Available online at sjuoz.uoz.edu.krd
Science Journal of University of Zakho
Vol. 12, No.1, pp. 34 –45 January-March, 2024
p-ISSN: 2410-7549
e-ISSN: 2414-6943
journals.uoz.edu.krd

ENHANCING WEBSITE USABILITY TESTING: CORRELATING EYE-TRACKING, GSR, AND SUS DATA WITH RESPECT TO GENDER PREFERENCES

Ismael A. Ali
Jazari Research Center, University Research Center, University of Zakho, Zakho - Iraq
Department of Computer Science, Faculty of Science, University of Zakho, Zakho - Iraq
ismael.ali@uoz.edu.krd
Received: 17 Oct., 2023 / Accepted: 20 Nov., 2023 / Published: 23 Jan., 2024. https://doi.org/10.25271/sjuoz.2024.12.1.1215

ABSTRACT:

Having a professional team of web developers can produce a professional website, but cannot guarantee an expected usable website. This study presents a comprehensive multilayer approach for examining the correlations between different layers of user consciousness in website usability testing. It utilizes visual attention data from eye-tracking, emotional engagement data from galvanic skin response, and self-reporting data from the system usability scale. Testing AUK and UoZ university websites with 18 users using the Gazepoint GP3 system revealed insightful correlations among different layers of user consciousness, such as high emotional engagement is associated with higher fixation counts and shorter time-to-complete and thus lower SUS scores. Whereas low emotional engagement is associated with lower fixation counts, longer time-to-complete, and thus higher SUS scores. Gender preferences verifies the results from the literature on female users generally experiencing higher emotional arousal thus having lower time-to-complete and lower SUS scores. Design problems are presented in the form of improvement recommendations. The findings of the study highlight the importance of considering different layers of user consciousness in website usability testing, as well as the importance of gender preferences. Finally, current limitations and future works are presented.
KEYWORDS: Website Usability Testing, Galvanic Skin Response, Eye-Tracking, System Usability Scale, Gender Differences.

1. INTRODUCTION

The usability feature of the websites is not mainly provided by web developers. This is the reason behind involving eye-tracking technologies in the field of website usability testing as it can answer usability questions of how users look at your website, whether they are looking at spots you need them to find and click on, or how much they pay attention to them. However, the eye-tracking data only captures our visual attention through our eye movements when observing a visual stimulus, but what about unconscious emotional engagement in the usability settings? This work addresses the effectiveness of the correlation between heterogeneous sources of website usability testing data toward a more comprehensive usability evaluation of the websites.

To draw a wider picture of our subconscious and ultimately conscious attention, this study examines the parallel correlation of three dimensions of user consciousness while performing website usability testing as (1) the user's unconscious emotional engagement captured by Galvanic Skin Response (GSR) biometric data, (2) the user's subconscious visual attention captured by eye-tracking data, (3) and user's conscious self-reporting feedback collected with System Usability Scale (SUS). The proposed approach is tested against the usability of the websites of two universities, The American University of Kurdistan (AUK) and the University of Zakho (UoZ).

The findings of this study revealed insightful correlations among different layers of user consciousness. Also, improvement recommendation is extracted from usability patterns that in turn affect the usability performance in terms of the tasks assigned to the users. The rest of the paper is structured as follows. Section 2 presents the background of the study. Section 3 reviews the related literature. Section 4 presents the proposed approach. Section 5 presents the results. Section 6 presents the study conclusions and future works.

2. BACKGROUND

2.1 Website Usability Problem

The usability problem is a subfield in Human-Computer-Interaction (HCI), measuring how successful, effective, and efficient an application is. The original usability problem is defined by ISO Standard (of 9241-1) as "The extent to which a product can be used by specified users to achieve specific goals with effectiveness, efficiency, and satisfaction in a specified context of use" (Green, 2006). Historically, the website usability problem has been observed, investigated, and practiced in the context of the field of e-commerce (Rinder, 2012).

Due to usability facts: any given website has 50 milliseconds to make a good first impression, and users do not usually tolerate a bad user usability experience (Barnum, 2020). Therefore, usability is a real issue and not a luxury. As a conclusion, the term website usability refers to the ability of a website to be effectively used by designated users within a particular timeframe to accomplish predetermined objectives with efficiency, effectiveness, user engagement, tolerance for errors, and ease of learning within a specified usage context (Green, 2006). Based on this definition website usability can be formulated with the following dimensions (Green, 2006; Abran, 2003; Buchanan, 2009):

- *Errors*: The tolerance measurement of how often errors are made by the user, as well as the severity, preventability, and recovery of the errors.
- *Satisfaction*: The engagement measurement of how pleasant the interface is to use by the user.
- *Efficiency*: The resource-usage measurement of how much time and cognitive load the user has consumed to accomplish their goals.

* Corresponding author

- *Effectiveness*: The achievement measurement of how accurately a given user has accomplished given goals.
- *Learnability*: The comprehension measurement of how easy it is for the user to achieve basic tasks for the first time after encountering the system.
- *Memorability*: The recallability measurement of how efficiently a given task is re-accomplished by a given user after a period of not using the same system.

Website usability testing can be done during the design phase or afterward in the testing or production phase. Thus technically, website usability testing is an evaluation practice of both functionality and design of the websites for given selected users and tasks/goals to accomplish. One of the main products of the website usability testing is a recommendation report on how to re-design and re-conceptualize the website for optimally better user usability experience and overall to meet organization objectives.

2.2 Eye-Tracking

Eye-tracking helps in re-designing more usable websites by measuring users' visual attention and behavior while using the websites (Djamasbi, 2014). Eye movements, eye positions, and places of focus can all be measured utilizing eye-tracking equipment. This valuable data can answer valuable questions such as, where exactly visitors are looking at or ignoring and for how long (Bergstrom, 2014). The most fundamental procedure of conducting an eye-tracking experiment for studying the usability of users toward visual stimuli has the following three elements (Goldberg, 2003): (1) selection of representative sample users from the user population, such as shoppers for an e-commerce website, (2) selection of representative actual technical tasks from tasks pool, such as asking the users to go through ordering an item in an e-commerce website, (3) monitoring and recording users usability behavior while performing and accomplishing the assigned task. Besides website usability testing, eye trackers are interestingly being used for solving a wide range of problems in academia and industry such as in the Internet of Things (Klaib, 2021), virtual reality (Clay, 2019), gaming (Stahlke, 2021), and software engineering (Sharafi, 2020).

Technically eye-tracking instrument needs to monitor gaze points and directions, where the eyes are looking by combining near-infrared technology with a high-resolution camera. Whereas, the underlying idea is known as Pupil Center Corneal Reflection, as the camera tracks the pupil's center and the cornea's reflection of light (Carter, 2020). This way the eye-tracking device can quantify and collect data about the visual and cognitive behavior regarding the attention or ignorance of the users toward presented visual stimuli. The following are the two main eye-tracking metrics used in this study (Carter, 2020; Blascheck, 2014): Fixation Count, which is a spatial and temporal cluster of very close gaze points that can be used to identify an area of interest by its visual attention, and Time-To-Complete (TTC), which is measured in seconds as the time the user spends to perform a given task or a finite set of tasks.

