[Inner endpaper]

HEC MONTRÉAL

Exploring the roles of interaction frequency and cognitive load on the aesthetics-usability effect

By:

Bowen Li

HEC Montréal

Under the direction of Professor. Hyung-Koo Lee

Master of Science

User Experience

A Thesis submitted in Partial Fulfillment of Requirements for a Master of Science Degree

April 2021

Résumé

Dans la recherche UX, il existe un impact positif de l'esthétique sur l'utilisabilité qui est appelé effet esthétique-utilisabilité. Cependant, il n'est pas encore pleinement examiné que lorsque l'effet d'utilisabilité esthétique est applicable dans des contextes empiriques. L'objectif de cette mémoire est d'examiner les rôles modérateurs de la fréquence d'interaction et de la charge cognitive sur l'effet esthétique-utilisabilité. Nous avons mené une expérience en ligne avec 50 participants qui ont effectué 8 tâches à caractère utilitaire en ligne sur une courte période. Les résultats indiquent: (1) l'esthétique visuelle a influencé à la fois l'utilisabilité perçue et la performance des tâches au début; (2) l'influence de l'esthétique visuelle sur l'utilisabilité perçue et la performance des tâches a diminué après plusieurs premières interactions; (3) et la charge cognitive a modéré la relation entre l'esthétique visuelle et l'utilisabilité perçue de telle sorte que l'impact de l'esthétique visuelle sur l'utilisabilité perçue diminue avec une charge cognitive réduite. Ces résultats aident à comprendre l'applicabilité de l'effet d'utilisabilité esthétique dans des contextes empiriques et fournir des implications pour le développement de produits numériques.

Mots clés : effet esthétique-utilisabilité, conditions aux limites, fréquence d'interaction, charge cognitive, effet modérateur, phase d'orientation

Abstract

In UX research, there exists a positive influence of aesthetics on usability which is referred to as "the aesthetics-usability effect." However, when the aesthetics-usability effect is applicable to different empirical contexts has not been fully examined. The objective of this thesis is to examine the moderating roles of interaction frequency and cognitive load on the aesthetics-usability effect. We conducted a web-based experiment with 50 participants who completed 8 utilitarian-oriented tasks online over a short period of time. The results indicate that: (1) visual aesthetics influenced both perceived usability and task performance at first; (2) the influence of visual aesthetics on perceived usability and task performance diminished after several interactions; (3) and cognitive load moderated the relationship between visual aesthetics and perceived usability such that the impact of visual aesthetics on perceived usability diminished with reduced cognitive load. The findings of this thesis contribute to improving our understanding of the applicability of the aesthetics-usability effect in empirical contexts and provide implications for the development of digital products.

Keywords: aesthetics-usability effect, boundary conditions, interaction frequency, cognitive load, moderating effect, orientation phase

Table of Contents

List of tables

List of figures

List of abbreviations and acronyms

UX: User Experience

UI: User Interface

BC: Boundary Conditions

IA: Information Architecture

VisAWI: Visual Aesthetics of Websites Inventory

VisAWI-S: Visual Aesthetics of Websites Inventory: Short version

EA: Expressive Aesthetics

CA: Classical Aesthetics

DES-R: Display Evaluation Scale

SMEQ: Subjective Mental Effort Questionnaire

SUS: System Usability Scale

LSA: Latent Semantic Analysis

LA: Low Aesthetics

HA: High Aesthetics

EEG: Electroencephalogram

Acknowledgements

I would first like to thank my thesis mentor, Professor. Hyung-Koo Lee. I sincerely appreciated your careful and patient guidance through this long journey, and this invaluable experience will be deeply remembered in my mind. Thank you for your mentorship, my thesis could not be finished without your support.

Secondly, I would like to express my appreciation to the two reviewers, Professor. Sylvain Sénécal and Professor. Constantions K. Coursaris, for giving me insightful comments and suggestions to further improve my thesis.

I would also like to thank my parents, Minghua Dai and Weiming Li for your encouragement and financial support which helped me to finish my master program and thesis through this difficult time.

Chapter 1 Introduction

With the rapid development of information technology, digital products including desktop websites and mobile applications have become indispensable parts of our lives. However, competition in the digital products market is fierce and product development cost is high. As of 2019, there were over 2 million apps available at the Google Play store and 1.83 million at the Apple App store. In addition, the basic development cost of an application with a core set of features is more than $60,000 while complex mobile app development costs start at $ 230,000, based on the most recent surveys by MLSDev, in 2020.

Under these circumstances, it is important for businesses to understand their users' behavior to be able to survive. Prior research has shown that usability has positive impacts on both customer intention to buy (Hasbullah et al., 2016) and customer loyalty (Flavián et al., 2006). Therefore, designing products with high usability should be the top priority in digital product development. Now, enhancing the user experience (UX) through UX testing and UX design has been embraced by product developers to reduce costs and avoid the high risk of failure. Within the context of UX, those approaches rely heavily on the measurement of usability (Hornbæk, 2006). Thus, several studies have examined the factors that influence usability and found that people perceive a beautiful interface as more usable; this is called the aesthetics-usability effect (Norman, 2004).

Several studies have found that users frequently visit an online retail website several times in short sessions before they decide to make a purchase and become loyal customers (Moe, 2003). Also, their perceptions of usability, aesthetics and task performance fluctuate during this period (Lee & Ha, 2019). Based on previous findings, it is necessary to comprehend how aesthetics influences usability differently within this short time period so that designers can create a steadily perceived usable system, which could subsequently lead to better conversion and customer retention for businesses.

From a theoretical perspective, Zettl (2013) proposes that the visual aesthetic, a key element in UX, combines with usability to create an overall visual effect. Therefore, much research has studied the relationship between aesthetics and usability. However, the findings in the literature remain inconsistent; some studies found that aesthetics influences usability (e.g., Kurosu & Kashimura, 1995; Tractinsky et al., 2000), while others found that there is no significant relationship between aesthetics and usability (e.g., Hassenzahl, 2004). A possible reason for this inconsistency is that the boundary conditions (BC) of the aesthetics–usability relation were not taken into consideration (Tractinsky et al., 2000). More specifically, strong temporal conditions were frequently ignored in previous research (Forlizzi & Battarbee, 2004).

Consequently, the following research investigates how the aesthetics-usability effect changes over time. Karapanos et al. (2008) identify three phases of digital product use: orientation, incorporation and identification. Based on this time line, Sonderegger et al. (2012) find that the aesthetics-usability effect exists at the initial orientation phase and wanes after a long interaction (two weeks). That evidence suggests that the aesthetics-usability effect could possibly diminish in a short time period (orientation phase) as well. It is thus worth investigating how the aesthetics-usability effect changes in a short time period to increase accuracy when applying this effect in empirical contexts.

Since investigating changes in the aesthetics-usability effect requires continuous interaction with the system, another BC, cognitive load, should also be taken into consideration. As the amount of working memory resources used/needed to finish/complete the task (Sweller, 2011), cognitive load decreases through continuous interaction (Sénécal et al., 2015). Subsequently, reduced cognitive load leads to increments in both perceived usability (Longo et al., 2012) and task performance (Haier et al., 1992). Thus, we propose that reduced cognitive load weakens the effect of aesthetics on usability.

Therefore, this thesis aims to address the research gap concerning how the aesthetics-usability effect changes in a short time period of interaction with the system to confirm the applicability of this effect in terms of empirical context. The objective of this study is to examine the interaction frequency (how many times tasks have been finished) and cognitive load as contextual variables and to collect multiple observations from our experiment. We mainly address the following three research questions:

- Does the impact of visual aesthetics on usability diminish in a short time period?
- How does interaction frequency moderate the relationship between aesthetics and usability?
- How does cognitive load moderate the relationship between aesthetics and usability?

To address these questions, this thesis applies BC exploration as a/the theoretical tool, thus examining both interaction frequency and cognitive load as moderators in our experiment. Results indicate that visual aesthetics have a positive influence on both perceived usability and task performance at first and gradually diminish after first several interactions; interaction frequency has the moderating effect of weakening the influence of visual aesthetics on both perceived usability and task performance, while reduced cognitive load has a moderating effect that only dilutes the impact of visual aesthetics on perceived usability.

These results provide important theoretical contributions and have useful business implications. As a preliminary study on the changes of the aesthetics-usability effect in the orientation phase, we find that the positive impact of visual aesthetics on both perceived usability and task performance diminishes just after the first few interactions with the system and both interaction frequency and cognitive load have moderating effects on aesthetics-usability effect. For digital product development, aesthetics design could be used together with less complex tasks that only require short time to finish which could make products stand out among fierce competition as higher usability leads to stronger intention to purchase and user retentions.

