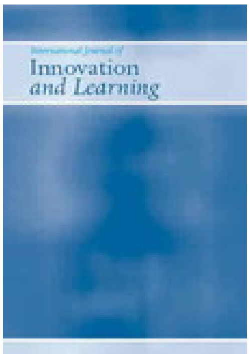


International Journal of Innovation and Learning

ISSN online: 1741-8089 - ISSN print: 1471-8197
https://www.inderscience.com/ijil

Observing cognitive load during online learning with various task complexities: an eye tracking approach

Prabaria Vesca Yulianandra, Suatmi Murnani, Paulus Insap Santosa, Sunu Wibirama

DOI: 10.1504/IJIL.2023.10051446

Article History:

Received:	18 February 2022
Accepted:	12 May 2022
Published online:	07 July 2023

Observing cognitive load during online learning with various task complexities: an eye tracking approach

Prabaria Vesca Yulianandra

Department of Electrical and Information Engineering,
Faculty of Engineering,
Universitas Gadjah Mada,
Yogyakarta, 55281, Indonesia
Email: prabaria.mti15@mail.ugm.ac.id

Suatmi Murnani

Department of Electrical Engineering,
Universitas Islam Indonesia,
Yogyakarta, 55584, Indonesia
Email: suatmi.murnani@uii.ac.id

Paulus Insap Santosa and Sunu Wibirama*

Department of Electrical and Information Engineering,
Faculty of Engineering,
Universitas Gadjah Mada,
Yogyakarta, 55281, Indonesia
Email: insap@ugm.ac.id
Email: sunu@ugm.ac.id
*Corresponding author

Abstract: E-learning has been used to support distance education during the COVID-19 pandemic. Unfortunately, little attention has been paid to the relationship between design complexity of an e-learning system, task complexity, and users' cognitive load. Here we conducted a novel investigation to observe effects of design complexity and task complexity towards users' cognitive load. Each group of participants was exposed to different interfaces of e-learning: low, medium, and high design complexity. Participants were asked to perform both simple and complex tasks. We used four instruments: eye tracking, cognitive load questionnaire, system usability scale (SUS), and user experience questionnaire (UEQ). Experimental results show that task complexity and design complexity significantly affect the eye tracking metrics (p < 0.05) and scores of cognitive load questionnaire (p < 0.05). Based on experimental results, we recommend an e-learning system with medium complexity to achieve minimum cognitive burden in online learning during the COVID-19 pandemic.

Keywords: user experience; e-learning; design complexity; task complexity; cognitive load; eye tracking.

Reference to this paper should be made as follows: Yulianandra, P.V., Murnani, S., Santosa, P.I. and Wibirama, S. (2023) 'Observing cognitive load during online learning with various task complexities: an eye tracking approach', *Int. J. Innovation and Learning*, Vol. 34, No. 1, pp.96–117.

Biographical notes: Prabaria Vesca Yulianandra received her BEng from the Department of Informatics Engineering, Yogyakarta State University, Indonesia. She received an MEng from the Department of Electrical and Information Engineering, Faculty of Engineering, Universitas Gadjah Mada Indonesia. Her research interests are learning technologies, human-computer interaction, and ICT in education.

Suatmi Murnani received her BEng and MEng in Electrical Engineering from the Universitas Gadjah Mada, Indonesia, in 2011 and 2020, respectively. She is currently a Junior Lecturer at the Department of Electrical Engineering, Universitas Islam Indonesia. Her general topics of interest include signal processing and machine learning.

Paulus Insap Santosa serves as an Associate Professor in the Department of Electrical and Information Engineering, Faculty of Engineering, Universitas Gadjah Mada, Indonesia. He received his BEng from the Universitas Gadjah Mada, Indonesia. He earned his MSc from the University of Colorado Boulder, USA. He received his PhD from the Department of Information Systems, School of Computing, National University of Singapore. His research interests include human-computer interaction, user experience and usability, web design, e-learning, and ICT in education.

Sunu Wibirama received his BEng from the Universitas Gadjah Mada, Indonesia, in 2007, MEng in Electronics Engineering from the King Mongkut's Institute of Technology Ladkrabang, Thailand, in 2010, and DrEng in Science and Technology from the Tokai University, Japan, in 2014. He is an Associate Professor with the Department of Electrical and Information Engineering, Faculty of Engineering, Universitas Gadjah Mada, Indonesia. His main research interests are intelligent user experience engineering, human-computer interaction, and eye-gaze tracking applications.

1 Introduction

Only a few weeks after the World Health Organization announced COVID-19 as a global pandemic (WHO Regional Office for Europe, 2020), more than 100 countries took a countermeasure to prevent the spread of COVID-19 in their educational systems. According to UNESCO (2020), the pandemic has negatively affected experience of 1.6 billions students worldwide, particularly in a country that implements nationwide schools and universities closure. In Indonesia, this policy impacted more than 68 million students across the country (UNESCO, 2021).

To mitigate this adverse situation, e-learning system has been inevitably used to support distance education in Indonesia (Rahiem, 2020). E-learning system is a web-based software that facilitates delivery of learning contents and manages the educational progress of online learners (Islam, 2016). The use of e-learning system in distance education provides several advantages over traditional or face-to-face education (Yang, 2018). E-learning enables borderless education between teachers and students regardless of their geographical locations – an apparent situation during the COVID-19 outbreak. In addition, various gamification strategies can be adopted in an e-learning system to improve satisfaction during learning activities (Park et al., 2022). More importantly, e-learning also provides flexibility in accessing teaching materials – promoting independent and asynchronous learning (i.e., non-real-time interaction) for students who encounter difficulty to access high-speed internet connection (Chang, 2016).

Despite of these advantages, there are several factors for successful implementation of e-learning system in distance education. In a recent study, Siron et al. (2020) collected online response from 210 university students in Indonesia to investigate some underlying factors of e-learning system adoption during the outbreak. Siron et al. found that perceived ease of use, students' experience, and self-efficacy were among several factors that affected e-learning implementation during distance learning. In the same year, Sukendro et al. (2020) conducted a survey of e-learning usage in five Indonesian higher education institutions. The study had successfully gathered data from 974 students majoring in sport science education. Similar with Siron et al., Sukendro et al. suggested that perceived ease of use significantly predicted perceived usefulness.

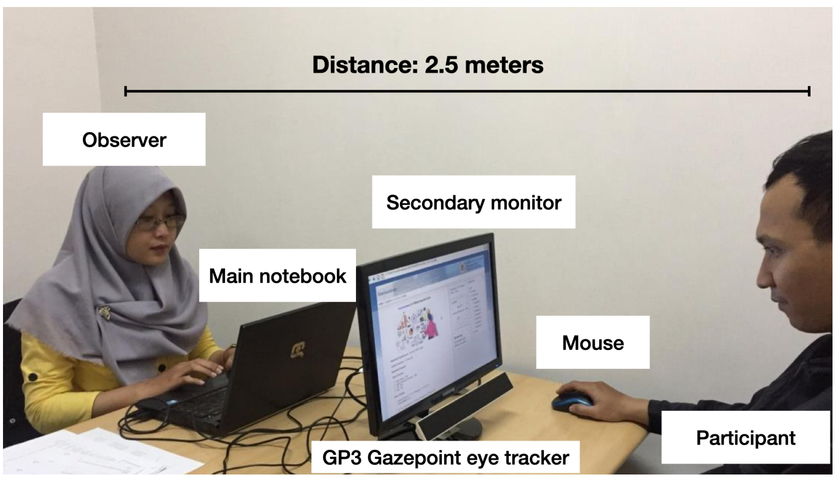
On the other hand, interaction between students and an e-learning system can also be understood in terms of their cognitive load and perception (Seufert, 2018). Normally, cognitive load can be defined if the user is equipped with tasks at hand (Schneider et al., 2018). As the cognitive load of students increases, their ability to work effectively will slowly decrease until it reaches an overload point (Kalyuga, 2007). The increment in students' cognitive load leads to a negative perception of e-learning website. In addition, understanding the cognitive load during interacting with e-learning can lead to a better understanding of what is considered to be an effective e-learning system (Lambert et al., 2009). According to Van Merriënboer and Sweller (2005), design complexity and task complexity are two main contributing factors that affect cognitive load of e-learning users.