2.3 Galvanic Skin Response

Although eye-tracking measures the underlying mental processes related to human visual attention, there is still a concern about the correlation between patterns of eye movements and the related website usability and cognitive problems. To solve this vagueness, psychosociological Electrodermal Activity (EDA), also known as Galvanic Skin Response (GSR) is used (Tyler, 2015). GSR measures the activity of the autonomic nervous system as the level of emotional arousal in human skin changes based on the current stimuli that produce emotional responses (Boucsein, 2012), which in turn causes an increase in eccrine sweat gland activity.

Thus, GSR is a good quantifier of human emotional arousal reporting users' unconscious psychological processes toward the stimuli triggers presented to them in different application domains (Boucsein, 2012; Satti, 2021; McNeal, 2020; Lin, 2005; Fraiwan, 2018; Baig, 2019; Foglia, 2008), thus making physiological data fitting into usability evaluation (Lin, 2005). The EDA has two main components (Dawson, 2007). First is the general tonic level with a slower-acting component of the signal measured by the Skin Conductance Level (SCL) reflecting general changes in autonomic arousal. The second component is the phasic component with a faster-changing element of the signal measured by Skin Conductance Response (SCR). The literature suggests that SCR is associated with emotional arousal (Dawson, 2007; Fowles, 1981; Anders, 2004; Bakker, 2011) as a result of the fluctuation of both underlying EDA components. Then, EDA Peaks are the sudden changes in phasic activity above tonic activity, which quantifies the level of emotional arousal during usability testing. Such as the more frequent wave peaks present in the GSR data the more emotionally arousing the stimulus is to the user. The most frequently used and studied parameter of EDA is the skin conductance (SC) with the standard unit of *microsiemens* (μS). However, the biometric sensor of GP3 provides the skin resistance feedback of the users, which is just the reciprocal of conductance, with the standard unit of *kiloohm* (kΩ) (Blascheck, 2014; Bari 2018; Bari 2023; Critchley, 2002). It is worth mentioning that GSR is an ideal measure to track emotional arousal, however, it is not able to reveal the emotional valence. Therefore, to gain the most out of GSR data is necessary to combine it with other sources of human-interaction data, such as eye-tracking data in this study. As eye-tracking reflects the visual attention toward the system usability being tested, GSR data can unfold the emotional dimension of the system usability testing and thus validate and complement each other.

2.4 System Usability Score

At the end of the experimental sessions with the collection of eye-tracking and GSR data, the self-reporting SUS questionnaire is conducted by the users, which is a widely utilized tool in the field of website usability testing (Aziz, 2021). In terms of easiness and functionality, the SUS questionnaire serves as a valuable resource for assessing the conscious usability of websites. SUS consists of ten questions addressing different aspects of usability, such as ease of use, efficiency, and overall satisfaction. Users respond to these questions by assigning rating scores on a scale ranging from 1 (strong disagreement) to 5 (strong agreement). This rating system allows users to express their level of agreement or disagreement with statements related to the website's usability. This study adapted the standard SUS statements (Brooke, 1996) as follows:

1. I think that I would like to use the Website frequently.
2. I found the Website unnecessarily complex.
3. I thought the Website was easy to use.
4. I think that I would need the support of a technical person to be able to use the Website.
5. I found the various functions on the Website are well integrated.
6. I thought there was too much inconsistency in the Website.
7. I would imagine that most people would learn to use the Website very quickly
8. I found the Website very cumbersome to use.
9. I felt very confident using the Website.
10. I needed to learn a lot of things before I could get going with the Website.

To calculate the raw SUS score for a particular user, a simple formula is applied. As for odd-numbered questions (1, 3, 5, 7, 9), one is subtracted from the users' rating score. In contrast, for even-numbered questions (2, 4, 6, 8, 10), the users' rating is subtracted from 5. Summing up these adjusted scores results in the raw score, which can range from a minimum of 0 to a maximum of 40. However, the raw score alone does not provide a meaningful interpretation of usability. To transform the raw score into the final SUS score, raw scores from all users are added and then multiplied by 2.5. This conversion process results in a final SUS score, which can range from 0 to 100. The SUS score is not just a numerical value; it comes with qualitative descriptors that help provide a richer understanding of the usability assessment. These adjectives, ranging from "worst imaginable" to "best imaginable," offer a qualitative perspective on website usability. Higher SUS scores correspond to better usability, often described as "excellent" or "acceptable." In contrast, lower scores might be associated with terms like "poor" or "not acceptable", Figure 1.

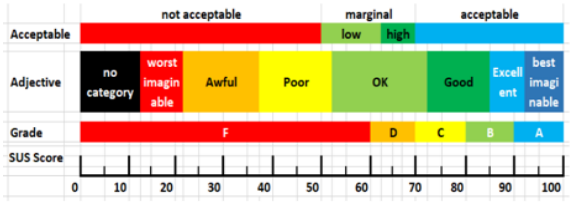

Figure 1: SUS Score (Bangor, 2009; Sauro 2011)

3. RELATED WORKS

Although eye-tracking based website usability testing is the leading website usability testing technique, one of the main challenges can be referred to as the hardship of correlating the patterns of eye-tracking user behavior with the usability and design problems, as well as the user self-reporting data. Some related works from the literature addressing website usability testing from different perspectives are presented. Foglia (2014) studied the potential use of physiological signals, including galvanic skin response, heart rate variability, and respiration, for evaluating user satisfaction with web systems with and without avatars. The findings suggested that the presence of an avatar effectively increases user engagement, and there exists a correlation between the rate of respiration overshoots and user approval of the avatar. This study implied the importance of physiological metrics, but adding the value of eye-tracking concepts to this context is worth a step toward deeper website usability testing.

Wang (2019) investigated the correlation between eye-tracking data and the SUS scores in the context of an online learning platform. The study observed different results in self-reported task difficulties and completion times, aiding in the identification of interface elements that need improvements. As the eye movements data provide insights into how users respond to visual elements, however, this study considered neither users' emotional engagement, nor their visual attention.

Zardari (2020) presented a hybrid user experience evaluation using heuristic evaluation, a user questionnaire, and an eye-tracking technique to identify and resolve complex usability issues in investigated e-learning. The results highlighted that the hybrid approach uncovered more issues compared to the single technique of usability testing. Compared to eye-tracking based usability testing, the presented hybrid approach provided a holistic view of usability issues, however adding the emotional dimension of galvanic skin response (GSR) to this hybrid approach is worth attempting.

Huang (2021) employed heuristic performance evaluation techniques to investigate how gender preferences influence the perception of usability when using online travel websites, indicating that women often have higher website usability expectations, while men have a less detailed approach. Therefore, women expressed a need for enhanced usability support while men were more task-oriented. These gender-based findings can be further enhanced by the involvement of emotional and self-reporting approaches.

De Carolis (2023) presented a study to explore the potential relationship between stress levels based on physiological signals and the usability of two versions of a registration website of "bad-designed" and a "well-designed" version. Then participants completed the SUS questionnaire. The study yielded interesting findings that stress values revealed a negative correlation with the SUS scores. Although the study proves the importance of the physiological dimension of the participants, appending the visual-attention dimension to the study will provide a wider understanding of the user usability experience as well as usability issues.