The remaining content of this thesis is organized as follows: Chapter 2 provides a comprehensive literature review, including a discussion about the roles of aesthetics and usability in UX; a summary of the main studies that do and do not support the findings of the aesthetics-usability effect; a review of studies on changes in the aesthetics-usability effect over time and cognitive load. In Chapter 3, we propose our theoretical model using BC approaches and formulate the hypotheses for this study. Chapter 4 introduces the experimental design in detail including aesthetics manipulation, information architecture (IA) construction, and task design. The results of a pre-test are also briefly described at the end of this chapter. In Chapter 5, we mainly present the results of Pearson's correlation test and mixed model to test hypotheses. Chapter 6 concludes our study with key findings, theoretical contributions, business implications, limitations and future research. Chapter 7 summarizes what has been done in this thesis.

Chapter 2 Literature review

In this chapter, a comprehensive review of the relevant literature is conducted. The literature review is arranged as follows: the concept, definition and role of both usability and visual aesthetics in UX are explored first. Next, prior research is reviewed, both supporting and not supporting a significant aesthetics-usability effect. After summarizing the limitations of previous studies, we review papers regarding changes in the aesthetics-usability effect over time and briefly introduce cognitive load. Finally, we summarize earlier research to provide direction for our experimental design. All reviewed papers were searched in online databases provided by the HEC Montréal library using the following keywords: aesthetics, usability, correlation, cognitive load, human-computer interaction, user experience, over time, boundary conditions.

2.1 The role of aesthetics in UX

Aesthetics, an ancient word that can be traced back to the old Greek word *aesthetikos*, has long been associated with human beings. Traditionally, product designs for furniture, clothes and vehicles were made beautiful by using appealing colors or symmetrical shapes to attract the eye of customers; these decorations work together with functionality. For example, door handles usually are designed with a downward curve, providing a hint about the right direction for opening the door. As with the design of traditional physical goods, the importance of aesthetics cannot be neglected in digital product design given the advent of digital technology. Many researchers have tried to provide a precise and clear definition of aesthetics in digital products; one of the most commonly used in the field of UX, from Hoffmann and Krauss (2004), describes it as the positive sense or emotion of people when appreciating something that has a pleasant, positive appearance. That is to say, aesthetics is a particular approach to what is pleasing to the senses. Different from physical goods, digital products are intangible, meaning that *visual* aesthetics are mostly applied to product design. Therefore, understanding the concept of visual aesthetics has been a topic of study in the context of UX (Coursaris & Kripintris, 2012). Lavie and Tractinsky (2004) conducted four studies and separate aesthetics into two dimensions that measure visual aesthetics: classical aesthetics and expressive aesthetics. Classical aesthetics follows the traditional design rules of visual clarity (Johnson, 1994) in terms of orderly and clear design; expressive aesthetics emphasizes original and creative design to disturb conventional design rules (Nasar, 2016). Moshagen and Thielsch (2013) developed an 18-item VisAWI and then abbreviated it to form the 4-item short version (VisAWI-S) to measure aesthetics. Therefore, we understand the concepts and definitions of aesthetics within the field of UX by reviewing those items mainly describing visual aesthetics as symmetry, clearness, sophistication, etc. (see Table 1 for details).

Table 1. Description of different aesthetics metrics

Items	EA (Lavie & Tractinsky, 2004)	CA (Lavie & Tractinsky, 2004)	VisAWI-S (Moshagen & Thielsch, 2013)
1	Creative design	Aesthetic design	Everything goes together on this website
2	Fascinating design	Pleasant design	The layout is pleasantly varied on this website
3	Use of special effects	Clear design	The color composition is attractive on this website
4	Original design	Clean design	The layout on this website appears professionally designed
5	Sophisticated design	Symmetric design	

With better understanding of the concept of visual aesthetics, prior research focused on understanding the function the function of aesthetics in UX. Hoffmann and Krauss (2004) propose that a visual aesthetic is about effective communication, and that the perceptions of the viewer are manipulated by applying visual aesthetics. The aim of visual aesthetics is to induce the user to unknowingly, unconsciously, and unsuspectingly choose to become involved in the message of a particular website. This is achieved through involving the user in the communication process by using coordinated elements of visual aesthetics that support the intended message. Zettl (2013) concludes that visual aesthetics works together with usability and other UX elements to create an overall visual effect.

2.2 The role of usability in UX

Usability is undoubtably one of the most important components in UX. It is human nature to be reluctant to expend a lot of effort on things that are hard to do, which means usability is a necessary condition for a system to be successful. For example, if users find that a website is hard to use, they will leave; if users cannot find the product they want to purchase, they will leave; even worse, there are seemingly infinite alternatives for users to choose from in the market. These examples all demonstrate that usability should be the top priority in UX development (Hornbæk, 2006).

At the start of the information age, the CPU was regarded as the heart of the informatics system; less attention was paid to user demands. But with the development of computer systems, emphasis changed from the CPU to the users of the system. This is because end-users were often neglected as an important part of design, resulting in frequent failures to meet their demands. Nicholls (1979) advocates that the center of a system should be the user, and that the former orientation of designs focused on the CPU must be reversed. Therefore, the concept of *usability* was formulated to describe how users interact with the system. The earliest definition of usability, from Miller (1971), describes it as "ease of use." Shackel (1984) expanded the definition, stating that usability should include both the user's subjective feelings about the system and objective performance while accessing the system. In this framework, there are four principal components of any user-system context: user, task, tool and environment (Shackel, 2009). Therefore, the evaluation of usability can be based on the following criteria: 1. success rate in meeting the specified ranges of users, tasks and environments; 2. user's judgement about ease of use; 3. effectiveness of human use in terms of performance (Shackel, 2009). Given the above, we summarize the basic elements that should be contained in usability (see Table 2).

Table 2. Basic elements and definitions of usability[1] (adapted from Norman, 2004, ISO 9241 and Shackel, 2009)

Elements	**Definitions**
Effectiveness	How the system helps users achieve specified goals with accuracy and completeness in particular environments
Efficiency	Users can perform tasks quickly with fewer resources and the easiest process
Satisfaction	How pleasant and acceptable it is to users affected by using the system
Learnability	Users can easily learn how to use the system with training
Tolerance	Users can easily recover from an error and accept levels of human cost in terms of negative emotions

In the field of UX, many approaches such as usability testing and user interviews depend critically on the measurement of usability (Hornbæk, 2006) to produce a system which is easy and pleasant to use (Gould & Lewis, 1985). That is to say, we should measure both objective usability and subjective usability. The reason these two dimensions are distinct is that they can lead to different conclusions about system usability in different contexts of use (Hornbæk, 2006). In UX, we care not only about better task performance, but also about the user's perception of the system. For example, when users browse a website for heuristic purposes, a longer session indicates that they are interested in the website, which is not consistent with objective usability concerning efficiency but rather reflects a good perception of usability by users. Therefore, we should pay attention to both objective and subjective usability and carefully use those metrics to draw unbiased conclusions when evaluating the quality of the whole system.

2.3 Aesthetics-usability effect on perceived usability

Prior research tried to establish the connections between UX elements and usability, aesthetics being one of those elements. As mentioned above, visual aesthetics

[1] The elements are summarized from Norman (2004), Shackel (2009) and ISO 9241. Among those elements, effectiveness, efficiency and satisfaction are included in the ISO 9241 definition of usability (see https://www.w3.org/2002/Talks/0104-usabilityprocess/slide3-0.html for detail). Norman (2004) proposes ease of learning, error tolerance and engagement in his definition of usability, and engagement has been conceptualized similarly to satisfaction in ISO 9241. Shackel (2009) proposes attitude in terms of error tolerance and flexibility, which has the same definition as efficiency.

works together with usability to create an overall visual effect for the system, thus many studies have tried to determine the actual relationship between aesthetics and usability. The first study providing evidence of a correlation between aesthetics and usability was conducted by Kurosu and Kashimura (1995). They developed 26 different ATM layouts and recruited 252 participants to rate usability and the beauty of the interface. Their result shows high correlation between apparent beauty and apparent usability – how attractive the interface looks and how easy it is to use.[2]

Tractinsky et al. (2000) developed an experiment with ATM layouts similar to that of Kurosu and Kashimura. They conducted an experiment with 132 participants, manipulating aesthetics and usability levels. Usability has three levels (i.e., low, medium and high) while aesthetics has two levels (i.e., low and high). Participants were asked to rate aesthetics and usability before and after using the interface. Their result reveals a strong aesthetics-usability correlation both before and after using the system. The authors explain this phenomenon as a halo effect, which means interface-aesthetics outshines all other features of the interface and therefore influences users' evaluation of the entire system.

