged, emphasizing that AI tools should complement human creativity and empathy.

Overall, the study demonstrates the potential of combining Design Thinking with advanced AI technologies to systematically address usability challenges, fostering a more effective and inclusive digital assessment environment, and promoting higher adoption rates and satisfaction among educators.

Interpretation of Findings

Usability Test

The findings from the usability tests and interviews underscore significant areas for improvement within the OAP. Despite achieving a "Good" rating on the SUS with an average score of 77.08, the data suggests that there are still numerous challenges that prevent an "Excellent" usability rating.

The average times recorded for completing various tasks reflect a substantial variability in the ease of use, highlighting a need for a more intuitive design that could facilitate quicker and more efficient user interactions.

After normalizing the values of each response on a scale from 0 to 10 and plotting the results, Figure 5 shows that Question 2 has a score of 5.62 and Question 4 has a score of 5.42. These even-numbered questions indicate that users perceive the platform as moderately complex and believe they may require technical support to use it effectively.

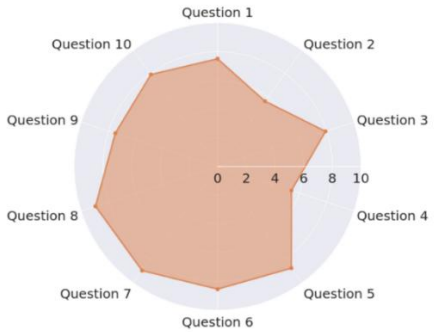

Figure 5. Per item normalized scores (0-10) contributing to the SUS study score.

The Net Promoter Score (NPS) is a widely used metric for measuring customer loyalty. According to Fisher and Korduplesk [10] NPS is calculated based on customer responses to the question "How likely are you to recommend our platform to a friend or colleague?", using a scale from 0 to 10. Customers are categorized into three groups based on their responses: promoters (9-10), passively satisfied (7-8), and detractors (0-6). The NPS is derived by subtracting the percentage of detractors from the percentage of promoters, resulting in a score that ranges from -100 to +100.

Sauro [15] identifies a significant correlation between the SUS and NPS through the LTR metric. In addition, Sauro suggests that LTR can be approximated by dividing the SUS score by 10, providing an initial estimate of NPS. For instance, a SUS score of 77.08 translates to an LTR of 7.7, placing users in the "passively satisfied" category.

However, Sauro cautions that the conversion is neither linear nor straightforward, and the relationship between LTR and NPS must be empirically adjusted to accurately reflect the actual distribution of user responses.

Using the MeasuringU calculator, an LTR of 7.7 from a SUS score of 77.08 corresponds to an approximate NPS of 16%. An NPS of 16% indicates a modest margin of more promoters than detractors, suggesting that while a decent number of users are willing to recommend the product, there is also a considerable proportion of neutral or dissatisfied users.

The insights derived from Empathy and User Journey Maps have been instrumental in illustrating the emotional and practical struggles users face while navigating the platform. These tools have effectively identified specific pain points in the user experience, such as difficulties in exam creation and settings management, which are not immediately obvious through quantitative measures alone.

Empathy Maps

From the Empathy Maps, it is evident that users experience a range of emotions and thoughts while interacting with the OAP. New users often feel insecure and stressed due to the lack of intuitive guidance and unclear instructions. For instance, one user expressed: *"I feel stressed because I have to review the entire screen to see where to click"* highlighting the need for a more user-friendly interface and better instructional support. Furthermore, users found certain elements, such as the section and title fields, confusing and difficult to understand without explicit guidance.

On the other hand, experienced users appreciated the practicality of the platform but pointed out persistent issues that hinder their workflow. For example, some users mentioned the need for better integration with other institutional systems to streamline group creation and exam scheduling. One user suggested: "If groups created in SICEUC could be automatically linked with EvPraxis, it would make the exam assignment process much easier."

Real User Journey, Derived from Empathy Maps

The designed persona by the researchers to trace the real user journey was Sofia Guerrero, an experienced higher education professor. This user journey provides a detailed account of her first encounter with the OAP and during her initial use, Sofia faced significant challenges, particularly in navigating the platform and understanding how she could create and manage exams; she felt stressed and frustrated due to the platform's complexity and lack of intuitive guidance. For example, Sofia struggled with the interface while trying to create her first exam, not knowing where to start or how to proceed with drafting questions.