4. METHODOLOGY

4.1 Approach

Eye movement is the commonly used technique when it comes to the field of website usability testing. However, our visual attention is not the only driver of our website usability behavior as our emotional experience has also a major role in directing our overall attention or avoidance in the use of websites. Therefore, to generate a wider picture of the website usability testing experience, including other sources of usability data such as a biometric metric and self-reporting usability feedback besides eye-tracking is worth attempting. Therefore, unlike the literature, which barely focuses on the use of eye-tracking data, this study pays attention to the emotional and verbal usability experience of users toward the website besides observing the eye-tracking feedback. To examine the proposed approach, a website usability testing setting is designed against the websites of two universities: the American University of Kurdistan and the University of Zakho. The investigated sources of data are Galvanic Skin Response, eye-tracking, and System Usability Scale, as different layers of users' usability consciousness. Figure 2 depicts the proposed multi-layer bottom-up approach, along with the aligned website usability dimensions:

- *Unconscious Usability Data*: Including observable quantitative data about users' emotional arousal status in terms of skin conductance to evaluate the website for the metrics of satisfaction and errors.
- *Subconscious Usability Data*: Including observable quantitative data about users' visual attention using eye-tracking data to evaluate the website for effectiveness and efficiency metrics.
- *Conscious Usability Data*: Including observable qualitative data from users' usability experience using the standard SUS reports to evaluate the website for the metrics of memorability and learnability.

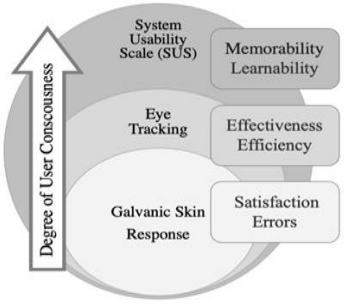

Figure 2: The proposed multilayer usability testing approach

4.2 Materials and Instruments

This study used the Gazepoint GP3 HD bundle[1], which is a research-grade instrument with an eye tracker and a biometric kit for real-time recording of GSR data. The sampling rate of 60Hz has been selected for this study to capture and record 60 samples of both eye-tracking and GSR data per second. It also comes with a [0.5-1] degree of visual angle accuracy and a 5/9-point calibration function.

4.3 Participants

According to the literature, more than 80% of the usability problems can be found with five or six users (Nielsen, 1999; Nielsen 2010), and as this number approaches 15, almost all of the usability problems can be identified, as in Figure 3. Therefore, to avoid missed usability problems, this study enrolled 18 users to identify usability problems on both AUK and UoZ university websites.

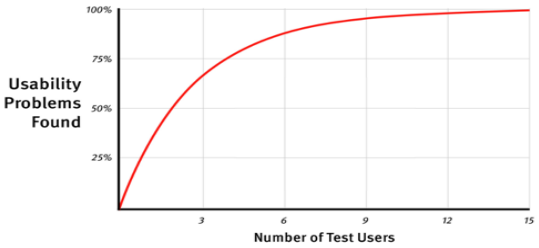

Figure 3: The number of users vs discovered usability problems.

The users were college students (average age of 22) and university staff (average age of 35) and they were experienced web users. Furthermore, to avoid any bias in the study, users are selected from AUK and UoZ universities. Their demographics are described in Table 1:

Table 1: Users' demographics

Group	Size	Gender	Working
AUK Users	9	5 Male, 4 Female	6 Staff, 3 Student
UoZ Users	9	5 Male, 4 Female	6 Staff, 3 Student

4.4 Experimental Protocol and Tasks

The experiments were conducted using the protocol adapted from Pernice (2009) and Białowąs (2021):

1. The users were provided with the consent form informing them that the GP3 device is unharmful to their bodies.
2. The experiments were performed with single-based sessions to minimize emotional stress.
3. To minimize the artifacts that dramatically can lower the quality of the biofeedback, the users are informed with special instructions such as breathing normally, with no or minimum movement, and talking.
4. The user sits in front of the monitor positioned properly for the eye-tracker device and has GSR sensors attached to the index and middle fingers.
5. The user starts the session with the calibration of eye-tracking and the biometric kits.
6. While the GP3 is recording GSR and eye-tracking data, the user starts performing website usability tasks.
7. At the end of the session, the SUS data is collected.

The users performed all the tasks on both the UoZ and AUK websites. Also, the order in which the website is used first is randomized among the users to collect more realistic data. The order of the website usability tasks remained the same as follows:

1. *Find "last News" published on the university website.*
2. *Open the Academic Calendar and find the "Date of Final Exam in Spring Semester."*
3. *Open the Department of Petroleum Engineering and find the "list of courses/study-plan."*
4. *Find the "contact information" of the university, email or phone number.*
5. *Open the IT/ICT department of the university, then find their "email address contact."*

Relaxation periods are inserted in between the tasks to emotionally separate each task. Then the total emotional experience of each user is calculated as the total number of the GSR peaks for the whole website usability testing session.

5. DISCUSSION OF RESULTS

In a multi-layer approach, this section presents the website usability results and their correlations with respect to the users' gender-based preferences.

5.1 Galvanic Skin Response

The raw GSR signal values of skin resistance collected from the biometric toolkits cannot be directly interpreted and thus need further signal pre-processing (Bakker, 2011). Following is an example of an actual female user's GSR data while using the UoZ website:

1. The GP3 biometric kit captures the GSR data with a sampling rate of 60Hz, meaning it records 60 samples of skin resistance per second in the unit of *ohms*, see Figure 4.
2. For ease of understanding and analysis of the collected GSR data, raw skin resistance samples are down-sampled with a factor of 60 by the use of *decimate* (in *scipy.org*) functions. Since it is a lowpass, it prefilters the high-frequency components of the signal and avoids aliasing effects by keeping the original wave trends from the original signal raw. This step aggregates each 60-signal sample into a 1-second signal sample, Figure 5.
3. Converting down-sampled signal of skin resistance to skin conductance with the standard unit of *microsiemens* (µS), which are just the reciprocal of each other, see Figure 6.
4. Skin conductance data was then smoothed by the use of a non-linear moving-median filter (Bakker, 2011), by subtracting the moving average value of the signal from the skin conductance signal value to remove unwanted tonic components of high wave spikes generated from different sources of the noise and keeping actual phasic component. Such noises include any movement, as in Figure 7.