As stated by Berlyne (1960), design complexity refers to amount of variety or diversity in patterns of stimulus – in this case, the user interface of an e-learning system. Lambert et al. (2009) suggested that a complex e-learning has a potential to provide more engaging learning experiences to the students compared with a simpler one. On the other hand, Rosch and Vogel-Walcutt (2013) believed that e-learning should be designed to contain only information needed to complete the designated tasks. Al-Rahmi et al. (2019) found that e-learning complexity is directly correlated with perception on less usefulness. Interestingly, another study found an inverted U relationship between design complexity with preference of young students (Wang and Lin, 2019). Thus, an e-learning system with medium complexity is preferred over simple and high complexity.

Task complexity can be defined as the amount of information related to task an individual has to process when performing a task (Wood, 1986). The more information needs to be processed, the more complex the task is. Yanson and Johnson (2016) observed that students who received training with simpler tasks performed better than who received training with more complex tasks. A more recent study by O'Brien et al. (2020) investigated relationship between task complexity with user engagement during online search behaviour. They found that greater task complexity led to negative effect on engagement, suggesting that simpler task was preferable over the complex one. Wang et al. (2020) conducted a study in 3D virtual environment to investigate whether there was any relationship between learning styles and task complexity. They found that there was no significant relationship between task complexity and learning styles in virtual environment.

Despite of these extensive studies on design complexity and task complexity, prior studies have not considered how design complexity and task complexity affect cognitive load during interaction with an e-learning system. In addition, prior studies were mostly based on self-reported evaluations or written questionnaires. Hence, results of prior study are limited to subjective point of view without comparable measures from behavioural metrics such as eye movements or heart rate variability (Furnham and Henderson, 1982; Wibirama et al., 2018, 2020).

Figure 1 Experiment setup of this study (see online version for colours)



Notes: To ensure physical distancing, the participant was positioned two and a half metres away from the observer.

Source: Yulianandra (2018)

To address this research gap, we present a novel study to enrich the understanding of how the complexity of e-learning design and task complexity affect users' cognitive load. We combined eye tracking and cognitive load questionnaire to provide more objective measures on cognitive process during interaction with an e-learning system (Djamasbi et al., 2010; Lai et al., 2013). To provide general recommendation on e-learning design complexity, system usability scale (SUS) questionnaire and user experience questionnaire (UEQ) were used to complement the eye tracking results.

2 Materials and methods

2.1 Apparatus

Figure 1 shows configuration of the experiment. GP3 Gazepoint Eye Tracker (Gazepoint Research Inc., Canada) was used to record and to analyse fixational eye movements

(Wijayanto et al., 2016). The eye tracker was mounted beneath a standalone stimulus unit. The eye tracker implemented the pupil-corneal reflection (PCR) method based on morphological image processing and ellipse fitting to measure the direction of the gaze (Nugroho et al., 2015; Satriya et al., 2016; Setiawan et al., 2018). The eye tracking system was equipped with near-infrared (NIR) camera, infrared illumination sources, and visible light filter to produce high contrast video regardless of light intensities of the experiment room. The sampling rate was 60 Hz, with 0.5°–1° of accuracy and less than 0.1° of spatial resolution. The operation range was 50–80 cm from the computer screen with less than 50 ms of system latency, providing sufficient support for this study. Recording and analysis of eye movements data were performed in Gazepoint Analysis UX Edition (Wibirama et al., 2020).

Table 1 Group design in this study

	Design complexity		
	High	*Medium*	*Low*
Simple task	Group 1 (14 participants)	Group 2 (14 participants)	Group 3 (14 participants)
Complex task	Group 1 (14 participants)	Group 2 (14 participants)	Group 3 (14 participants)

Source: Yulianandra (2018)

Table 2 Design of e-learning complexity

Design complexity	*Page*	*Page length (px)*	*Amount of hyperlinks*	*Amount of pictures*	*Amount of blocks*
Low	Home	1,349 × 1,270	30	1	1
	Course	1,349 × 2,028	29	1	1
Medium	Home	1,349 × 2,148	42	3	3
	Course	1,349 × 2,923	40	5	3
High	Home	1,349 × 3,223	54	7	6
	Course	1,349 × 4,371	50	11	6

Source: Yulianandra (2018)

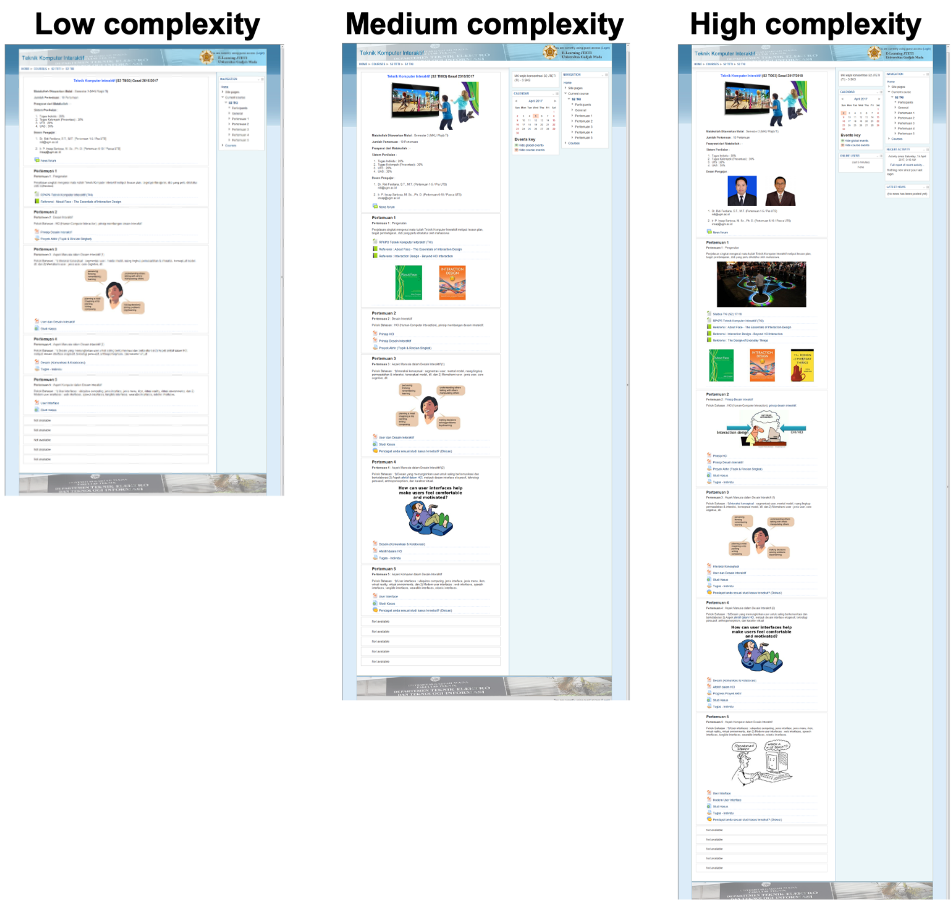
2.2 Participants

Participants were students of Universitas Gadjah Mada, Indonesia who were recruited on a voluntary basis (Yulianandra, 2018). The participants joined the experiment after filling in an informed consent form. Among 42 participants, 25 were males and 17 were females with an average age of 24.31 years old ranging from 20 to 32 years old. Among them, 11 participants were undergraduate students and 31 participants were graduate students. All participants were familiar with web-based e-learning. All participants were healthy, with normal or corrected eyes. The authors confirmed that all experimental procedures were arranged according to the WMA Declaration of Helsinki (*Ethical Principles for Medical Research Involving Human Subjects*).

2.3 Design of experiment

In this research, we conducted an experiment with a 3 × 2 design experiment (3 levels of e-learning design complexity × 2 levels of task complexity). The experiment implemented a split-plot design – applying between-subject and within-subject design for e-learning design complexity and task complexity, respectively. The participants were divided into three groups based on e-learning design complexity as shown in Table 1. Each group consisted of 14 participants and was exposed to only one out of three e-learning designs: high complexity, medium complexity, or low complexity (see Figure 2). During interacting with each design, the participant had to perform two tasks: simple and complex tasks. We applied counterbalance for both tasks to avoid systematic bias and learning effect.