Sofia's journey highlighted several specific areas for improvement:

- **Platform navigation:** Sofia found the navigation less intuitive compared to other platforms, like Google Forms, which led to initial stress and confusion.
- **Exam creation:** The process of creating an exam was complicated by unclear instructions and the layout of fields, leading her to make errors, such as writing answers in the wrong fields.
- **Group and exam scheduling:** Assigning exams to groups and scheduling them was not straightforward, causing further frustration and errors in the process.

Despite these challenges, Sofia recognized the potential of the OAP once she became more familiar with it. She appreciated the platform's capabilities, such as the ability to create varied question types and integrate images into questions but emphasized the need for better user support and clearer instructions.

Ideal User Journey

In envisioning an ideal user journey for Sofia Guerrero, the goal was to create a streamlined and user-friendly experience that addresses the identified pain points. This improved journey includes several key enhancements:

- Pre-platform exploration: In the ideal scenario, Sofia receives comprehensive training through academic workshops, which provide her with a clear understanding of the OAP before she uses it independently. Additionally, the platform is introduced with positive feedback from colleagues, encouraging Sofia to explore it with confidence.
- Platform navigation: When Sofia logs into the OAP, she is greeted by a visually appealing and intuitive interface. A step-by-step tutorial or guided tour helps Sofia understand the platform’s main features, significantly reducing her initial stress and confusion. This guidance ensures that she feels comfortable navigating the OAP from the start.
- Exam creation: In this ideal journey, Sofia easily finds the exam creation option, guided by clear, contextual help prompts. The platform provides a user-friendly editor with predefined templates for various question types, minimizing confusion and errors in the main process. Additionally, a real-time preview feature allows Sofia to see how the exam will appear to students, enhancing her confidence in the process. This makes the task of creating exams more straightforward and less intimidating.
- Group and exam scheduling: Sofia can directly link her LMS or Institutional School Control System groups to the platform, streamlining the process of assigning exams. Scheduling exams is simplified with an intuitive calendar interface and clear instructions, reducing the likelihood of errors. This enhancement ensures that Sofia can efficiently manage group assignments and exam schedules without frustration.
- Post-exam management: After the exam Sofia receives automated notifications confirming the successful submission and grading of exams. The platform provides detailed analytics on student performance, accessible through an easy-to-navigate dashboard. Throughout the

process Sofia feels supported, with access to ongoing training and resources, and very confident in using the platform effectively and efficiently.

These qualitative insights underscore the importance of addressing the usability issues in the OAP. The Empathy Maps and User Journey Maps revealed that both new and experienced users encounter significant hurdles that affect their overall satisfaction and efficiency. By addressing these pain points through improved interface design, better instructional support, and more intuitive navigation the user experience can significantly be enhanced.

Figure 6 below illustrates the User Journey Map, created using DALL-E via ChatGPT, highlighting key user interactions and challenges encountered by Sofia Guerrero throughout her experience with the OAP.

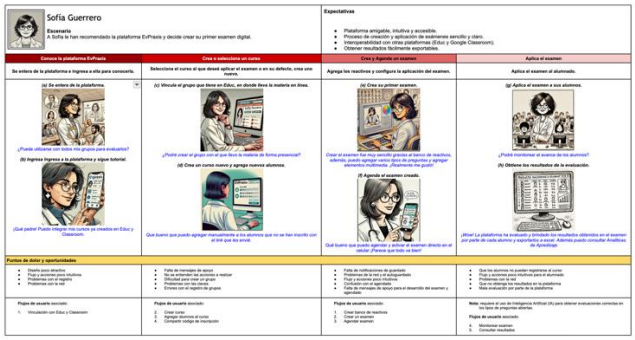

Figure 6. User Journey Map created using DALL-E via ChatGPT, highlighting key user interactions and challenges.

This approach is consistent with the principles of Design Thinking and ensures that the OAP meets the diverse needs of its users, creating a more effective and user-friendly digital assessment environment.

Design Thinking Impact

The implementation of Design Thinking in this context has proven useful, as it has led to a deeper understanding of user needs and the subsequent identification of design deficiencies. This approach has facilitated a User-Centered view in diagnosing usability issues, ensuring that solutions are aligned with user expectations and requirements.