[1] https://www.gazept.com

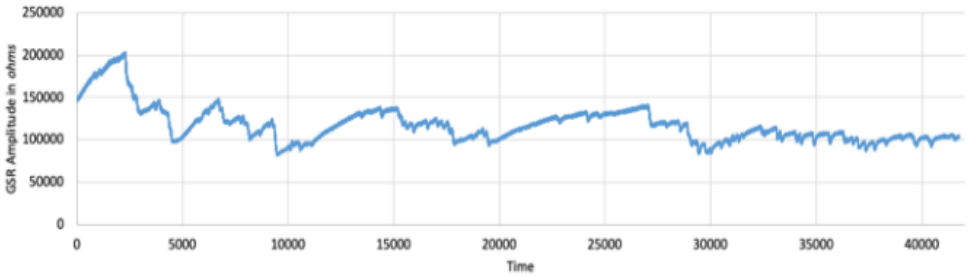

Figure 4: Raw skin resistance data in *ohms.*

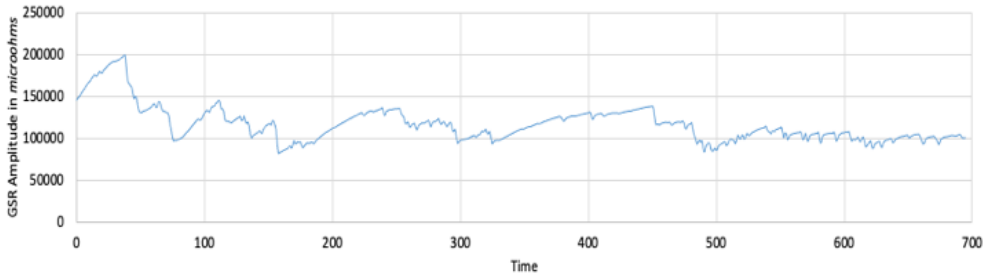

Figure 5: Down-sampled skin resistance data in *ohms.*

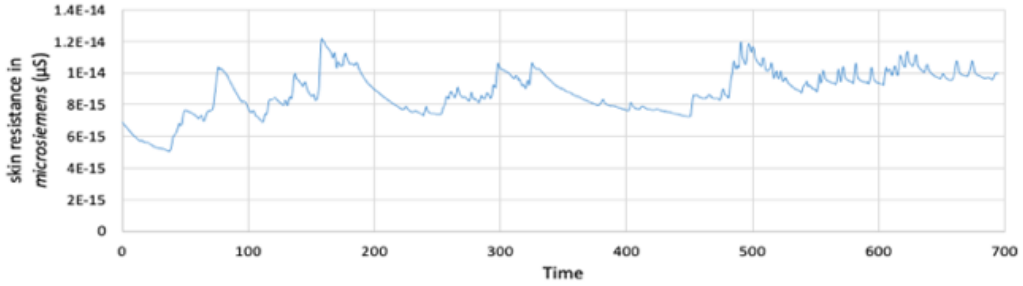

Figure 6: Skin resistance to skin conductance in *µS.*

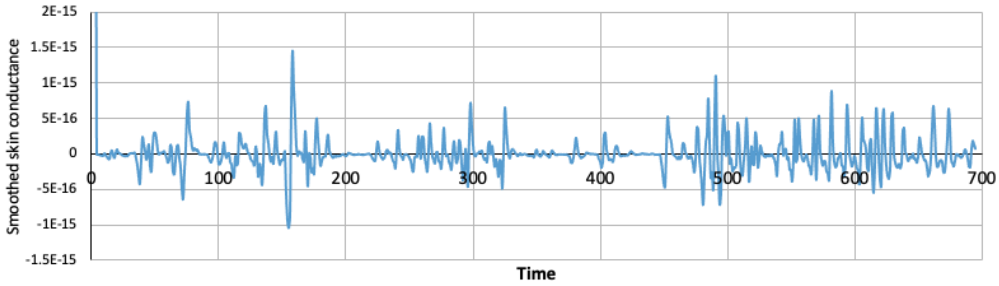

Figure 7. Smoothed skin resistance signal.

Emotional arousal is captured from skin response, which is considered by the number of peaks. The more peaks the user is experiencing the higher the emotional arousal is. Thus, the number of GSR peaks quantifies the emotional engagement of the users while using both websites. Figure 8 shows the number of GSR peaks observed per user for both AUK and UoZ websites. On average and across all users and on average, the AUK website produced more GSR peaks, (42.9 peaks) for the collective website users, meaning the users were experiencing more emotional arousal rather than relaxation with the AzK website than with the UoZ website. Whereas, on average the UoZ website produced a smaller number of peaks, (23.5 peaks), as in Figure 9.

As far as gender emotional differences, Figure 10 depicts both genders were more relaxed and preferred the UoZ website over the AUK website by producing a smaller number of peaks, Table 2. In general female users experienced a higher amount of emotional arousal with the (37.88) number of GSR peaks than the male users with the (29.4) number of GSR peaks.

5.2 Eye-tracking Data

Eye-tracking data uncovered the visual attention of users while browsing both websites and performing the usability tasks. Figures 11 and 12, present the Time-To-Complete (TTC) in seconds and the number of Fixation Counts (FC) metrics for both the AUK and UoZ websites respectively.

For the TTC metric measured in seconds, the users on average spent slightly more time accomplishing the tasks on UoZ than on AUK, table 3. In the case of the AUK website, the shorter TTC of tasks pushed the users to produce more fixation counts also experiencing a higher number of GSR peaks and FC counts. In other words, subconscious engagement of the visual attention did affect the unconscious emotional engagement of the users with the websites.

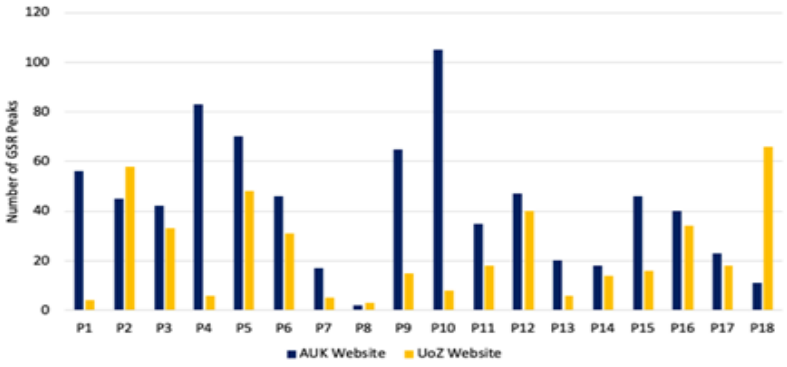

Figure 8. Number of GSR peaks for all users.

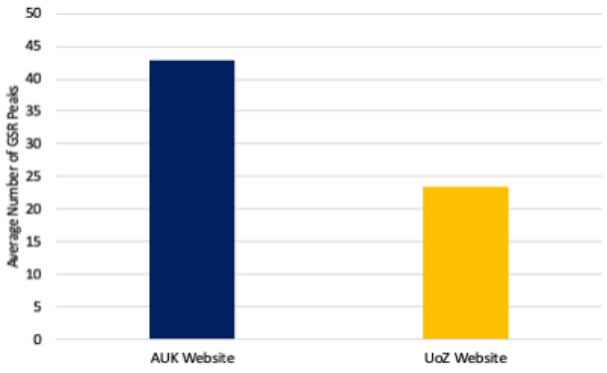

Figure 9. Average GSR peaks number per website.

Figure 10. Gender-emotional differences of both websites

Table 2: Gender-emotional differences between both websites

Gender	AUK Websites	UoZ Websites
Male	33.8	25
Female	54.1	21.6

Table 3: The results of TTC and FC for both websites.