Figure 2 Manipulation of e-learning design complexity: low complexity, medium complexity and high complexity (see online version for colours)



(a) (b) (c)

Source: Yulianandra (2018)

2.3.1 Manipulation of design complexity

Table 2 presents manipulation of e-learning design complexity as suggested by Geissler et al. (2001). The manipulation was performed on the home page and the course page by varying the length of the page, the number of hyperlinks, the number of pictures, and the number of blocks – such as calendar or clock navigation. Figure 2 shows examples of e-learning homepage with three levels of design complexity: low complexity, medium complexity, and high complexity.

A pretest study involving 15 participants was performed to ensure the e-learning designs had been manipulated according to the specified complexity (Wang et al., 2014). All participants were asked to assess the complexity of each design. The complexity was rated using a questionnaire consisted of single question with seven-scale rating. The result showed there was a significant difference between each design complexity ($F(1, 14) = 211.413$, $p < 0.05$). Based on this pretest study, e-learning designs with low, medium, and high complexity were rated with average complexity scores of 1.73, 4.27, and 6.13, respectively. These results implied that design manipulation was valid for further used in this research.

2.3.2 Manipulation of task complexity

Manipulation for task complexity was performed according to the studies of Leuthold et al. (2011) and Wang et al. (2014). Each task complexity – both simple and complex – was designed with three activities according to a study by Ramakrisnan et al. (2012): finding a certain subject, finding a link to download certain lecture notes, and finding a link to upload a certain assignment.

The simple task was designed to be a clear and direct task. Participants were instructed to finish the task with clear guidance on the required activities to be performed. In addition, the guidance explicitly mentioned name of the subject, file to be downloaded in specific location inside the subject, and file to be uploaded in the subject.

On the other hand, the complex task was designed to be more complicated as it required higher cognitive load, such as comparing several subjects to be selected as a target based on specific criteria. The guidance explicitly mentioned name of file to be downloaded in the designated target subject. However, there was no exact information on lecture session that contained the target file. Same approach was applied for the file to be uploaded. The guidance specifically referred the name of file to be uploaded, but the participants should find relevant lecture session to upload the designated assignment.

Similar to manipulation of design complexity, manipulation of task complexity was assessed by 15 participants. Participants were asked to fill in a questionnaire containing single question with seven-scale rating (Wang et al., 2014). Average scores for simple and complex tasks were 2.10 and 5.77, respectively. Statistical analysis showed that there was significant difference between both tasks ($t(14) = -30.471$, $p < 0.05$). Based on these results, we concluded that both simple and complex tasks were appropriate to be used in this research.

2.4 Data analysis method

2.4.1 Eye tracking

First, we defined area of interest (AOI) to extract eye tracking metrics in specific regions. Similar to previous study by Wang et al. (2014), AOI was specified to be the entire page to observe the interface design complexity of a full e-learning page. AOI was located on each home page and course page accessed by participants which regards to the given task. If a participant accessed a page other than the page related to the given task, this page was considered invalid and was not included in data analysis.

This study used split-plot (mixed) ANOVA to analyse eye tracking data. The independent variables were design complexity and task complexity. The dependent variables were number of fixation on AOI, fixation duration on AOI, and time on task of users when interacting with e-learning environment (Perkhofer and Lehner, 2019; Katona, 2022). The results included descriptive statistics, tests of within-subjects and between-subjects (ANOVA), and the marginal means of each dependent variable.

2.4.2 Cognitive load questionnaire

The cognitive load questionnaire consists of four statements with seven-scale score adapted from previous studies (Ayres, 2006; Cierniak et al., 2009). Score 1 implies ‘strongly disagree’ while score 7 suggests ‘strongly agree’. The greater the score given to each statement, the more participants agreed with the statement. The scores of four questions were then summed to a cognitive load score. The greater the score indicated the greater cognitive load felt by participants. Questions list of the cognitive load questionnaire was presented in Appendix A.

2.4.3 SUS questionnaire

SUS aims to measure the usability of three e-learning designs with different levels of complexity (Sauro, 2011). SUS consists of ten statements with a five Likert scale. Score 1 refers to ‘strongly disagree’, score 2 refers to ‘disagree’, score 3 is ‘neutral’, score 4 is ‘agree’, and score 5 is ‘strongly agree’. To calculate a participant’s SUS score, the score of the odd numbered statement is reduced by 1. On the contrary, in the even numbered statement, the calculation is 5 minus the score. The final score for each statement is summed together and multiplied by 2.5. SUS scores range from 0 to 100 with average value of 68. Questions list of the SUS was presented in Appendix B.

2.4.4 User experience questionnaire

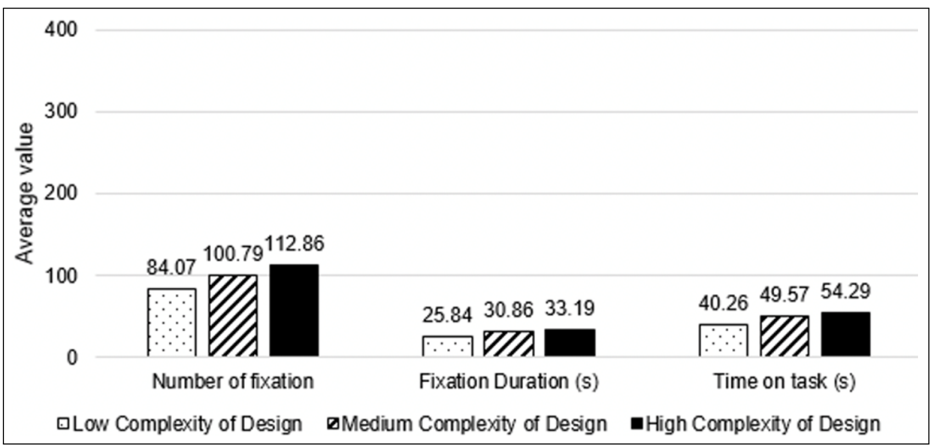
UEQ data analysis was performed using the UEQ data analysis template in Microsoft Excel format (Laugwitz et al., 2008). The UEQ covers a comprehensive impression of user experience in terms of attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty. There are 27 questions with a range of scale from 1 to 7. The score of each question is subtracted with a value of 4 in order to obtain scores ranging from –3 to +3. An average score between –0.8 to +0.8 implies a neutral impression. An average score of less than –0.8 indicates a negative impression. Finally, an average score of more than +0.8 implies a positive impression. Questions list of the UEQ was presented in Appendix C.

3 Results

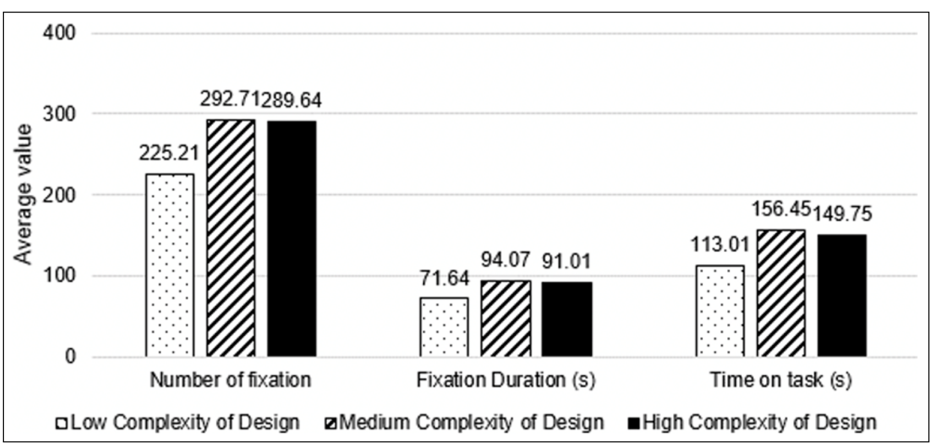
3.1 Eye tracking

This study extracted three metrics of eye tracking to be analysed: the number of fixation, fixation duration, and time on task. A more number of fixation, a more fixation duration, or a longer time for task completion implies a higher cognitive load (Perkhofer and Lehner, 2019; Katona, 2022). Figure 3 shows eye tracking results in different levels of the task and design complexities for each eye tracking metric. The values on the figure denote the average values from each group of participants.