However, it is important to note that while AI tools like ChatGPT have contributed to creating detailed user interaction maps, providing a richer and more nuanced understanding of the user experience, their effectiveness in this process should be approached with caution. According to Hicks et al. [12], large language models (LLMs) like ChatGPT can produce content without any genuine concern for the truth, a phenomenon better described as 'bullshit' rather than 'hallucinations.' Therefore, insights derived from AI should be carefully evaluated and supplemented with human expertise to avoid misinterpretations and ensure that design interventions are both valid and effective.

Practical Implications

The findings have several practical implications for the ongoing development of the OAP. Firstly, the identified usability issues necessitate a redesign of the platform's navigation schema and user interface. This redesign should aim to simplify the user journey, facilitating more efficient completion of common tasks.

Furthermore, the study highlights the importance of user training and support. The implementation of enhanced training materials and potentially real-time support features has the prospective to markedly improve user satisfaction and reduce the reliance on trial-and-error learning.

Limitations and Future Research

While the findings are compelling, this study has some limitations. The sample was primarily composed of faculty staff from a public university, which may limit the generalizability of the results to other educational institutions or user demographics. Therefore, future research should consider a more diverse sample population and may benefit from comparing usability across different educational technologies to validate these findings.

Moreover, further research should also investigate the long-term effects of the implemented design changes. A subsequent study could assess whether the improvements have sustained positive impacts on usability and user satisfaction.

Additionally, while AI-driven tools such ChatGPT have been valuable in analyzing user interactions, their limitations, particularly the potential to generate misleading content, should be acknowledged. As Hicks et al. (2024) suggest, insights from large language models must be carefully evaluated. Future research should explore ways to better integrate these tools with human expertise to ensure reliable and effective user feedback.

Conclusions and Future Work

The findings of this study highlight the value of integrating Design Thinking with advanced AI tools to optimize the user experience of this specific OAP. The comprehensive analysis, which included usability tests and in-depth interviews, revealed critical usability issues and areas for improvement. By prioritizing a User-Centered Design (UCD) approach and employing AI for feedback analysis, the research has stablished a foundation for systematic enhancements that address both technical and emotional aspects of user interactions.

The following sections summarize the key findings and outline the future steps necessary to implement and evaluate the proposed improvements, with the objective of ensuring that the platform evolves into a more efficient and user-friendly system.

Conclusions

A series of usability tests, in-depth interviews, and analytical tools were employed to identify critical areas for improvement in the OAP. Findings from the SUS and user feedback indicate that, while the platform is generally effective, several usability issues remain.

Key issues include the necessity for clearer navigation, more intuitive interface elements on the exam creation page, and simplified exam management processes. The participants' emotional responses (captured through Empathy Maps and User Journey Maps) reveal user frustration and confusion, identifying crucial areas for improvement.

The use of Generative AI (GenAI) and automated transcription accelerated analysis, reducing the time required for interview analysis and aiding in the discovery of valuable insights. However, AI should complement, not replace, human skills and decision-making processes, so regular review and validation of AI-generated results are essential.

In conclusion, while AI tools facilitate data analysis, the goal is to enhance human capabilities and ensure that design decisions are User-Centered and reflect genuine user needs. This balanced approach will help achieve a more intuitive, user-friendly, and efficient OAP.

Future Work

The next phase of this project will entail two primary tasks: the implementation of the suggested improvements and a subsequent re-evaluation of the platform to assess the impact of these changes on the user experience. The proposed enhancements will be designed to directly address the usability challenges identified in this study, focusing on streamlining the exam creation process, upgrading the clarity of the user interface, and reducing the cognitive load on users.

- **Implementation of improvements:** Based on the study's findings, the development team will prioritize modifications that simplify navigation and interaction, such as revising ambiguous commands and reorganizing the structure of exam settings. The incorporation of more intuitive design elements is expected to reduce the time users spend on common tasks and decrease the overall complexity of the platform.
- **Evaluation of implemented changes:** After these improvements are implemented, a second round of usability testing and interviews will be conducted. This will not only assess the effectiveness of the new designs in enhancing usability but will also provide data on user satisfaction and system performance subsequent to implementation. The SUS will be employed once more to quantify improvements in usability, and further user feedback will be solicited to refine the system further.