Eye-tracking Metric	AUK Website	UoZ Website
Average TTC (Sec)	41.7	48.8
Average FC (#)	99.5	91

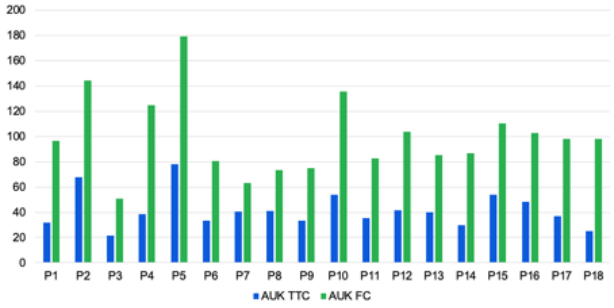

Figure 11. The TTC and FC results for the AUK website.

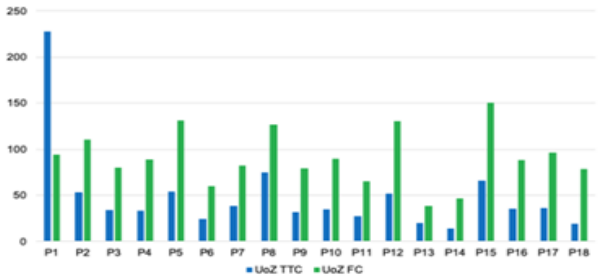

Figure 12. The TTC and FC results for the UoZ website.

On the other hand, the TTC metric of the UoZ website tells that the users spent more time while performing usability tasks therefore generating a smaller number of FC counts as an indicator of relaxation, which in turn interprets the reason for generating a smaller number of GSR peaks while using the UoZ website. To shorten, the AUK website overall appeared to be more emotionally arousal than the UoZ website, Figure 13.

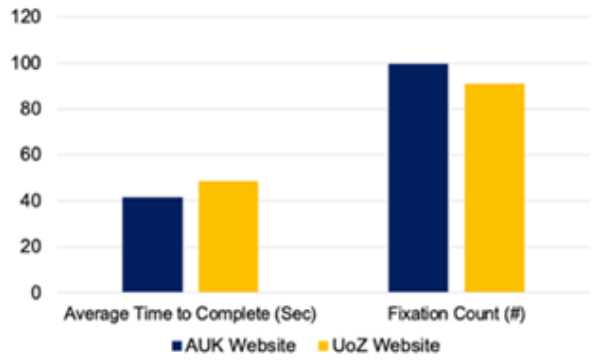

Figure 13. The results of TTC and FC for both websites.

As far as gender differences and in terms of speed of usage, based on the TTC metric male users performed slightly faster than female users' especially on the UoZ website, while on average both genders generated a similar number of fixation counts, see Table 4. In terms of GSR results, this can be related to the male users being more relaxed on average while performing usability tasks on both websites. Figure 14 shows the overall gender differences with respect to TTC and FC metrics.

Table 4: Eye-tracking usability results by genders.

Gender	TTC (Sec) AUK	TTC (Sec) UoZ	FC (#) AUK	FC (#) UoZ
Male	42.1	39.7	96.8	93.9
Female	41.3	60.3	103.1	87.3

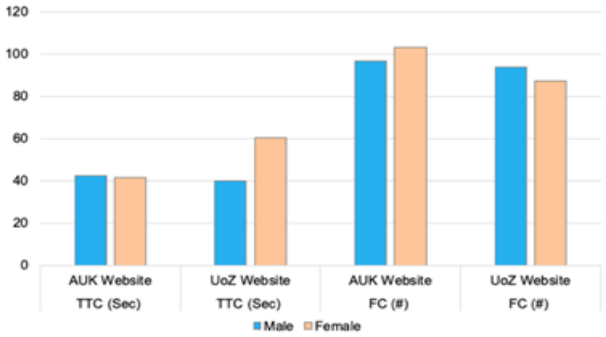

Figure 14. Gender differences in eye-tracking data.

5.3 System Usability Score (SUS)

The SUS is conscious usability feedback, which can interpret other layers of usability consciousness of emotional and visual attention engagements.

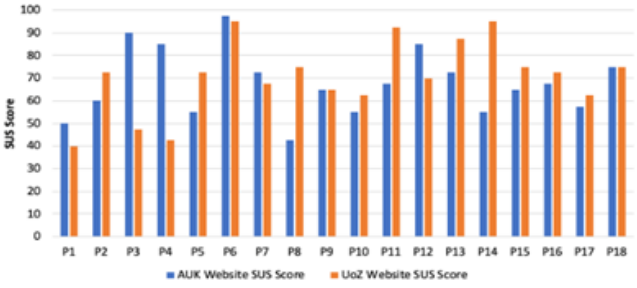

Figure 15: SUS scores for both websites across all users.

As far as the observed SUS usability scores in Table 5, on average the users scored the UoZ website higher than the AUK website. This conscious scoring behavior is considered to be related to the unconscious results of emotional engagement of the users, as in the case of the UoZ website users generated fewer GSR peaks than the AUK website, meaning that the overall users were more relaxed with the UoZ website. Based on the acceptability rating depicted in Figure 1, the UoZ SUS score of (70.6) is considered "acceptable" while the AUK SUS score of (67.6) is considered to be "marginal high", yet on the adjective scale both are considered "okay".

Table 5: Average SUS scores for both websites for all users.

Website	Average SUS Score
AUK	67.6
UoZ	70.6

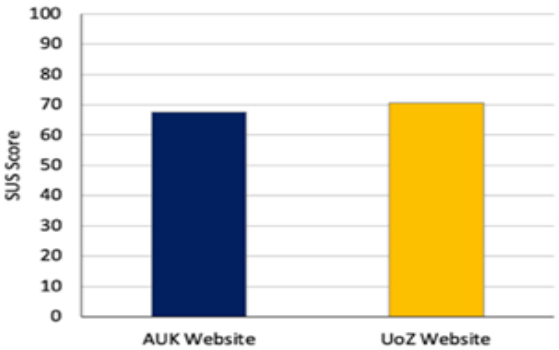

Figure 16 Average SUS scores for both websites for all users.

On the other hand, the AUK website gained a lower SUS score, which can be related to having higher fixation counts and a

higher number of GSR peaks, Figure 16. However, this does not reject the fact that on average users spent less time completing the tasks on the AUK website, which in turn can be related to having higher emotional arousal (by the higher number of GSR peaks) causing the users to perform usability tasks faster than UoZ website, with a higher number of fixation counts.

Eventually, this unrelaxing unconscious experience forced the users to assign lower conscious SUS scores to the usability of the AUK website despite performing tasks faster than the UoZ website, as in Table 6. This confirms the previous finding in the literature on poor usability experience which showed higher levels of emotional arousal (Lin, 2005; De Carolis, 2023) and that improved website design requires fewer user fixations (Çınar, 2009), Figure 17. In terms of gender preferences and based on SUS scoring, male users mostly preferred the UoZ website over the AUK website whereas female users preferred the AUK website over the UoZ website. This also can be related to the fact that compared to female users, male users experienced a lower number of GSR peaks and FC values than female users on the UoZ website.