Figure 3 Eye tracking metric results in different level of task and design complexities, (a) simple task (b) complex task



(a)



(b)

During simple tasks, shown in Figure 3(a), users of high complexity e-learning design performed more number of fixation and longer duration of fixation as well as required more time to complete the task. These results imply that the users experienced the highest cognitive load (Wang et al., 2014) compared with both medium and low complexity designs. Complex design required a longer completion time as users needed to search the target within many displayed elements and information.

On the other hand, during complex tasks as shown in Figure 3(b), users of medium complexity design performed more number of fixation, longer duration of fixation, and longer duration of task completion compared with both high and low complexity design users. However, it was not because the users of medium design complexity experienced the highest cognitive load. Instead, the users of the complex design experienced a higher cognitive load (Wang et al., 2014). This experience more likely led to cognitive overload causing the users to overlook irrelevant information (Lavie et al., 2004).

A statistical test was performed on the eye tracking results to observe the impact of both task and design complexities. Normality tests were performed to each of the eye tracking metric results beforehand using the Shapiro-Wilk test. The results show that each data were normally distributed. Therefore, a two-way mixed ANOVA was used in this study.

Based on the results of statistical test on the number of fixation data, the number of fixation is significantly affected by both the task complexity ($F(1, 39) = 216.661$, $p < 0.05$) and design complexity ($F(2, 39) = 6.345$, $p < 0.05$). However, there is no significant interaction between task complexity and design complexity ($F(1, 39) = 1.700$, $n.s.$). Based on the post-hoc test, design with high complexity possesses the most number of fixation during simple tasks but not significantly higher than design with medium complexity ($MD = 12.071$, $p = 0.634$) and low complexity ($MD = 28.786$, $p = 0.087$). During complex tasks, design with medium complexity has the most number of fixation and significantly higher than the design with low complexity ($MD = 67.500$, $p = 0.030$), yet not significantly higher than the design with high complexity ($MD = 3.071$, $p = 0.992$).

Based on the statistical test, the duration of fixation is significantly affected by both task complexity ($F(1, 39) = 194.619$, $p < 0.05$) and design complexity ($F(2, 39) = 3.791$, $p < 0.05$). There is also no significant interaction between task complexity and design complexity ($F(1, 39) = 1.666$, $n.s.$). For simple tasks, the post-hoc test shows that the design with high complexity has the longest duration of fixation but not significantly longer than the design with low complexity ($MD = 7.354$, $p = 0.139$) and medium complexity ($MD = 2.336$, $p = 0.811$). Similarly, for complex tasks, design with medium complexity has the longest duration of fixation yet not significantly longer than the design with high complexity ($MD = 3.054$, $p = 0.949$) and low complexity ($MD = 22.427$, $p = 0.073$).

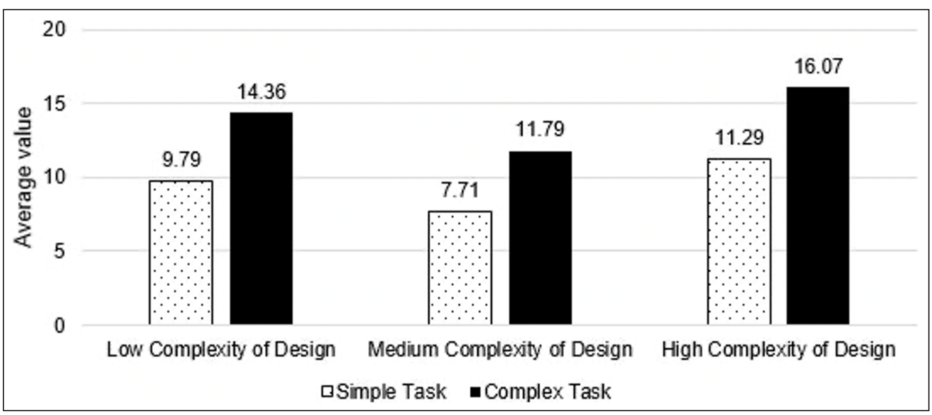
Lastly, the task completion time is also significantly affected by both task complexity ($F(1, 39) = 244.949$, $p < 0.05$) and design complexity ($F(2, 39) = 8.719$, $p < 0.05$). There is also no significant interaction between task complexity and design complexity. The post-hoc test result shows that the design with high complexity required the longest time to complete the simple task and significantly longer than the design with low complexity ($MD = 14.024$, $p = 0.041$) but not significantly longer than the design with medium complexity ($MD = 4.712$, $p = 0.675$). For the complex tasks, the design with medium complexity required the longest time to complete and significantly longer than the design with low complexity (MD = 43.434, p = 0.06) but not significantly lower than the design with high complexity (MD = 6.692, p = 0.869).

The statistical test results show that both task and design complexities significantly affect the number of fixation, fixation duration, and time on task. Therefore, based on an eye tracking approach, it is justifiable to argue that both task and design complexities significantly affect the user's cognitive load.

3.2 Cognitive load questionnaire

While eye tracking was used as a more objective approach to observe a user's cognitive load, the cognitive load questionnaire was used as a subjective approach. The questionnaire was given to each participant in each group according to Table 1. Figure 4 presents the average cognitive load score of each group based on the questionnaire. Design with high complexity contained too much irrelevant information for completing a task. Thus, it generated the highest score of cognitive load. On the other hand, design with medium complexity delivered the lowest cognitive load score despite its more complex design compared with the low complexity design. This was most likely due to medium complexity design had a sufficient number of elements and information according to the user's perception.

Figure 4 Cognitive load questionnaire score in different levels of task and design complexities



A normality test with Shapiro-Wilk test was performed in order to select an appropriate statistical test. The results show that each group data is normally distributed. Therefore, we used a two-way mixed ANOVA to observe the effect of the different levels of task and design complexity to the user's cognitive load. The statistical test result shows that cognitive load scores are significantly affected by both task complexity ($F(1, 39)$ = 68.272, $p < 0.05$) and design complexity ($F(2, 39)$ = 4.649, $p < 0.05$). However, there is no significant interaction between task complexity and design complexity ($F(1, 39)$ = 0.153, $n.s.$). For simple tasks, the post-hoc test result shows that medium complexity design has the lowest cognitive load and is significantly lower than the high complexity design (MD = –3.571, p = 0.007) but not significantly lower than the low complexity design (MD = –2.071, p = 0.159). Similarly, during complex tasks, medium complexity design also has the lowest cognitive load and is significantly lower than the high complexity design (MD = –4.286, p = 0.047) but not significantly lower than the low complexity design (MD = –2.571, p = 0.312).

3.3 System usability scale

In this study, SUS has an objective to measure the usability of each level of the design complexity. Figure 5 presents the SUS score of each level of the task and design complexities. In both simple and complex tasks, the medium complexity design obtained the highest usability score. The users might perceive medium complexity design to have sufficient complexity, neither too simple nor too complex. It was most likely due to users could perform both simple and complex tasks in medium complexity design with ease and without too much unnecessary information. On the contrary, the high complexity design suffered the lowest score for both simple and complex tasks due to excessive elements or information displayed on the design.