Other studies have explored how design elements influence the aesthetics-usability correlation and optimal design solution. For example, Coursaris and Kripintris (2012) conducted an empirical study on the white space of a website. They manipulated website white space to three levels: 25%, 50% and 75%, and collected 90 subjects from the experiment. Their results show that aesthetics significantly correlates with usability,[3] and that a website with 50% white space is most attractive. Coursaris et al.

[2] It is worth noting that the authors measured both perceived usability and objective usability. Objective usability was named "inherent usability" in their article and was measured by 7 metrics. However, the results show that only one metric is correlated with apparent usability, which means that there is no significant correlation between inherent usability and apparent usability. Therefore, objective usability is not correlated with aesthetics as well.

[3] In this study, the authors divide usability into three dimensions following ISO standards: effectiveness, efficiency and satisfaction. It is worth noting that website attractiveness correlates with satisfaction but not with effectiveness and efficiency. However, according to regression results, satisfaction accounted for the majority of usability, thus visual aesthetics correlates with usability.

(2008) also investigated how color temperature affects the aesthetics-usability correlation.

Norman (2004) names this correlation the "aesthetic-usability effect," which means users tend to perceive attractive products as more usable. People tend to believe that things that look better work better – even if they aren't actually more effective or efficient. This effect has become a very useful guiding principle for designers. However, it has certain issues, such as usability problems that are minor (rather that critical). For example, if consumers cannot find products on an e-commerce website, they will lose patience and choose to leave no matter how beautiful the interface looks (Flavián et al., 2006).

While several studies have investigated the correlation between aesthetics and usability, results have not always been consistent. In a widely cited article by Hassenzahl (2004), which evaluated the usability, hedonic attributes, goodness and beauty of four MP3-player skins, the author did not find any significant correlation between beauty and perceived usability. Instead, satisfaction was dependent on usability while beauty was dependent on how beautiful the product itself looks.[4] Similar results were also found by Hamborg et al. (2014) who propose that visual aesthetics would affect usability perceptions of the screen of a mobile device as well. In their study, they conducted an experiment with 88 participants who completed two mobile website navigating tasks. Two mobile prototypes with low- and high-aesthetics were developed, and usability was manipulated at two levels (i.e., low and high). However, the results indicate that there is no significant correlation between aesthetics and usability and, instead, that *perceived* usability influences aesthetics. Therefore, we can draw the following conclusion: what is usable is beautiful. Table 3 summarizes the main studies, including those that succeed and fail in proving the aesthetics-usability effect.

[4] In this article, the author uses "goodness" as a general term to measure satisfaction. In the conclusion, the author uses "self-oriented" to indicate that the perception of aesthetics is determined by a product's self-beauty.

Table 3. Summary of the main studies on an aesthetics-usability correlation

Source	Interface	Task	Usability Metrics	Usability Manipulation	Aesthetics Metrics	Aesthetics Manipulation	Correlation(r)
Kurosu and Kashimura (1995)	26 different ATM layouts	View	Self-made (1 item 10-point scale)	No	Self-made (1 item 10-point scale)	No	0.589*
Tractinsky et al. (2000)	9 different ATM layouts	View and use	Self-made (1 item 10-point scale)	High and low	Self-made (1 item 10-point scale)	High, medium and low	Pre-use: 0.66* Post-use: 0.71*
van Schaik and Ling (2003)	Websites	Information retrieval	Display Evaluation Scale (DES-R)	Two different link colors	Self-made (1 item)	No	0.49*
Parizotto-Ribeiro et al. (2004)	7 different ATM layouts	View and use	Self-made (1 item 5-point scale)	No	Self-made (1 item 5-point scale)	No	0.76*
Hassenzahl (2004) Study 1	4 MP3 player skins	View	Pragmatic quality (PQ)^{a}	No	Beauty (1 item 7-point scale)	No	0.07
Hassenzahl (2004) Study 2	4 MP3 player skins	Use	Pragmatic quality (PQ)^{a}	No	Beauty (1 item 7-point scale)	No	Pre-use: 0.14 Post-use: 0.08
Lavie and Tractinsky (2004)	Websites	Browse and view	Self-made (1 item 10-point scale)	No	CA and EA	No	EA: 0.40 - 0.46* CA: 0.68 - 0.78*
Chawda et al. (2005)	Search tool	Information search	SUS	No	Self-made (No detail)	No	Pre-use: 0.76* Post-use: 0.71*

De Angeli et al. (2006)	2 Art websites	Information retrieval	Self-made (5 items 5-point scale)	No	CA and EA	Casual and formal style	CA: 0.38 - 0.48* EA: 0.38*
Cyr et al. (2006)	Mobile website	Information retrieval	PEOU (perceived ease-of-use)	No	Self-made (4 items)	No	0.23*
Hartmann et al. (2007)	3 websites	Browse	Self-made (1 item)	No	Self-made (1 item)	No	0.43*
van Schaik and Ling (2009)	Websites	Information retrieval	SMEQ	No	CA and EA	High and low	Pre-use: -0.11(CA) 0.18(EA) Post-use: -0.13(CA) -0.02 (EA)
Katz (2010)	Search engine	Information search	Self-made (1 item 5-point scale)	High, medium and low	Self-made (pre-use: 1 item post-use: 3 items 5-point scale)	High, medium and low	Pre-use: 0.61* Post-use: 0.10
Quinn and Tran (2010)	Mobile phone	Use	SUS	No	Self-made (7 items 7-point scale)	No	0.50 - 0.53*
Sánchez‐Franco and Martín‐Velicia (2011)	2 Websites	Browse	SMEQ	Utilitarian and hedonic services	Self-made (4 items 7-point scale)	No	Test 1: 0.448* Test 2: 0.201*
Lavie et al. (2011) Study 2	In-vehicle navigation map	Track and navigate	Self-made (5 items)	No	Self-made (14 items)	No	0.97*
Lindgaard et al. (2011)	50 Website homepages	View	Self-made (1 item 9-point scale)	No	Self-made (1 item 9-point scale)	From highest to lowest	Test 1: 0.67* Test 2: 0.87*

Coursaris and Kripintris (2012)	Websites	Information search	ISO*	No	6 items partially cited from CA and EA	High, medium and low	0.662^{b}*
Tuch et al. (2012)	Websites	Pick up items	SUS	High and low	CA and EA	High and low	Pre-use: 0.33 Post-use: 0.46
Hamborg et al. (2014)	Mobile prototype	Causal navigation	AttrakDiff2[c]	High and low	AttrakDiff2	High and low	0.06
Zhuang et al. (2016)	Website	Information retrieval	User Engagement Scale (UES)	No	UES	No	0.471*
Oyibo and Vassileva (2017a)	Mobile website	Causal navigation	Self-made (1 item 5-point scale)	High and low	CA	High and low	0.703*
Oyibo and Vassileva (2017b)	Mobile website	Causal navigation	Self-made (1 item 5-point scale)	High and low	CA and EA	High and low	0.745*(CA) -0.02(EA)
Peng et al. (2017)	Website	View garments	Self-made (5 items 7-point scale)	No	Self-made (6 items 7-point scale)	Different product colors	0.34
Thielsch et al. (2019)	10 chairs	View and judge	Self-made (1 item 7-point scale)	Different design	Self-made (1 item 7-point scale)	Different design	-0.178

Note: Not all studies used the same confidence level; *represents significant at its designated confidence level

a: PQ questionnaire contains twenty-one 7-point items

b: This is the mean score of 3 dimensions of usability in ISO standard: efficiency, effectiveness and satisfaction

c: Author used AttrakDiff2 questionnaires; see http://attrakdiff.de/ for detail

2.4 Aesthetics-usability effect on objective usability

The aesthetics-usability effect is also applicable to objective usability, which is reflected in the literature by task performance (Hornbæk, 2006). Many previous studies confirm that perceived higher visual aesthetics have positive impact on task performance. Sonderegger and Sauer (2010) developed two mobile prototypes, one aesthetically appealing and the other aesthetically unappealing. Results from 60 subjects indicate that, when using the aesthetically appealing prototype, there is less task-completion time required as well as fewer clicks and errors compared with the aesthetically unappealing prototype.