Table 6 :Gender preferences based on SUS scores for both websites.

Gender	Website	SUS Score
Male	AUK	65.3
	UoZ	75.5
Female	AUK	68.9
	UoZ	64.4

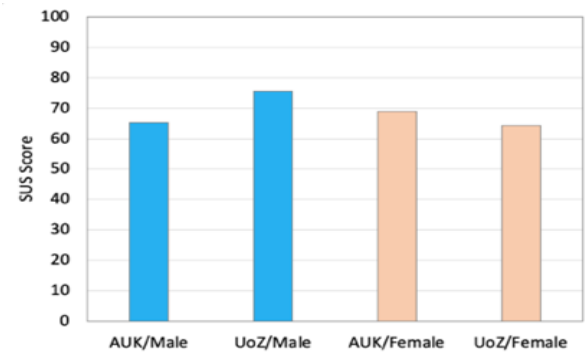

Figure 17. Gender preferences based on SUS scores.

5.4 Usability Design Recommendations

Website usability testing typically serves as a means of identifying areas for redesign and improvement. While quantitative usability testing numerically answers the questions of "What are the usability problems and who encountered them?", qualitative usability evaluation verbally answers the questions of "Why and how the users have experienced usability problems?". In this context, qualitative usability evaluation is done using two methodologies.

The first is by verbally collecting users' self-reporting feedback (post-test questionnaire) by answering both classes of questions about why they have had a special experience (positive or negative) and how they have en

The qualitative post-test questionnaire included a set of questions as follows, adapted from Kokil (2017):

1. *Why did you like or dislike the site?*
2. *Why you were confused while navigating the website?*
3. *Why is this website easy or hard to use?*
4. *How was the information organized on the website?*
5. *How pleasant was the appearance of the website?*
6. *How the content was useful and informative?*

Second, the use of heatmap visualizations of eye-tracking data, like those seen in Figures 18 and 19, provided insights into the users' visual attention on the UoZ and AUK websites. These heatmaps utilize color-coded representations, where red areas signify points of interest with a greater concentration of gaze points, indicating higher user interest. Conversely, yellow and green areas represent lower user interest (Jiang, 2020). Other visualizations, such as *fixation maps*, *opacity maps*, and *bee swarms*, also contribute to the understanding of user attention patterns (Isokoski, 2018).

Below is a list of design and content-oriented problems and recommendations captured in the AUK and UoZ websites as the results of performing qualitative usability evaluation:

- The users were confused by the mixed and less unstructured information on the AUK website about the organizational structure of colleges, departments, and programs. While in UoZ, the information on college, department, and program was more separated and organized.
- Unclear grouping of content, such as IT department in the AUK website.
- Existence of the hardly readable content and visual elements, such as having the text and its background either both dark or light, especially on the UoZ website, produced longer TTC timing.
- About 80% of the users did not scroll down the web pages, instead, they insisted on finding their target elements in the top part of the webpages, which increased the TTC results in UoZ and higher FC in AUK websites.
- Other confusing factors that did not meet user expectations regarding content placement were:
- The News section was expected to be on the far right of the menu on the UoZ website.
  - The Academic Calendar was expected to be in the Academics menu on the UoZ website.
  - Users expected to have A-to-Z full lists of programs or university-wide staff and units on both websites.
  - For both websites, the homepage *slideshow* distracts users' attention toward concentrating on their visual scanning path.
  - Having dynamic structural visualizations for both websites' main menu and other content.
  - Expecting further shortcuts for the commonly used content in the form of icons on the AUK website, such as the academic calendar and university e-learning system.

What is also notable from visual patterns is that the users did not follow the traditional visual patterns of type F-Pattern or E-Pattern (Djamasbi, 2011) as in e-commerce websites, instead, it was mostly center-based, where the users started looking at the center of the page and then moving around to search their target elements. As part of the content placement strategy, this center-based pattern suggests placing high-priority content in the center of the web pages to maximize the usability results.

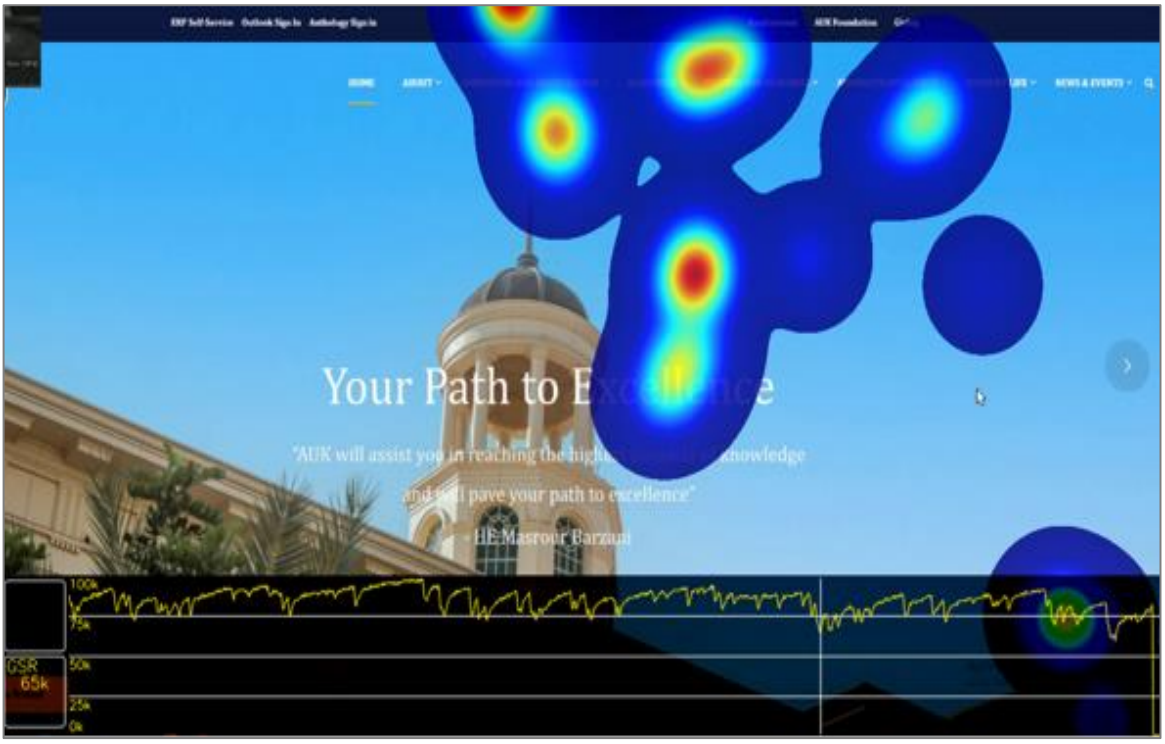

Figure 18: Heatmap visualization on the AUK homepage.