Figure 5 SUS score in different levels of task and design complexities

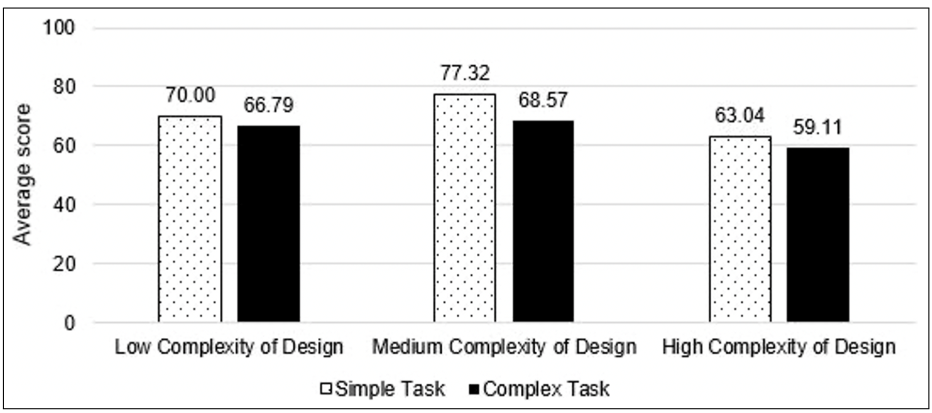


Figure 5 also shows that in each level of design complexity, the SUS score is higher for the simple task. The lower score during complex tasks is presumably due to the complex task required a higher ability of the user to accomplish. Users who lacked the ability would experience more difficulties. Hence, reducing the user's perception of the system usability (Sauro, 2011).

To determine whether each interface design is acceptable, we used the interpretation guideline by Bangor et al. (2009). During simple tasks, the medium complexity design is considered acceptable while the others are marginally acceptable. Meanwhile, during complex tasks, all three designs are marginally acceptable. Marginally acceptable indicates sufficient usability according to the users.

3.4 User experience questionnaire

Figure 6(a) shows the interpretation of the UEQ score for the low complexity design given simple tasks. The perspicuity and efficiency aspects are considered good presumably due to the low complexity of the design lead to easiness for users to quickly find and to understand the information on e-learning. As a consequence, however, the attractiveness and dependability aspects are only above average due to the few elements or information that lead to limited potential interactions. Due to a similar reason, the stimulation and novelty aspects obtained below-average scores.

Figure 6 The UEQ score during simple task in different design complexities, (a) low complexity (b) medium complexity (c) high complexity (see online version for colours)

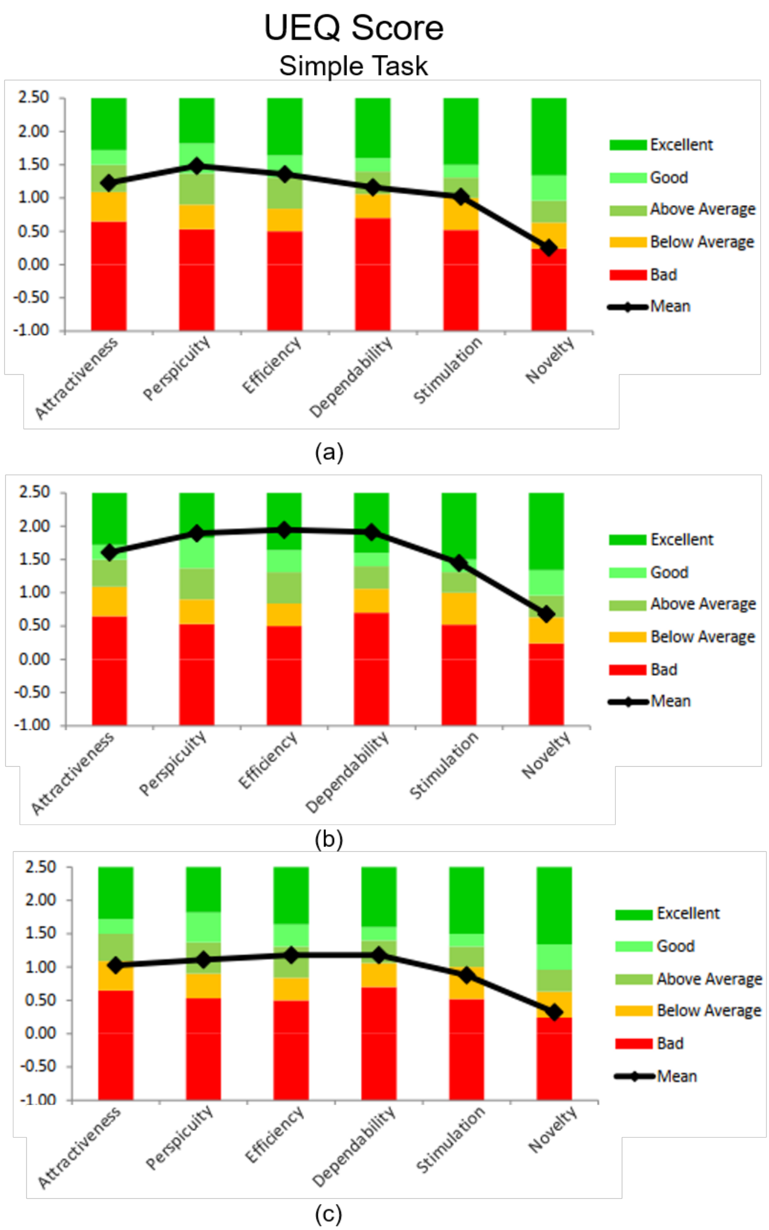
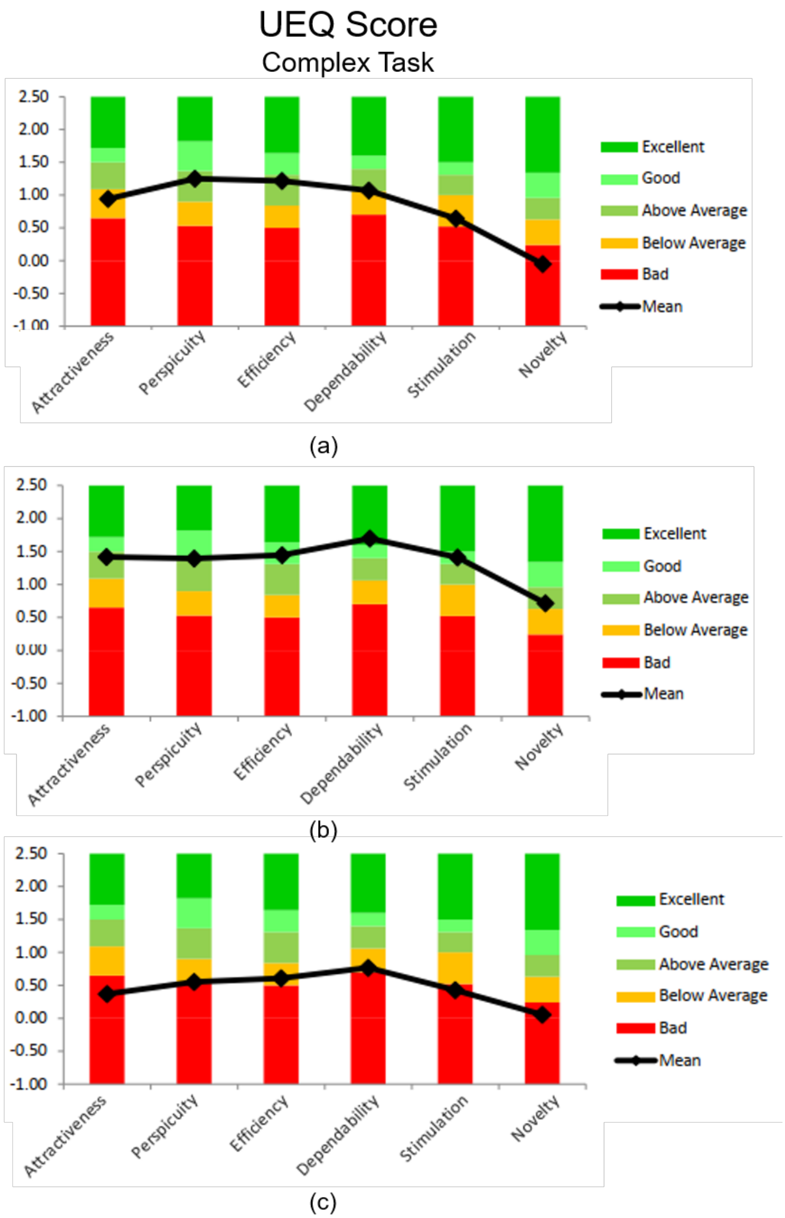
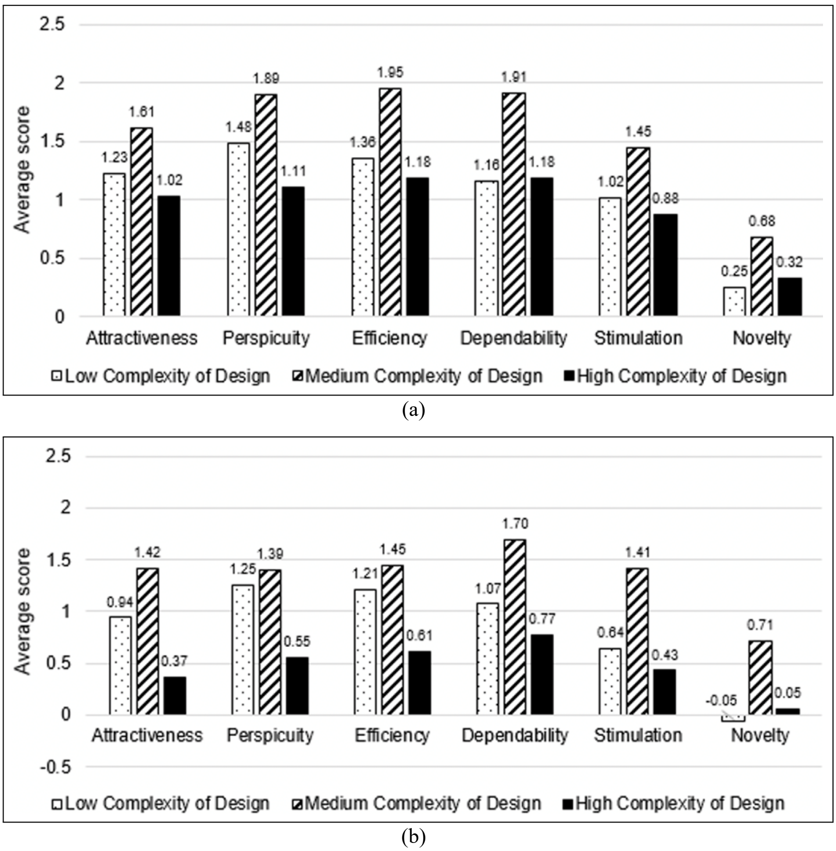