However, there exists an opposite view: visual aesthetics has a negative influence on task performance. Products are often aesthetically designed, which increases their attractiveness; for users however, aesthetic design sometimes serves as a distraction from the task (Marcus, 1992). Thus, many argue that, while we should be in favor of widening the scope of UX by adding more elements such as aesthetics and emotion, it is not inevitable that task performance will subsequently deteriorate (De Angeli et al., 2002). The findings of another study, conducted by Sauer and Sonderegger (2009), support this view. In this study, the authors designed three types of mobile prototypes, each with a different fidelity level (paper, computer and fully operational), and manipulated aesthetics to two levels (high and moderate) to measure the effect of prototype fidelity on usability testing. The prototype with moderate aesthetics had less task completion time and fewer interaction commands[5] than the prototype with high aesthetics.

Finally, several studies found no effect of aesthetics on task performance. For example, Thüring and Mahlke (2007) conducted three studies examining the interactions between usability, aesthetics and emotions. They developed two mobile prototypes for desktops and manipulated both usability and aesthetics to two levels: high and low. They designed a set of browsing tasks to record the number of accomplished tasks and the task completion time to investigate a user's behavior tendency. The results indicate no significant difference in

---

[5] In this study, mobile phone prototypes were still using real keyboards. The authors built comprehensive metric "interaction efficiency" by dividing the optimal number of user inputs by the actual number of user inputs to measure this metric. We use interaction commands, which are much easier to understand.

low- and high-aesthetics prototypes in terms of task performance metrics. Therefore, we summarize the main studies regarding the effect of aesthetics on objective usability- task performance (see Table 4).

2.5 Possible reasons for divergence in previous research

The reasons that may cause inconsistency in providing evidence of the aesthetics-usability effect are sophisticated. First, most studies have very different usage scenarios. For example, we mainly *touch* the screen while using a mobile interface and *click* while using a desktop. These touch-based and mouse-based interactions could lead to different subjective feelings and task performance results (Travis & Murano, 2014). Moreover, larger desktop screen size and smaller mobile screen size, each with a different resolution, influence the results as well (Bridgeman et al., 2003; Raptis et al., 2013). Also, some studies designed tasks for utilitarian use; others designed tasks for hedonic use. Task purpose can influence participants' mood (Seo et al., 2016) such that positive emotions lead to better task performance and perception while negative emotions lead to worse task performance and perception.

In addition to those listed above, the most probable reason for divergence in previous research is that the boundary conditions (BC) of the aesthetics–usability effect were often neglected (Tractinsky et al., 2000). Boundary conditions refer to the different empirical contexts in which the aesthetics-usability relation would change differently. In other words, we should not ignore the impact of temporal and spatial conditions (Busse et al., 2017) when applying the aesthetics-usability effect to empirical contexts. Temporal conditions have an especially strong impact, which should be taken into consideration in usability testing (Forlizzi & Battarbee, 2004). That is to say, experiments adopted by most researchers that only use a one-time test are inappropriate, because having participants who are unfamiliar with systems can lead to the high randomness of results (Sonderegger et al., 2012).

Table 4. Summary of the main studies of the aesthetics effect on task performance

Source	Interface	Task	Aesthetics Manipulation	Aesthetics Metrics	Usability Manipulation	Usability Metrics	Results
Ben-Bassat et al. (2006)	Mobile prototypes	Data entering	High and low	Self-made (1 item 7-point scale)	High, medium and low	Number of items entered	(−)
Thüring and Mahlke (2007)	Mobile prototypes	Browse	High and low	Self-made (1 item 7-point scale)	High and low	Number of completed tasks Task completion time	Study 1: / Study 2: /
Hartmann et al. (2007)	3 websites	Browse	No	Self-made (1 item)	No	Task completion time	/
Sauer and Sonderegger (2009)	Mobile prototypes	Type text message and phone number	High and moderate	Self-made (1 item 7-point scale)	Low. medium and high fidelity	Task completion time Interaction efficiency[a]	(−)
Moshagen et al. (2009)	Website	Answer questions	High and low	Self-made (1 item 7-point scale)	High and low	Task completion time Number of errors	(+) /
Schmidt et al. (2009)	Websites	Read articles	Appealing and diverse	Self-made (1 item 10-point scale)	Image and text balance	Interaction speed	/
Sonderegger and Sauer (2010)	Mobile prototypes	Type text message and phone number	Appealing and unappealing	Self-made (1 item 7-point scale)	No	Task completion time Interaction efficiency[b] Number of errors	(+)

Katz (2010)	Search engine	Information search	High, medium and low	Self-made (1 item 5-point scale)	High, medium and low	Number of successful answers Task completion time Number of search iterations Number of links clicked	/
Quinn and Tran (2010)	Mobile phone	Use	No	Self-made (7 items 7-point scale)	No	Task success rate Task completion time	(+)
Miller (2011)	Virtual learning software	Record videos Sign descriptions Name photos	High and low	CA and EA	High and low	Fluency, expression, linguistics and accuracy[c]	(+)
Lavie et al. (2011)	In-vehicle navigation map	Track and navigate	No	Self-made (5 items)	No	Task completion time	(+)
						Accuracy	/
Tuch et al. (2012)	Websites	Pick up items	High and low	CA and EA	High and low	Task completion time Task success rate Number of clicks[d]	/
Phillips et al. (2014)	Pictures	Decision making with visual-aids	High and low	Self-made (5 items 7-point scale)	No	Response time Answer accuracy	(+)
Douneva et al. (2015)	Chat screen	Virtual team diagnose	Attractive and less attractive	Self-made (4 items 7-point scale)	No	Number of diagnosed patients	/
Reppa and McDougall (2015)	Icons	Search and localize	Appealing and unappealing	Self-made (1 items 5-point scale)	Simple and complex	Task completion time Task errors	(+)

Thielsch et al. (2019)	Website	Search, learn and answer questions	High and low	VisAWI-S	No	Response time Answer accuracy	/
Reppa et al. (2021)	Desktop software	Listen to music and rate	Appealing and unappealing	No	Complex and simple	Response time	Study 1：(+) Study 2：(+)

Note: (-) represents aesthetics leading to task performance decrement; / represents aesthetics having no impact on task performance; (+) represents aesthetics leading to task performance increment.

a: Interaction efficiency is calculated by dividing the optimal number of user inputs by the actual number of user inputs.

b: Interaction efficiency is calculated by dividing the optimal click number by the actual number of clicks.

c: Authors built a comprehensive metric, “task performance,” ranging from 0-100 with these 4 measures.

d: Authors extracted an objective usability score from these 4 measures by using exploratory factor analysis for the correlation test.

2.6 Changes in the aesthetics-usability effect over time

Few previous studies collected multiple observations of the aesthetics-usability correlation. Tractinsky et al. (2000) observed in their study that usability perceptions decrease in the high aesthetics group while they increase in medium and low aesthetics groups by measuring pre-use and post-use contexts. Several studies that measured twice also have shown the dip in correlation. In their study, van Schaik and Ling (2009) conducted two experiments to explore how contexts of use change users' perceptions of aesthetics over time; usability and aesthetics were measured both before and after use. The results indicate that the aesthetics-usability correlation weakened after use.[6]

These findings provide evidence that users' perceptions of digital products change dynamically during product use. Karapanos et al. (2008) propose three phases of product use (orientation, incorporation and identification) based on their four-week usability testing experiment. The first phase, orientation, refers to a user's first experience with the product. In this phase, aesthetics plays an important role as it has the most influence on the user's first impression. Another important quality is learnability, which is one of the dimensions of usability. Because the user is not familiar with the product, the product should have good usability; to increase user satisfaction, it should be easy to learn to use. The second phase is incorporation, when the product is integrated into the user's daily life. As a user gets used to the product, both aesthetics and usability become less important than the experience during the orientation phase. Thus, usefulness in terms of how frequently the product can be used in daily life outweighs all other aspects. In the last phase, identification, users want to be differentiated from their friends by having a unique product. As such, social aspects matter most in the user's consideration, and usability still matters as a user spends more time interacting with the product.[7]

---

[6] It is worth noting that the main purpose of this study was to measure how aesthetics perceptions change over time in different contexts of use, so usability was not the main measurement. Also, the authors did not indicate p-value in their study; they only indicated which correlations were significant. The correlation between aesthetics and usability was not significant (see Table 4 for details) but the actual p-value was not indicated. However, the decrease in correlation was revealed in the data.

[7] The authors used stimulation to represent aesthetics and built a regression model with satisfaction as predicted. They reported significant parameters: usability (0.43), stimulation (0.43) in orientation phase; usability (0.19), stimulation (0.22) in incorporation phase; usability (0.44) in identification phase.