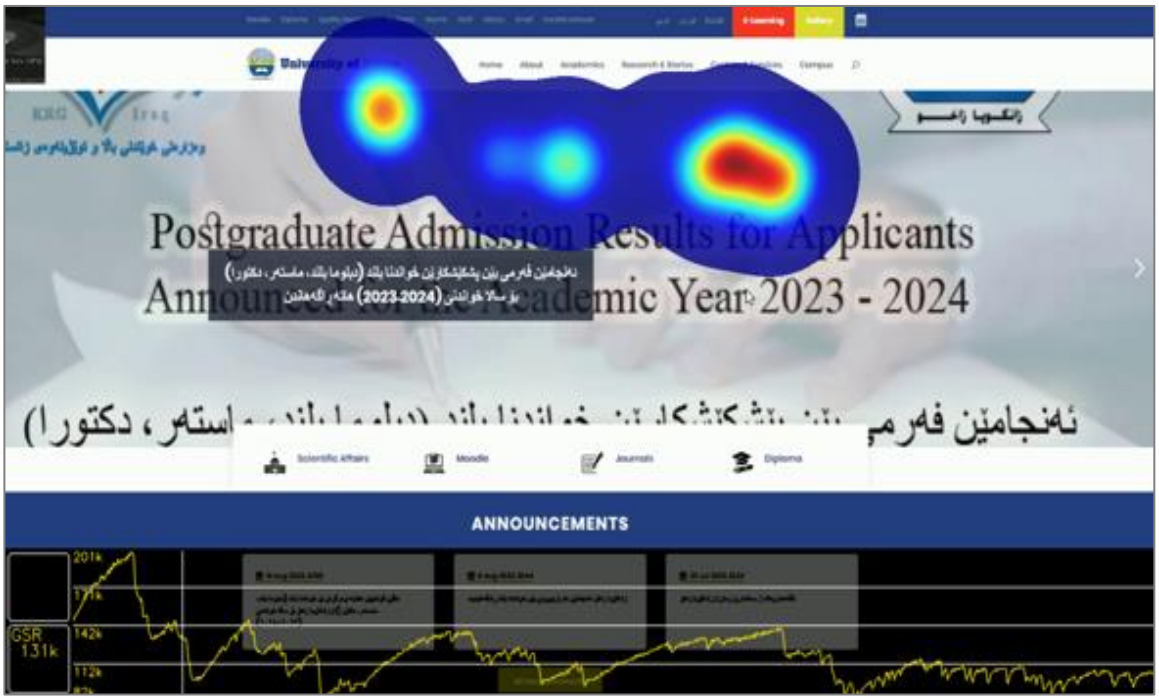

Figure 19: Heatmap visualization on the UoZ homepage.

6. CONCLUSION

This study presented a comprehensive usability testing approach, considering users' emotional engagement, visual attention, and self-reported feedback. It conducted usability testing on the AUK and UoZ university websites with 18 participants using the Gazepoint GP3 system. The results revealed a pattern of correlations between different user consciousness levels, highlighting the influence of emotional engagement on fixation counts and completion time, resulting in varying SUS scores. As the discussion of the results depicted, some notable correlations were observed among the three layers of user consciousness depicted in Figure 2, as follows:

1. *A negative correlation between emotional arousal and each of the time-to-complete, and SUS scoring.*
2. *A positive correlation between emotional arousal and fixation counts.*
3. *Female users experience higher emotional arousal.*

This was obvious in the case of the AUK website with its shallow information hierarchy where the high emotional engagement was related to reduced TTC and increased FC, leading to unsatisfied and lower SUS scores, which confirms previous findings that higher emotional arousal is related to unpleasant user experiences (Ward, 2002; Wang, 2014). On the other hand, the UoZ website with its better-organized structure offered a lower emotional engagement which in turn related to the extended TTC and decreased FC, resulting in higher SUS scores, thus confirming the early findings that better information organization yields higher usability satisfaction (Çınar, 2009; De Carolis, 2023). The gender-based findings confirmed the existing literature (Bari, 2020), indicating generally female users experience higher emotional arousal.

Conclusively, the study emphasized the importance of involving different user levels of user consciousness in website usability testing and the importance of considering gender preferences in website design and development. This study is limited to academic website settings. Additional research work is needed to examine the proposed approach in other contexts of mobile and web applications. The notable challenge was not having a single hub for interpreting and correlating the eye-tracking, GSR, and SUS data. Future research includes considering the amplitude and recovery time of the GSR peaks. Also, performing task-based analysis of the results.

7. ACKNOWLEDGMENTS

The author would like to thank all the users who were part of this study including the undergraduate students of the Department of General Psychology who helped in organizing the experiments.

REFERENCES

Abran, A., Khelifi, A., Suryn, W., & Seffah, A. (2003, April). Consolidating the ISO usability models. In Proceedings of 11th International Software Quality Management Conference (Vol. 2003, pp. 23-25).

Anders, S., Lotze, M., Erb, M., Grodd, W., & Birbaumer, N. (2004). Brain activity underlying emotional valence and arousal: A response-related fMRI study. Human brain mapping, 23(4), 200-209.

Aziz, N. S., Sulaiman, N. S., Hassan, W. N. I. T. M., Zakaria, N. L., & Yaacob, A. (2021, May). A Review of Website Measurement for Website Usability Evaluation. In Journal of Physics: Conference Series (Vol. 1874, No. 1, p. 012045). IOP Publishing.

Baig, M. Z., & Kavakli, M. (2019). A survey on psycho-physiological analysis & measurement methods in multimodal systems. Multimodal Technologies and Interaction, 3(2), 37.

Bakker, J., Pechenizkiy, M., & Sidorova, N. (2011, December). What's your current stress level? Detection of stress patterns from GSR sensor data. In 2011 IEEE 11th International Conference on Data Mining Workshops (pp. 573-580). IEEE.

Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. Journal of usability studies, 4(3), 114-123.

Bari, D. S. (2020). Gender differences in tonic and phasic electrodermal activity components. Science Journal of University of Zakho, 8(1), 29-33.

Bari, D. S., Aldosky, H. Y. Y., Tronstad, C., Kalvøy, H., & Martinsen, Ø. G. (2018). Electrodermal responses to discrete stimuli measured by skin conductance, skin potential, and skin susceptance. Skin Research and Technology, 24(1), 108-116.

Bari, D. S., Rammoo, M. N. S., Aldosky, H. Y., Jaqsi, M. K., & Martinsen, Ø. G. (2023). The Five Basic Human Senses Evoke Electrodermal Activity. Sensors, 23(19), 8181.

Barnum, C. M. (2020). Usability testing essentials: Ready, set... test!. Morgan Kaufmann.

Bergstrom, J. R., & Schall, A. (Eds.). (2014). Eye-tracking in user experience design. Elsevier.

Berkovsky, Shlomo, et al. "Detecting personality traits using eye-tracking data." Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019.

Białowąs, S., Pieranski, B., Szyszka, A., & Reshetkova, A. (2021). Experimental design and biometric research. Toward innovations. Poznań University of Economics and Business Press. Chicago.

Blascheck, T., Kurzhals, K., Raschke, M., Burch, M., Weiskopf, D., & Ertl, T. (2014, June). State-of-the-art of visualization for eye-tracking data. In Eurovis (stars) (p. 29).

Boucsein, W. (2012). Electrodermal activity. Springer Science & Business Media.

Brooke, J. (1996). SUS: A Quick and Dirty Usability Scale. Usability evaluation in industry, 189(3), 189-194.