Figure 7 The UEQ score during complex task in different design complexities, (a) low complexity (b) medium complexity (c) high complexity (see online version for colours)



The UEQ score interpretation for medium complexity design – given simple tasks – is shown in Figure 6(b). The perspicuity, efficiency, and dependability aspects are considered excellent due to adequate information on medium complexity design. Therefore, the user could quickly find and understand the user information yet still manage to provide enough interactions. The sufficient number of elements and information on the medium complexity design also leads to good attractiveness as well as provides a quite pleasant and innovative appearance. Hence, above-average scores for stimulation and novelty aspects.

Figure 8 Summary of the UEQ score for each level of task in different design complexities, (a) simple task (b) complex task



The last design for the given simple tasks is the high complexity design with the UEQ score interpretation shown in Figure 6(c). The above-average scores for perspicuity, efficiency, and dependability aspects imply that the users still managed to quickly find and to understand the information as well as had sufficient interaction with the e-learning. However, despite having more elements and information, this design has below average attractiveness, stimulation, and novelty.

Figure 7(a) shows the interpretation of the UEQ score for the low complexity design given complex tasks. The perspicuity, efficiency, and dependability aspects achieved above-average scores which imply that in low complexity design, the users managed to find and to understand the information. In addition, the elements and information were well perceived by the users. However, the users did not find the design interesting and were reluctant to use e-learning in the future. Thus, the attractiveness and stimulation scores are below average. In addition, the users also found the design is ordinary with only a few elements that lead to a bad score of novelty aspect.

For medium complexity design, shown in Figure 7(b), the dependability aspect achieved an excellent score indicating the users were able to perform various interactions with the e-learning. The other five aspects are in either good or above-average scores. The results indicate that medium complexity design is interesting and innovative as well as easy to find and to understand the information displayed in the e-learning.

The worst UEQ score among all the given designs is the high complexity design during complex tasks as shown in Figure 7(c). All of the aspects achieved either below-average or bad scores. The results indicate that during complex tasks, the users had a hard time finding and understanding the information due to the excessive elements and irrelevant information on the e-learning. Hence, inhibiting the users in completing a task. In addition, the users also perceived this design to be less attractive despite provided many elements and information.

The summary of UEQ score for each design complexity during simple tasks is shown in Figure 8(a). In all three designs, all aspects achieved positive evaluation from the users except for the novelty aspect which the users give a neutral evaluation. Similarly, during complex tasks, shown in Figure 8(b), the users also gave a neutral evaluation for the novelty aspect. However, in high complexity design, while the users gave a positive evaluation during simple tasks, during complex tasks the users only gave a neutral evaluation. By comparing Figure 8(a) and 8(b), simple tasks obtained higher UEQ scores in all levels of design complexities. It implies that when given simple tasks, the users earned better experiences during interaction with e-learning. In addition, the best experience was obtained in medium complexity design in both simple and complex tasks.

3.5 Correlation between eye tracking analysis and questionnaires analysis

3.5.1 Simple task

During simple tasks, there are significant positive linear correlations between eye tracking results. The number of fixation result has linear and significant correlation with both fixation duration result (r = 0.895; p = 0.0001) and time on task (r = 0.961; p = 0.0001). However, there is no significant correlation between cognitive load questionnaire results and each eye tracking results: number of fixation (r = 0.177; p = 0.263), fixation duration (r = 0.208; p = 0.186), and time on task (r = 0.250; p = 0.110). SUS results also have no significant correlation with the number of fixation (r = –0.304; p = 0.050), fixation duration (r = –0.252; p = 0.107), and time on task (r = –0.327; p = 0.035). Similarly, each of eye tracking result also has no significant correlation with each of UEQ aspect.

3.5.2 Complex task

Similar to the correlation results during simple tasks, there are significant positive linear correlations between eye tracking results during complex tasks. The number of fixation result has linear and significant correlation with both fixation duration result (r = 0.950; p = 0.0001) and time on task (r = 0.938; p = 0.0001). However, there is no significant correlation between cognitive load questionnaire results and each eye tracking results:

number of fixation (r = 0.132; p = 0.406), fixation duration (r = 0.042; p = 0.792), and time on task (r = 0.014; p = 0.931). SUS results also have no significant correlation with the number of fixation (r = –0.226; p = 0.149), fixation duration (r = –0.155; p = 0.326), and time on task (r = –0.143; p = 0.365). Similarly, each of eye tracking result also has no significant correlation with each of UEQ aspect.

4 Discussion

E-learning design that requires a lower cognitive load to process will be desirable to be developed. Our study shows that cognitive load can be observed using eye tracking metrics (Figure 3). During both simple and complex tasks, design with low complexity needs the lowest cognitive load to process. However, based on cognitive load questionnaire results (Figure 4), medium complexity design requires the lowest cognitive load. To specify a recommended design, the SUS and UEQ results are also taken into consideration to observe the users' perception of the given designs.

As shown in Figure 5, medium complexity design of e-learning – both in simple and complex task – achieved the highest SUS score indicating that it has the highest usability among others. Similarly, the UEQ results shown in Figure 8 also conclude that medium complexity design of e-learning obtained the highest UEQ score in both simple and complex tasks. Although the eye tracking results favour the low complexity design due to its lowest cognitive load, the design is too simple and ordinary that can possibly cause users to have less involvement in the learning activities (Lambert et al., 2009). Therefore, we recommend an e-learning system with medium complexity design to achieve minimum cognitive burden during online learning in the era of COVID-19 pandemic.

Despite the upper hand of using both objective and subjective approaches in analysing the most appropriate e-learning design based on cognitive load, this study also holds some limitations. We only provide two types of e-learning pages; a home page and a course page. There are only two levels of the task of complexities and three levels of design complexities. More levels of a task or design complexities might also need to be observed in the future. There are also some uncontrolled environment conditions during experiments such as light intensity and room temperature. We have yet observed how those conditions affect the experimental results. In addition, we also have not explored the difference between novice and expert users of e-learning.