Study results from Karapanos et al. (2008) indicate that the effect of usability always remains constant while the effect of aesthetics weakens when assessing user satisfaction during all three product use phases. To examine how the aesthetics-usability effect changes, Sonderegger et al. (2012) conducted a two-week usability-testing experiment covering the entire product use phase. Results show significant positive correlation between aesthetics and perceived usability at first, and the correlation diminishes after two weeks. However, their study finds no significant effect of aesthetics on task performance. In summary, the aesthetics-usability effect wanes over time.

Sonderegger et al. (2012)'s study supports the finding that the aesthetics-usability effect diminishes over a long time period (two weeks). More precisely, the significant impact of aesthetics on perceived usability is only found after two tasks in one-day usability testing. Given this finding, the aesthetics-usability effect could possibly diminish, which is worth investigating by evaluating more tasks during the orientation (first) phase.

2.7 Cognitive load as another BC in the aesthetics-usability effect

From the neurophysiological perspective, the disappearance of the aesthetics-usability effect could also be explained by human brain activities. Cognitive load, which refers to the amount of working memory resources used to finish the task (Sweller, 2011), could be another BC of the aesthetics-usability effect. Satpute and Lieberman (2006) suggest a dual-process model of brain activities: controlled processing that is conscious and requires a lot of mental effort, and automatic processing that is unconscious and requires less mental effort. As people become familiar with the interface though continuous interaction, brain activities move from controlled (new task) to automatic processing (routine task) (Sénécal et al., 2015). Prior research found that cognitive load decrement leads to increments in both perceived usability (Longo et al., 2012) and task performance (Haier et al., 1992). Therefore, when moving from controlled processing to automatic processing, it is possible that the effect of reduced cognitive load dilutes the impact of aesthetics on usability.

2.8 Summary of the literature review

In summary, the/our literature review shows that aesthetics plays an important role in the UX field and the correlation between aesthetics and usability is confirmed by previous research. Few studies have not found a correlation; the probable reason is that BC and contingency were not fully explored (Tractinsky et al., 2000).

Consequently, the strong effect of temporal conditions should be taken into consideration (Forlizzi & Battarbee, 2004) because it may influence results. Sonderegger et al. (2012)'s study provides evidence that the effect of aesthetics on perceived usability is significant and diminishes after a long time period (two weeks). We propose that the time period during which the aesthetics-usability effect diminishes could be further narrowed to the orientation (first) phase. Another BC is cognitive load because reduced cognitive load could possibly dilute the impact of aesthetics on usability.

Given these questions, this thesis is dedicated to finding answers for the research gap regarding how the aesthetics-usability effect changes over a short time period (orientation phase), and how two BCs – interaction frequency and cognitive load – influence the aesthetics-usability effect. In the next chapter, we present our research model and develop our hypotheses.

Chapter 3 Research Model and Hypotheses

In this chapter, we propose our theoretical model using BC approaches and formulate the hypotheses for this study.

3.1 Theoretical approach and research model

In our study, we applied theoretical tools using a BC framework to answer our research questions. The purpose of using a BC approach is to increase the accuracy of applicability of the aesthetics-usability effect in terms of empirical contexts. To account for temporal conditions as contextual variables, we first recorded how many times tasks have been finished as the interaction frequency. As the two variables that were not correlated with visual aesthetics and usability, interaction frequency and cognitive load were amended as moderators. We therefore built our research model (see Fig. 1) based on the BC theoretical approach. In our model, visual aesthetics played the role of the independent variable, which was divided into two dimensions, low and high; usability was the dependent variable, which was measured using both objective and subjective dimensions. The tasks of the experiment were repeated eight times (see Chapter 4 for details), seeking answers to two questions: in which time period does the aesthetics-usability effect diminish, and how do interaction frequency and cognitive load moderate their correlation.

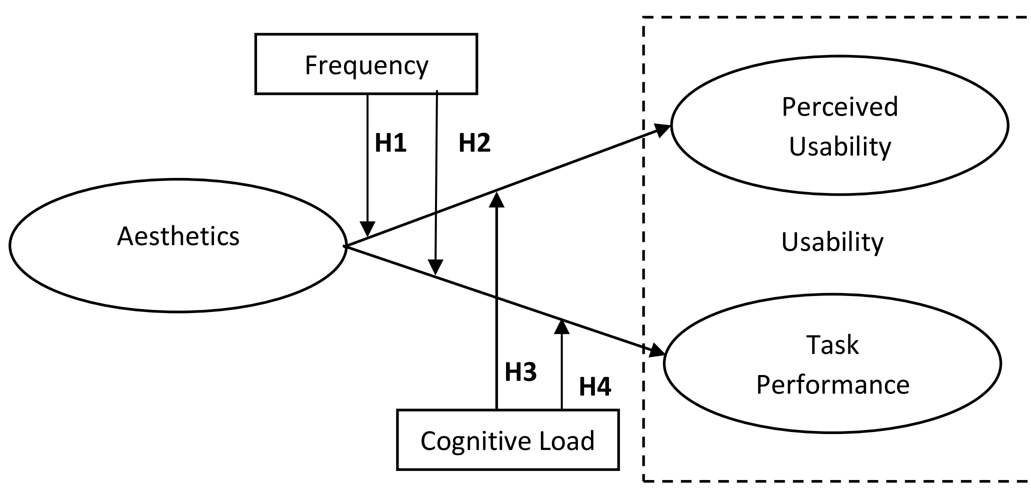

Figure 1. Proposed research model

3.2. The effect of aesthetics on usability over time

Prior research found the positive impact of aesthetics on both perceived aesthetics (e.g., Kurosu & Kashimura, 1995; Tractinsky et al., 2000) and task performance (e.g., Quinn & Tran, 2010; Sonderegger & Sauer, 2010), which is called the aesthetics-usability effect (Norman, 2004). However, several studies found no evidence to support the existence of the aesthetics-usability effect (e.g., Hartmann et al., 2007; Hassenzahl, 2004; van Schaik & Ling, 2009). A possible reason is that temporal conditions were sometimes ignored during usability testing (Forlizzi & Battarbee, 2004). To find reasons for this divergence, Sonderegger et al. (2012) conducted an experiment in usability testing that covered all three product use phases and found a positive correlation between aesthetics and perceived usability during the orientation phase (first day). Thus, we expected visual aesthetics to be initially correlated with usability. The experiment (Sonderegger et al., 2012) also provided evidence that the aesthetics-usability effect gradually diminishes after the orientation phase. Subsequent studies have tried to further investigate changes in the aesthetics-usability effect during short time periods (orientation phase). Limited results indicate that perceived usability in the low aesthetics group gradually increases to the same level as it was in the high aesthetics group, and task performance improves during short time periods (Lee & Ha, 2019). Therefore, we suggest that, by collecting a few more observations (8 times in our study) during this short time period, it is possible to see the aesthetics-usability effect diminish during short time periods. Also, we suggest that the diminished aesthetics-usability effect should be ascribed to interaction frequency, which moderates the correlation between visual aesthetics and usability. As such, we present our first and second hypotheses as follows:

H1: Interaction frequency moderates the effect of aesthetics on perceived usability such that the effect will diminish with more interaction with the interface.

H2: Interaction frequency moderates the effect of aesthetics on task performance such that the effect will diminish with more interaction with the interface.

3.3 The moderating role of cognitive load on the aesthetics-usability effect

From the neurophysiological perspective, we suggest that cognitive load, considered as another BC in our study, also has a moderating effect on the aesthetics-usability effect. Prior research provides evidence that brain activities transform from controlled processing, requiring higher cognitive load, to automatic processing, requiring lower cognitive load, through continuous interaction with the interface (Sénécal et al., 2015). As the brain activities are transformed to the automatic processing, users tend to complete the tasks unconsciously, which means they will pay less attention to the visual aesthetics. As such, the impact of visual aesthetics on both perceived usability and task performance may gradually diminish with reduced cognitive load. Thus, we suggest that reduced cognitive load dilutes the impact of visual aesthetics on usability, which means that cognitive load moderates the correlation between aesthetics and usability. Therefore, we present the third and fourth hypotheses of this study as follows:

H3: Cognitive load moderates the effect of aesthetics on perceived usability such that the effect will diminish with reduced cognitive load.

H4: Cognitive load moderates the effect of aesthetics on task performance such that the effect will diminish with reduced cognitive load.