Buchanan, S., & Salako, A. (2009). Evaluating the usability and usefulness of a digital library. Library Review, 58(9), 638-651.

Carter, B. T., & Luke, S. G. (2020). Best practices in eye-tracking research. International Journal of Psychophysiology, 155, 49-62.

Çınar, M. O. (2009). Eye-tracking method to compare the usability of university websites: A case study. In Human Centered Design: First International Conference, HCD 2009, Held as Part of HCI International 2009, San Diego, CA, USA, July 19-24, 2009 Proceedings 1 (pp. 671-678). Springer Berlin Heidelberg.

Clay, Viviane, Peter König, and Sabine Koenig. "eye-tracking in virtual reality." Journal of Eye Movement Research 12.1 (2019).

Critchley, H. D. (2002). Electrodermal responses: what happens in the brain. The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry, 8(2), 132–142.

Dawson, M. E., Schell, A. M., & Filion, D. L. (2007). The electrodermal system. Handbook of psychophysiology, 2, 200-223.

De Carolis, B., Loglisci, C., Giuseppe, M., & Trufanova, K. (2023, September). Analyzing Stress Responses Related to Usability of User Interfaces. In Proceedings of the 15th Biannual Conference of the Italian SIGCHI Chapter (pp. 1-9).

Djamasbi, S. (2014). Eye-tracking and web experience. AIS Transactions on Human-Computer Interaction, 6(2), 37-54.

Djamasbi, S., Siegel, M., & Tullis, T. (2011). Visual hierarchy and viewing behavior: An eye-tracking study. In Human-Computer Interaction. Design and Development Approaches: 14th International Conference, HCI International 2011, Orlando, FL, USA, July 9-14, 2011, Proceedings, Part I 14 (pp. 331-340). Springer Berlin Heidelberg.

Foglia, P., Prete, C. A., & Zanda, M. (2008, May). Relating GSR signals to traditional usability metrics: Case study with an anthropomorphic web assistant. In 2008 IEEE Instrumentation and Measurement Technology Conference (pp. 1814-1818). IEEE.

Foglia, P., Zanda, M., & Trading, I. O. N. (2014). Towards relating physiological signals to usability metrics: a case study with a web avatar. WSEAS Transactions on Computers, 13, 624.

Fowles, D. C., Christie, M. J., Edelberg, R., Grings, W. W., Lykken, D. T., & Venables, P. H. (1981). Publication recommendations for electrodermal measurements. Psychophysiology, 18(3), 232-239.

Fraiwan, L., Basmaji, T., & Hassanin, O. (2018, November). A mobile mental health monitoring system: a smart glove. In 2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS) (pp. 235-240). IEEE.

Goldberg, J. H., & Wichansky, A. M. (2003). Eye-tracking in usability evaluation: A practitioner's guide. In the Mind's Eye(pp. 493-516). North-Holland.

Green, D., & Pearson, J. M. (2006). Development of a web site usability instrument based on ISO 9241-11. Journal of Computer Information Systems, 47(1), 66-72.

Huang, Z., & Mou, J. (2021). Gender differences in user perception of usability and performance of online travel agency websites. Technology in Society, 66, 101671.

Isokoski, P., Kangas, J., & Majaranta, P. (2018, June). Useful approaches to exploratory analysis of gaze data: enhanced heatmaps, cluster maps, and transition maps. In Proceedings of the 2018 ACM Symposium on Eye-tracking Research & Applications (pp. 1-9).

Jiang, Y. (2020). A solution to analyze mobile eye-tracking data for user research in GI Science (Master's thesis, University of Twente).

Klaib, Ahmad F., et al. "eye-tracking algorithms, techniques, tools, and applications with an emphasis on machine learning and Internet of Things technologies." Expert Systems with Applications 166 (2021): 114037.

Kokil, U., & Scott, S. (2017, February). Usability testing of a school website using qualitative approach. In International Conference on Human Computer Interaction Theory and Applications (Vol. 3, pp. 55-64). SCITEPRESS.

Lin, T., Omata, M., Hu, W., & Imamiya, A. (2005, November). Do physiological data relate to traditional usability indexes?. In Proceedings of the 17th Australia conference on computer-human interaction: Citizens online: Considerations for today and the future (pp. 1-10).

McNeal, K. S., Zhong, M., Soltis, N. A., Doukopoulos, L., Johnson, E. T., Courtney, S., ... & Porch, M. (2020). Biosensors show promise as a measure of student engagement in a large introductory biology course. CBE—Life Sciences Education, 19(4), ar50.

Nielsen, J. (1999). Designing web usability: The practice of simplicity. New riders publishing.

Nielsen, J., & Pernice, K. (2010). Eyetracking web usability. New Riders.

Pernice, K., & Nielsen, J. (2009). How to conduct eyetracking studies. Nielsen Norman Group.

Rinder, J. (2012). The importance of website usability testing.

Satti, F. A., Hussain, M., Hussain, J., Kim, T. S., Lee, S., & Chung, T. (2021, January). User stress modeling through galvanic skin response. In 2021 15th International Conference on Ubiquitous Information Management and Communication (IMCOM) (pp. 1-6). IEEE.

Sauro, J., & Lewis, J. R. (2011, May). When designing usability questionnaires, does it hurt to be positive?. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 2215-2224).

Sharafi, Zohreh, et al. "A practical guide on conducting eye-tracking studies in software engineering." Empirical Software Engineering 25 (2020): 3128-3174.

Stahlke, Samantha N., et al. "Frontiers of immersive gaming technology: A survey of novel game interaction design and serious games for cognition." Recent Advances in Technologies for Inclusive Well-Being: Virtual Patients, Gamification and Simulation (2021): 523-536.

Tang, Wilson YF. "Application of Eye Tracker to Detect Visual Processing of Children with Autism Spectrum Disorder." Current Developmental Disorders Reports 9.4 (2022): 77-88.

Tyler, W. J., Boasso, A. M., Mortimore, H. M., Silva, R. S., Charlesworth, J. D., Marlin, M. A., ... & Pal, S. K. (2015). Transdermal neuromodulation of noradrenergic activity suppresses psychophysiological and biochemical stress responses in humans. Scientific reports, 5(1), 13865.

Wang, J., Antonenko, P., Celepkolu, M., Jimenez, Y., Fieldman, E., & Fieldman, A. (2019). Exploring relationships between eye-tracking and traditional usability testing

data. International Journal of Human–Computer Interaction, 35(6), 483-494.

Wang, Q., Yang, S., Liu, M., Cao, Z., & Ma, Q. (2014). An eye-tracking study of website complexity from cognitive load perspective. Decision support systems, 62, 1-10.

Ward, R. D., Marsden, P. H., Cahill, B., & Johnson, C. (2002, April). Physiological responses to well-designed and poorly designed interfaces. In Proceedings of CHI 2002 workshop on physiological computing.

Zardari, B. A., Hussain, Z., Arain, A. A., Rizvi, W. H., & Vighio, M. S. (2021). QUEST e-learning portal: Applying heuristic evaluation, usability testing and eye tracking. Universal Access in the Information Society, 20, 531-543.