The objective of this ongoing project is to guarantee that the OAP evolves into a more user-friendly platform that aligns with the most effective practices in digital assessment and meets the high standards of educational technology. By continuously engaging with the feedback loop and consistently applying the principles of Design Thinking, the development team aims to create a robust system that supports educators and students in an optimal manner.

Acknowledgments

We would like to express our gratitude to Sandra Mata Rodríguez, MSc, General Director of the Directorate General of Digital Educational Resources, for her unwavering support and efforts in facilitating this project, and to Martha E. Chávez González, PhD, General Director of the Directorate General for Academic Staff Development, for her invaluable assistance in coordinating the professors who participated in this study. Both organizational units are part of the public university where this study was conducted.

References

[1] Adikari, S., McDonald, C., & Campbell, J. (2013). Reframed Contexts: Design Thinking for Agile User Experience Design. In A. Marcus (Ed.), *Design, User Experience, and Usability. Design Philosophy, Methods, and Tools* (Vol. 8012, pp. 3-12). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-39229-0_1

[2] Asadi, A. R. (2023). *LLMs in Design Thinking: Autoethnographic Insights and Design Implications*. 2023 The 5th World Symposium on Software Engineering (WSSE), 55-60. https://doi.org/10.1145/3631991.3631999

[3] Blattgerste, J., Behrends, J., & Pfeiffer, T. (2022). *A Web-Based Analysis Toolkit for the System Usability Scale*. Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments, 237-246. https://doi.org/10.1145/3529190.3529216

[4] Coordinación de Planeación y Desarrollo Institucional (CPDI). (2024). *Numeralia Institucional 2023*. Universidad de Colima. https://portal.ucol.mx/cpdi/numeralia.htm

[5] Corbin, J. & Strauss, A. (2015). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory* (4th ed.). SAGE Publications.

[6] Dirección General de Recursos Educativos Digitales (DGRED). (2023). *Informe de labores 2023*. Universidad de Colima. https://portal.ucol.mx/dgred/informe.htm

[7] Fajardo-Flores, S., Gaytan-Lugo, L., Santana-Mancilla, P., & Rodriguez-Ortiz, M. (2021). *Accessibility Assessment for Online Education Tools: Towards Accessible Principles for a Mexican University*. EAI Endorsed Transactions on E-Learning. https://doi.org/10.4108/eai.17-2-2021.168720

[8] Faulkner, L. (2003). Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. *Behavior Research Methods, Instruments, & Computers*, *35*(3), 379-383. https://doi.org/10.3758/BF03195514

[9] Fischer, H., Dres, M., & Seidenstricker, S. (2023). *Application of ChatGPT in Design Thinking*. AHFE 2023 Hawaii Edition. https://doi.org/10.54941/ahfe1004312

[10] Fisher, N. I., & Kordupleski, R. E. (2018). Good and bad market research: A critical review of Net Promoter Score. *Applied Stochastic Models in Business and Industry*, *35*(1), 138-151. https://doi.org/10.1002/asmb.2417

[11] Hillner, M., & Lim, S. (2018). *Design thinking—Towards a new perspective*. https://doi.org/10.13140/RG.2.2.34625.04967

[12] Hicks, M.T., Humphries, J. & Slater, J. (2024). ChatGPT is bullshit. *Ethics Inf Technol*, *26*(38). https://doi.org/10.1007/s10676-024-09775-5

[13] Krippendorff, K. (2018). *Content Analysis: An Introduction to Its Methodology* (4th ed.). Sage Publications.

[14] Mortazavi, A., Andersson, T., Jankauskas, T., & Wever, R. (2023). *Enhancing User Experience Design workflow with Artificial Intelligence tools* [Master's thesis]. Linköping University.

[15] Sauro, J. (2012, January 3). Predicting Net Promoter Scores from System Usability Scale Scores. *MeasuringU–UX Research and Software*. https://measuringu.com/nps-sus/

[16] Silverman, D. (2011). *Interpreting Qualitative Data* (4th ed.). Sage Publications. Tullis, T. S., & Stetson, J. N. (2006). *A Comparison of Questionnaires for Assessing Website Usability*.

[17] Tullis, T. S., & Stetson, J. N. (2006). *A Comparison of Questionnaires for Assessing Website Usability*.

© 2024 by the authors. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/ or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.