Chapter 4 Method

To find answers to research questions and test hypotheses, we conducted a controlled experiment with participants recruited from MTurk. In this chapter, we present the details of the experimental design, the development of the experimental prototype, the visual aesthetics manipulation, the IA, and the experiment procedure.

4.1 Experimental design

Our study employed a usability testing experiment that referred to prior research using utilitarian tasks (e.g, Lee & Ha, 2019; Sauer & Sonderegger, 2009; Tuch et al., 2012) to examine the effects of aesthetics on perceived usability and task performance. To examine the moderating effect of interaction frequency and cognitive load, we repeated the task eight times and collected data from post-task questionnaires. This study has a 2 (visual aesthetics: low, high) × 8 (frequency: 1 to 8) mixed factorial design, where aesthetics is a between-subject factor with two levels: high aesthetics and low aesthetics, and interaction frequency is a within-subject factor that reflects the number of times the experimental task was completed.

4.1.1 Participants

Fifty participants were recruited via Amazon Mturk to take part in the experiment (see Appendix A for demographic information). To ensure that participants would be able to understand our instructions and use the website, we recruited subjects who use English as their primary language as well as those without any underlying health conditions. Also, we required participants to have their own computer with a stable and secure Internet connection of at least 5mbps download/upload since the experiment was conducted remotely. To ensure participant performance, US$15 compensation was allocated to participants after obtaining their experiment data. Each participant started the experiment by choosing one of two experimental conditions; then, they could not participate in the other condition.

4.1.2 Prototype

To create a realistic online shopping experience for the experiment, we designed prototypes of an online shopping website, *Watchyourspirit*, mainly selling ethical jewelry products including necklaces and earrings. This website has a homepage, a product catalog page, a product page, an "about us" section, a sign up/sign in page, and basic shopping cart functions. The two different versions (low aesthetics and high aesthetics) of the prototypes were developed by Axure RP 9 software (see fig 3, 4, and 5 for details). The two prototypes were only different in the level of visual aesthetics and were consistent in functionality.

4.2 Construction of information architecture

The purpose of building the IA was to ensure that both prototypes have similar objective usability levels to avoid creating bias about perceptions and influencing task performance while maintaining different levels of aesthetics. First, we built the IA by applying closed card sorting. This method is useful in understanding how users think about the content, thus helping us organize content to suit users' mental models (Wood & Wood, 2008). We recruited 4 UX experts to create 54 cards in total, then 20 participants were recruited to sort the cards. All UX experts and participants did not take part in the following experiment. There was a total of 5 groups that were categorized as the result of card sorting (see Table 5).

Table 5. Results of card sorting

Group 1	**Group 2**	**Group 3**	**Group 4**	**Group 5**
Labels				
Earrings	New In	My account	Conditions of Sale	About Us
Necklaces	Bestsellers	Wish List	Privacy	Social Responsibility
Gifts	Our Picks	Login/Sign up	Returning Items	Community
Collections	Trending Now	Order History	Shipping	FAQs
Accessories	Our Products	User profile		Follow Us
Special Editions				

To ensure that both low- and high-aesthetics prototypes have similar levels of usability, we built the same IAs by choosing several of the sorted cards above. Our experiment task was to find the target items (see section 4.4 for details). To ensure that there was no significant difference in the difficulties of each task, no search function was designed in the prototype and only one navigation path was designed to reach each target item (see Fig. 2 for example).

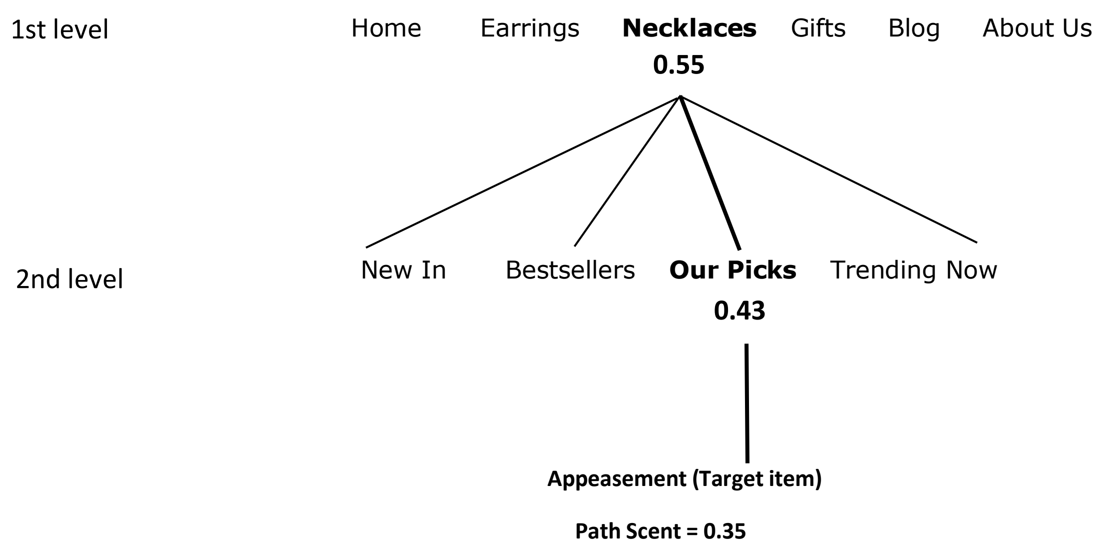

Figure 2. Typical example of target item navigation path

To ensure similar task difficulty of different items, we applied the latent semantic analysis (LSA) approach to calculate the information scents. Information scent refers to the extent to which users can predict what they will find if they follow a certain path through a website (Landauer et al., 1998). Latent semantic analysis is a mathematical method used to compare the similarity of words from different titles, paragraphs and texts; it is often used to build optimal IA to reduce users' cognitive loads. We calculated the information scents by using this application, provided by the University of Colorado Boulder (http://lsa.colorado.edu/). As shown in figure 2, the information scents indicate how much effort users need to reach from one page to another page, the path scent indicates the total effort users need to reach the designated page. The results indicate that all 4 items have similar path scents, ranging from 0.28 to 0.43, meaning that the difficulty of each task is similar.

Table 6. Path scent of each item

	Item 1	Item 2	Item 3	Item 4	Mean (SD)
Path Scent	0.35	0.28	0.30	0.43	0.34 (0.06)

*Path Scent ranges from 0 to1, 0.30-0.50 indicates path scent is medium

4.3 Manipulation of aesthetics

To ensure the success of aesthetics manipulation at two levels (high and how), we first referred to industry benchmarks (e.g., Tiffany & Co. and Bulgari) as templates and applied their appealing visual elements (icons, colors, etc.) to the interface design. Equally important, we referred to previous research results that provided the best design ideas for how to build beautiful websites. Figures 3, 4, and 5 are screenshots that compare our two designed prototypes. Following are the design features of the high-aesthetics website that distinguish it from the one with low-aesthetics:

1. Leave 25% white space in every page (Coursaris & Kripintris, 2012);

2. Apply a cool primary color (e.g., blue or green) and a warm secondary color (e.g., red or orange) as background colors (Coursaris et al., 2008);

3. Use sans-serif font style (Altaboli, 2013) and appropriate font size for each section;

4. Use symmetrical arrangements of product photos (Bauerly & Liu, 2008);

5. Use appealing icons exported from luxury jewelry websites.

Figure 3. Homepage (top: low aesthetics; bottom: high aesthetics)