Our research provides empirical proofs on how quantity of hyperlinks, pictures, and blocks affected cognitive load during learning activities. In fact, pictures provides effective aid to learners when they are spatially combined with text (Castro-Alonso et al., 2021). Ramadiani et al. (2019) suggested that e-learning materials are important factors of learning effectiveness in Indonesia. Our research however, did not measure how various combinations of pictures and text affect cognitive load and learning effectiveness. In future, our research can be extended by investigating impact of pictures-text combination towards cognitive load, learning effectiveness, and relationship between both of them.

5 Conclusions

E-learning systems are inevitably used to support distance education during COVID-19 pandemic. Unfortunately, little attention has been paid on relationship between design complexity of an e-learning system, task complexity, and users' cognitive load. Here we present a novel study to get a better understanding of how e-learning design complexity and task complexity affect users' cognitive load. We proposed the use of four research instruments: eye tracking, cognitive load questionnaire, SUS questionnaire, and UEQ. Three e-learning designs were developed by manipulating different levels of complexities. Beside design complexity, tasks with different levels of complexities – simple and complex tasks – were also taken into consideration as moderating factors. Experimental results show that both task complexity and design complexity significantly affected the number of fixation, fixation duration, and time on task; thus, significantly affected the user's cognitive load. Based on the cognitive load results from the eye tracking and questionnaire approach as well as the participants' perception gathered from the SUS and UEQ, our study recommends a medium complexity design of e-learning to be used in distance learning during COVID-19 pandemic. In the future, a better understanding of cognitive load between novice and expert e-learning users may be explored by relying on three eye tracking metrics used in this research.

References

Al-Rahmi, W.M., Yahaya, N., Aldraiweesh, A.A., Alamri, M.M., Aljarboa, N.A., Alturki, U. and Aljeraiwi, A.A. (2019) 'Integrating technology acceptance model with innovation diffusion theory: An empirical investigation on students' intention to use e-learning systems', *IEEE Access*, Vol. 7, pp.26797–26809, DOI: 10.1109/ACCESS.2019.2899368.

Ayres, P. (2006) 'Using subjective measures to detect variations of intrinsic cognitive load within problems', *Learning and Instruction*, October, Vol. 16, No. 5, pp.389–400.

Bangor, A., Kortum, P. and Miller, J. (2009) 'Determining what individual SUS scores mean: adding an adjective rating scale', *Journal of Usability Studies*, May, Vol. 4, No. 3, pp.114–123.

Berlyne, D.E. (1960) *Conflict, Arousal, and Curiosity*, McGraw-Hill Series in Psychology, McGraw-Hill Book Company, New York, NY, USA.

Castro-Alonso, J.C., de Koning, B.B., Fiorella, L. and Paas, F. (2021) 'Five strategies for optimizing instructional materials: instructor- and learner-managed cognitive load', *Educational Psychology Review*, Vol. 33, No. 4, pp.1379–1407.

Chang, V. (2016) 'Review and discussion: e-learning for academia and industry', *International Journal of Information Management*, Vol. 36, No. 3, pp.476–485.

Cierniak, G., Scheiter, K. and Gerjets, P. (2009) 'Explaining the split-attention effect: is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load?', *Computers in Human Behavior*, Vol. 25, No. 2, pp.315–324.

Djamasbi, S., Siegel, M. and Tullis, T. (2010) 'Generation Y, web design, and eye tracking', *International Journal of Human-Computer Studies*, Vol. 68, No. 5, pp.307–323.

Furnham, A. and Henderson, M. (1982) 'The good, the bad and the mad: response bias in self-report measures', *Personality and Individual Differences*, Vol. 3, No. 3, pp.311–320.

Geissler, G., Zinkhan, G. and Watson, R.T. (2001) 'Web home page complexity and communication effectiveness', *Journal of the Association for Information Systems*, Vol. 2, No. 1, p.2.

Islam, A.N. (2016) 'E-learning system use and its outcomes: moderating role of perceived compatibility', *Telematics and Informatics*, Vol. 33, No. 1, pp.48–55.

Kalyuga, S. (2007) ‘Enhancing instructional efficiency of interactive e-learning environments: a cognitive load perspective’, *Educational Psychology Review*, Vol. 19, No. 3, pp.387–399.

Katona, J. (2022) ‘Measuring cognition load using eye-tracking parameters based on algorithm description tools’, *Sensors*, Vol. 22, No. 3, p.912.

Lai, M.L., Tsai, M.J., Yang, F.Y., Hsu, C.Y., Liu, T.C., Lee, S.W.Y., Lee, M.H., Chiou, G.L., Liang, J.C. and Tsai, C.C. (2013) ‘A review of using eye-tracking technology in exploring learning from 2000 to 2012’, *Educational Research Review*, Vol. 10, pp.90–115.

Lambert, J., Kalyuga, S. and Capan, L.A. (2009) ‘Student perceptions and cognitive load: what can they tell us about e-learning Web 2.0 course design?’, *E-Learning and Digital Media*, Vol. 6, No. 2, pp.150–163.

Laugwitz, B., Held, T. and Schrepp, M. (2008) ‘Construction and evaluation of a user experience questionnaire’, in Holzinger, A. (Ed.): *HCI and Usability for Education and Work, USAB 2008, Lecture Notes in Computer Science*, pp.63–76, Springer, Berlin, Heidelberg.

Lavie, N., Hirst, A., De Fockert, J.W. and Viding, E. (2004) ‘Load theory of selective attention and cognitive control’, *Journal of Experimental Psychology: General*, Vol. 133, No. 3, p.339.

Leuthold, S., Schmutz, P., Bargas-Avila, J.A., Tuch, A.N. and Opwis, K. (2011) ‘Vertical versus dynamic menus on the world wide web: eye tracking study measuring the influence of menu design and task complexity on user performance and subjective preference’, *Computers in Human Behavior*, Vol. 27, No. 1, pp.459–472.

Nugroho, H.A., Listyalina, L., Setiawan, N.A., Wibirama, S. and Dharmawan, D.A. (2015) ‘Automated segmentation of optic disc area using mathematical morphology and active contour’, in *2015 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*, pp.18–22.

O’Brien, H.L., Arguello, J. and Capra, R. (2020) ‘An empirical study of interest, task complexity, and search behaviour on user engagement’, *Information Processing & Management*, Vol. 57, No. 3, p.102226.

Park, T.J., Whang, J., Watts, S. and Han, D.G. (2022) ‘Key success factors in the continuous use of MOOC education in South Korea’, *International Journal of Innovation and Learning*, Vol. 31, No. 2, pp.137–165.

Perkhofer, L. and Lehner, O. (2019) ‘Using gaze behavior to measure cognitive load’, in Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.M. and Randolph, A.B. (Eds.): *Information Systems and Neuroscience*, pp.73–83, Springer International Publishing, Cham.

Rahiem, M.D.H. (2020) ‘The emergency remote learning experience of university students in Indonesia amidst the COVID-19 crisis’, *International Journal of Learning, Teaching and Educational Research*, Vol. 19, No. 6, pp.1–26.

Ramadiani, R., Azainil, A., Frisca, F., Hidayanto, A.N. and Herkules, H. (2019) ‘An integrated model of e-learning continuance intention in Indonesia’, *International Journal of Innovation and Learning*, Vol. 26, No. 1, pp.1–26.

Ramakrisnan, P., Jaafar, A., Razak, F.H.A. and Ramba, D.A. (2012) ‘Evaluation of user interface design for leaning management system (LMS): investigating student’s eye tracking pattern and experiences’, *Procedia-Social and Behavioral Sciences*, Vol. 67, pp.527–537.

Rosch, J.L. and Vogel-Walcutt, J.J. (2013) ‘A review of eye-tracking applications as tools for training’, *Cognition, Technology & Work*, Vol. 15, No. 3, pp.313–327.