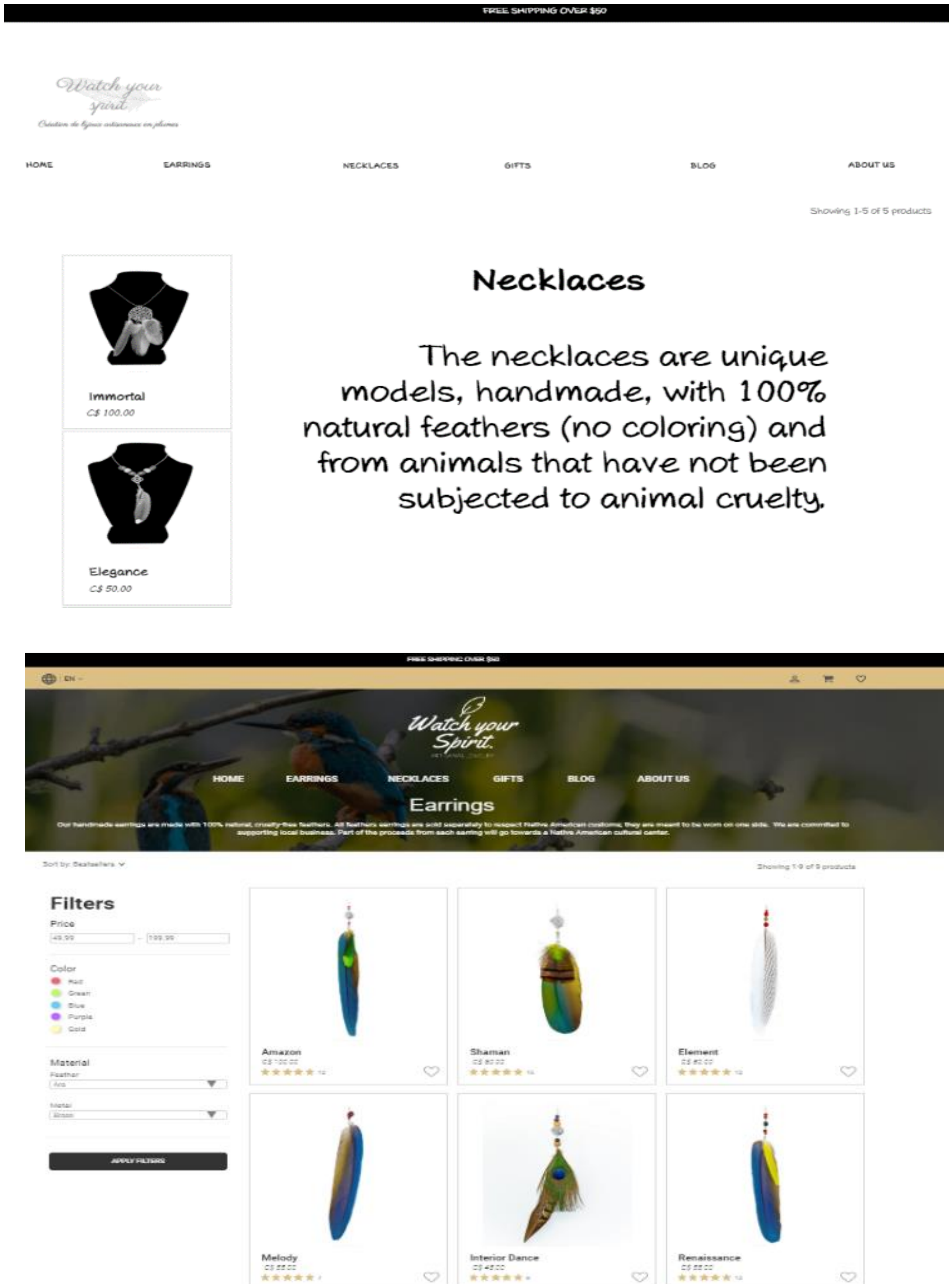

Figure 4. Product Catalog (top: low aesthetics; bottom: high aesthetics)

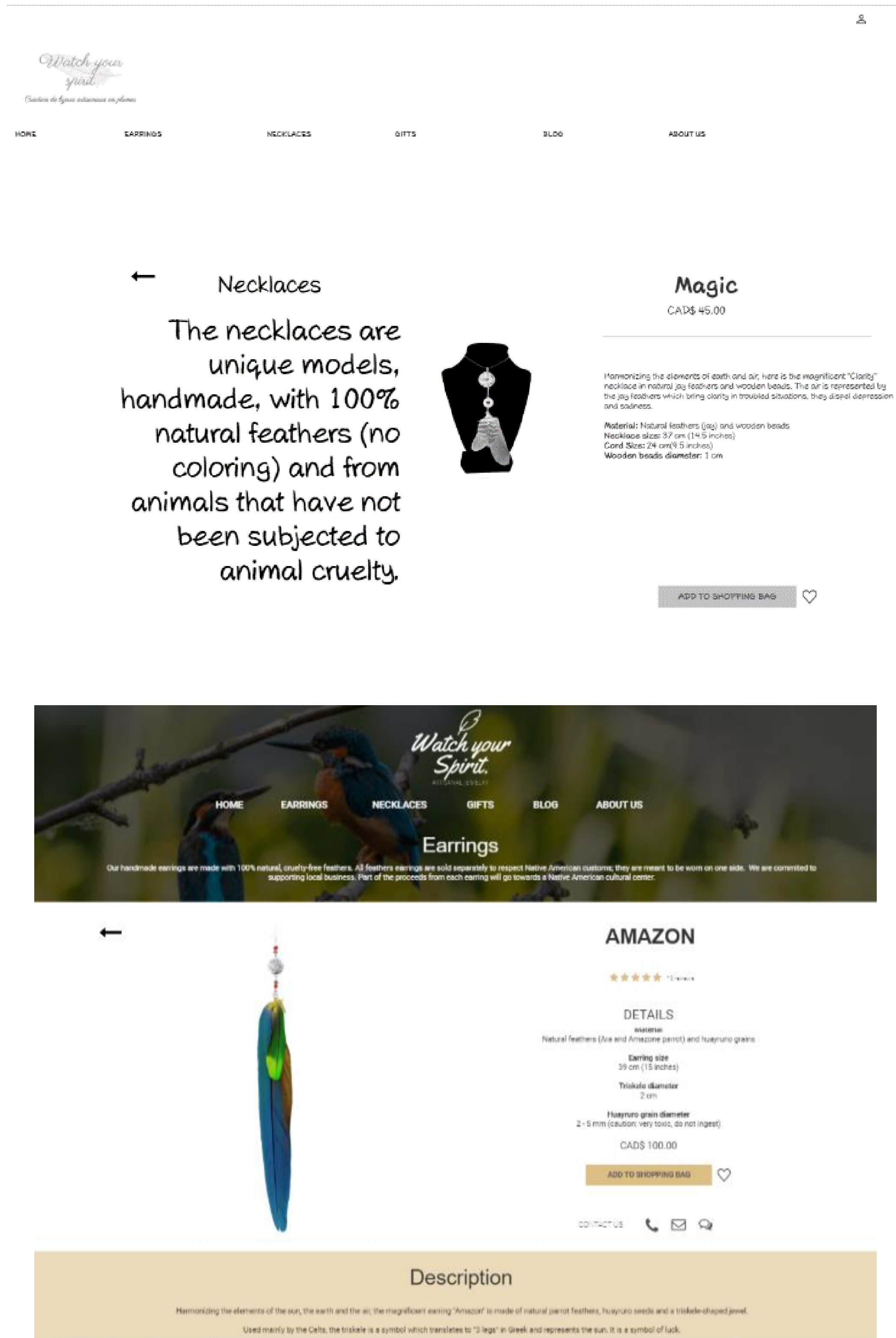

Figure 5. Product Page (top: low aesthetics; bottom: high aesthetics)

4.4 Experimental procedure

Our experiment was conducted remotely on Amazon Mturk. At the preliminary session, an electronic consent form was shown to participants at the Amazon Mturk website to briefly introduce the objectives and process of our study and to inform them about potential physical and psychological risks (see Appendix B for details). After that, participants could decide whether they want to take part in the experiment or not. The estimated time for completing the experiment was about an hour and participants could withdraw at any moment. However, they would receive no compensation if they quit during the experiment.

For the main session, 50 participants were randomly assigned to one of two conditions (25 participants for each prototype). Then, detailed instructions about how to finish the experiment were shown to the participants (see Appendix C for details). The whole procedure of the experiment took average about an hour for a participant to finish. The first step of the experiment was collecting basic demographic information through the pre-questionnaires (see Appendix D). Before the task, participants were shown a photo of the target item. The task was to add the target item to the shopping cart within three minutes. As shown before, the tasks were counterbalanced as the order of target items for each participant was randomized. There was a total of 4 target items that were randomly displayed to each participant (see Fig. 6). If participants found the correct item within the designated time, the prototype jumped to the result page and informed them that they had succeeded in the task; if participants found the wrong item, the prototype jumped to the result page and informed them that the task was a failure. If participants were not able to find the target item in the designated time, the task was also regarded as a failure. We designed a small window to count time; the “give up” button appeared after three minutes and participants could simply click it to end the task. Task completion time was recorded electronically after the task, and participants responded to the items about task success or failure, perceived usability, visual aesthetics, and perceived cognitive load in post-task questionnaires. Then the prototype jumped back to the new task at the start page. The post-task questionnaire process was repeated eight times. An overview of the entire procedure with measurements is shown below (see Fig. 7).