Satriya, T., Wibirama, S. and Ardiyanto, I. (2016) ‘Robust pupil tracking algorithm based on ellipse fitting’, in *2016 International Symposium on Electronics and Smart Devices*, pp.253–257.

Sauro, J. (2011) ‘Sustisfied? Little-known system usability scale facts’, *UX Magazine*, Vol. 10, pp.2011–2013.

Schneider, S., Nebel, S., Beege, M. and Rey, G.D. (2018) ‘The autonomy-enhancing effects of choice on cognitive load, motivation and learning with digital media’, *Learning and Instruction*, Vol. 58, pp.161–172.

Setiawan, M.T., Wibirama, S. and Setiawan, N.A. (2018) 'Robust pupil localization algorithm based on circular hough transform for extreme pupil occlusion', in *2018 4th International Conference on Science and Technology (ICST)*, pp.1–5.

Seufert, T. (2018) 'The interplay between self-regulation in learning and cognitive load', *Educational Research Review*, Vol. 24, pp.116–129.

Siron, Y., Wibowo, A. and Narmaditya, B.S. (2020) 'Factors affecting the adoption of e-learning in Indonesia: lesson from COVID-19', *Journal of Technology and Science Education*, Vol. 10, No. 2, pp.282–295.

Sukendro, S., Habibi, A., Khaeruddin, K., Indrayana, B., Syahruddin, S., Makadada, F.A. and Hakim, H. (2020) 'Using an extended technology acceptance model to understand students use of e-learning during COVID-19: Indonesian sport science education context', *Heliyon*, Vol. 6, No. 11, p.e05410.

UNESCO (2020) *COVID-19 Educational Disruption and Response* [online] https://en.unesco.org/news/covid-19-educational-disruption-and-response (accessed 8 April 2021).

UNESCO (2021) *Education: From Disruption to Recovery* [online] https://en.unesco.org/covid19/educationresponse (accessed 8 April 2021).

Van Merriënboer, J.J.G. and Sweller, J. (2005) 'Cognitive load theory and complex learning: recent developments and future directions', *Educational Psychology Review*, Vol. 17, No. 2, pp.147–177.

Wang, H.F. and Lin, C.H. (2019) 'An investigation into visual complexity and aesthetic preference to facilitate the creation of more appropriate learning analytics systems for children', *Computers in Human Behavior*, Vol. 92, pp.706–715.

Wang, Q., Yang, S., Liu, M., Cao, Z. and Ma, Q. (2014) 'An eye-tracking study of website complexity from cognitive load perspective', *Decision Support Systems*, Vol. 62, pp.1–10.

Wang, R., Lowe, R., Newton, S. and Kocaturk, T. (2020) 'Task complexity and learning styles in situated virtual learning environments for construction higher education', *Automation in Construction*, Vol. 113, p.103148.

WHO Regional Office for Europe (2020) *WHO announces COVID-19 Outbreak A Pandemic* [online] https://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/news/news/2020/3/who-announces-covid-19-outbreak-a-pandemic (accessed 8 April 2021).

Wibirama, S., Nugroho, H.A. and Hamamoto, K. (2018) 'Depth gaze and ECG based frequency dynamics during motion sickness in stereoscopic 3D movie', *Entertainment Computing*, Vol. 26, pp.117–127.

Wibirama, S., Santosa, P.I., Widyarani, P., Brilianto, N. and Hafidh, W. (2020) 'Physical discomfort and eye movements during arbitrary and optical flow-like motions in stereo 3D contents', *Virtual Reality*, Vol. 24, No. 1, pp.3433–3451.

Wijayanto, T., Wibirama, S., Maryoto, Z.Z., Winadi, M.N. and Bait, M. (2016) 'Effects of morning-night differences and sleep deprivation on situation awareness and driving performance', in *2016 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pp.267–271.

Wood, R.E. (1986) 'Task complexity: definition of the construct', *Organizational Behavior and Human Decision Processes*, Vol. 37, No. 1, pp.60–82.

Yang, S. (2018) 'Construction research on index system of teaching quality of distance education', *Journal of Discrete Mathematical Sciences and Cryptography*, Vol. 21, No. 6, pp.1431–1436.

Yanson, R. and Johnson, R.D. (2016) 'An empirical examination of e-learning design: the role of trainee socialization and complexity in short term training', *Computers & Education*, Vol. 101, pp.43–54.

Yulianandra, P.V. (2018) *Impact of Interface and Activity Complexity on an E-Learning Website Towards Cognitif Load Based on Eye Tracking*, Master's thesis, Universitas Gadjah Mada.

Appendix A

Cognitive load questionnaire

Instruction: for each statement below, circle an answer (1–7) that you think is the most appropriate. ‘1’ indicates strong disagreement, whereas ‘7’ suggests strong agreement.

1	I needed a high mental effort to finish a task on the e-learning page								
	Strongly disagree	1	2	3	4	5	6	7	Strongly agree
2	I think the task that I completed according to the given instructions was difficult								
	Strongly disagree	1	2	3	4	5	6	7	Strongly agree
3	In my opinion, the e-learning page view gave me a hard time in completing a given task								
	Strongly disagree	1	2	3	4	5	6	7	Strongly agree
4	I needed a lot of concentration during a task completion on the e-learning page								
	Strongly disagree	1	2	3	4	5	6	7	Strongly agree

Appendix B

SUS questionnaire

Instruction: for each statement below, circle an answer that you think is the most appropriate.

No.	*Question*	*Answer*				
		Strongly disagree	*Disagree*	*Neutral*	*Agree*	*Strongly agree*
1	I think that I would like to use this system frequently.	1	2	3	4	5
2	I found the system unnecessarily complex.	1	2	3	4	5
3	I thought the system was easy to use.	1	2	3	4	5
4	I think that I would need the support of a technical person to be able to use this system.	1	2	3	4	5
5	I found the various functions in this system were well integrated.	1	2	3	4	5
6	I thought there was too much inconsistency in this system.	1	2	3	4	5
7	I would imagine that most people would learn to use this system very quickly.	1	2	3	4	5
8	I found the system very cumbersome to use.	1	2	3	4	5
9	I felt very confident using the system.	1	2	3	4	5
10	I needed to learn a lot of things before I could get going with this system.	1	2	3	4	5

Appendix C

User experience questionnaire

Instruction: based on your experience in using the e-learning, please provide answers to each of the following statement items by ticking one circle per line.

	1	*2*	*3*	*4*	*5*	*6*	*7*		
Annoying	○	○	○	○	○	○	○	Enjoyable	1
Not understandable	○	○	○	○	○	○	○	Understandable	2
Creative	○	○	○	○	○	○	○	Dull	3
Aasy to learn	○	○	○	○	○	○	○	Difficult to learn	4
Valuable	○	○	○	○	○	○	○	Inferior	5
Boring	○	○	○	○	○	○	○	Exciting	6
Not interesting	○	○	○	○	○	○	○	Interesting	7
Unpredictable	○	○	○	○	○	○	○	Predictable	8
Fast	○	○	○	○	○	○	○	Slow	9
Inventive	○	○	○	○	○	○	○	Conventional	10
Obstructive	○	○	○	○	○	○	○	Supportive	11
Good	○	○	○	○	○	○	○	Bad	12
Complicated	○	○	○	○	○	○	○	Easy	13
Unlikable	○	○	○	○	○	○	○	Pleasing	14
Usual	○	○	○	○	○	○	○	Leading edge	15
Unpleasant	○	○	○	○	○	○	○	Pleasant	16
Secure	○	○	○	○	○	○	○	Not secure	17
Motivating	○	○	○	○	○	○	○	Demotivating	18
Meets expectations	○	○	○	○	○	○	○	Does not meet expectations	19
Inefficient	○	○	○	○	○	○	○	Efficient	20
Clear	○	○	○	○	○	○	○	Confusing	21
Impractical	○	○	○	○	○	○	○	Practical	22
Organized	○	○	○	○	○	○	○	Cluttered	23
Attractive	○	○	○	○	○	○	○	Unattractive	24
Friendly	○	○	○	○	○	○	○	Unfriendly	25
Conservative	○	○	○	○	○	○	○	Innovative	26