Figure 6. Target items for tasks

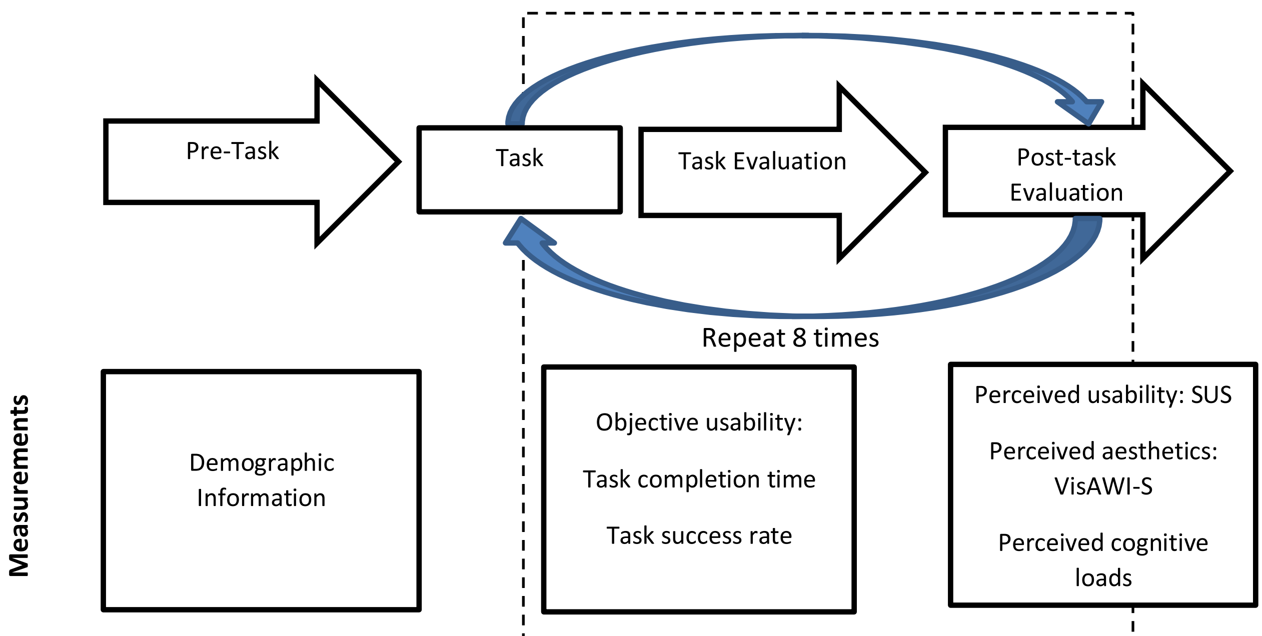

Figure 7. Overview of experimental procedure with measurements

4.5 Measurements

SUS – In this study, participants' perceptions of usability were recorded with this commonly used system usability scale (see Appendix E). In usability testing, SUS is a simple but comprehensive ten-item Likert scale measuring attitude that evaluates subjective system usability with a global view, based on ISO standards for effectiveness, efficiency and satisfaction. This measure has been widely adopted by previous researchers (e.g, Chawda et al., 2005; Grishin, 2018; Quinn & Tran, 2010; Tuch et al., 2012) and can effectively reflect a user's perception of usability from most aspects of the whole system.

Task Completion Time/Task Success Rates – In this study, we collected task completion time and task success rates as the metrics to measure objective usability - task performance. These two metrics were used in prior studies (e.g, Lavie et al., 2011; Quinn & Tran, 2010) as the main metrics for measuring task performance.

VisAWI-S – As tasks were repeatedly performed by participants, obtaining their instinctive and intuitive perceptions was key in determining the results of the study. The VisAWI-S (see Appendix F), a four-question short assessment extracted from the VisAWI, has been shown to be effective and comprehensive in measuring visual aesthetics while keeping assessment times to the minimum (Moshagen & Thielsch, 2013). Therefore, we applied this questionnaire to evaluate our participants' perceptions of aesthetics in this study.

Perceived Cognitive Load – We adapted a self-report questionnaire that was developed by Klepsch et al. (2017) and named "naïve ratings." (see Appendix G) This economical cognitive load measurement has been found to be effective in measuring users' straightforward feelings about their mental efforts when processing interaction tasks.

4.6 Pretest of aesthetics manipulation

To ensure that instructions were given clearly to participants, that there were no obvious usability issues in the prototypes, and that aesthetics was successfully manipulated, a pretest was conducted prior to the official test. There was a total of 10 participants recruited from HEC Montréal for the pretest (see Appendix J for demographic information); data collected from the pretest participants were not used in the main experiment.

The results show that aesthetics manipulations were successful during the pretest. To check for aesthetics manipulation, a t-test was conducted using the VisAWI-S scores. As shown in Table 7 below, the difference in the VisAWI-S scores between the two groups is significant (Cohen's *f* >0.5), which means the effect size of aesthetics manipulation is large and successful.

Table 7. Results for the pretest of the visual aesthetics manipulation

Variable	**Group**	**N**	**Mean**	**SD**	**t**	***P***	**Cohen's *f***
Aesthetics (VisAWI-S)	Low Aesthetics	5	3.6	0.68	-2.68	<.05	0.85
	High Aesthetics	5	5.0	0.95			

Chapter 5 Data Analysis

In this chapter, we examine the descriptive statistics of each variable to assess how they vary with frequency. Then, we conduct Pearson's linear correlation tests time by time to test the hypotheses of whether the aesthetics-usability effect diminishes over more interactions with the website for both perceived and objective usability. Finally, we build mixed models to test the moderating effect of interaction frequency and cognitive load.

5.1 Testing the aesthetics manipulation

Because we used a different sample for the experiment from that of the pretest, we checked whether the aesthetics manipulations were successful again before the main analysis. The result indicates that the effect size of visual aesthetics is large (Cohen's f>0.8).

Table 8 Results for manipulation check of visual aesthetics

Variable	**Group**	**N**	**Mean**	**SD**	**t**	***P***	**Cohen's *f***
Aesthetics	Low Aesthetics	200	3.91	1.23	-1.8	<.001	0.91
	High Aesthetics	200	5.67	0.59			

5.2 Descriptive statistics

All data were checked for normal distribution and linearity before the main analysis (see Appendix H for details). All distribution plots follow a curved bell pattern and Q-Q plots lie on straight diagonal lines. For all statistical tests an alpha level of 0.05 was used. The following two tables (Table 9 and Table 10) summarize the descriptive statistics of the experiment data.

Table 9. Descriptive statistics of the high-aesthetics group

	Perceived Usability		Aesthetics		Task Success Rate		Task Completion Time(S)		Cognitive Loads	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Task 1	66.7	17.03	5.61	0.68	80%	0.41	105	30.71	3.71	1.56
Task 2	69.2	18.21	5.71	0.69	76%	0.44	90.8	36.12	3.63	1.62
Task 3	68.2	18.17	5.65	0.58	88%	0.33	81.56	32.59	3.42	1.57
Task 4	70.3	14.65	5.78	0.60	88%	0.33	72.8	18.97	3.51	1.59
Task 5	69.4	17.58	5.7	0.58	80%	0.51	72.64	31.88	3.27	1.45
Task 6	67.3	15.47	5.69	0.52	96%	0.20	68.84	15.19	3.43	1.53
Task 7	64.7	16.39	5.59	0.59	96%	0.20	64.04	15.91	3.43	1.37
Task 8	65.1	16.72	5.66	0.53	96%	0.20	68.88	17.88	3.23	1.35

Table 10. Descriptive statistics of the low-aesthetics group

	Perceived Usability		Aesthetics		Task Success Rate		Task Completion Time(S)		Cognitive Loads	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Task 1	54.44	20.89	3.74	1.29	72%	0.46	129.16	45.18	3.85	1.65
Task 2	55	21.94	3.65	1.33	64%	0.49	116.16	42.71	3.86	1.39
Task 3	57.96	22.17	3.85	1.38	92%	0.28	99.64	25.61	3.73	1.49
Task 4	61.88	23.70	3.91	1.34	84%	0.37	89.88	28.20	3.68	1.44
Task 5	65.54	23.03	3.98	1.10	76%	0.44	87.4	31.12	3.48	1.36
Task 6	67.06	20.98	4.05	1.17	92%	0.28	79.28	25.90	3.37	1.44
Task 7	69.46	19.81	4.04	1.11	96%	0.2	81.16	24.29	3.38	1.42
Task 8	68.86	21.03	3.93	1.19	100%	0	78.04	27.71	3.41	1.48

To better understand how variables fluctuated with times, we drew line charts and conducted t-tests to examine whether there is significant difference between the two groups.

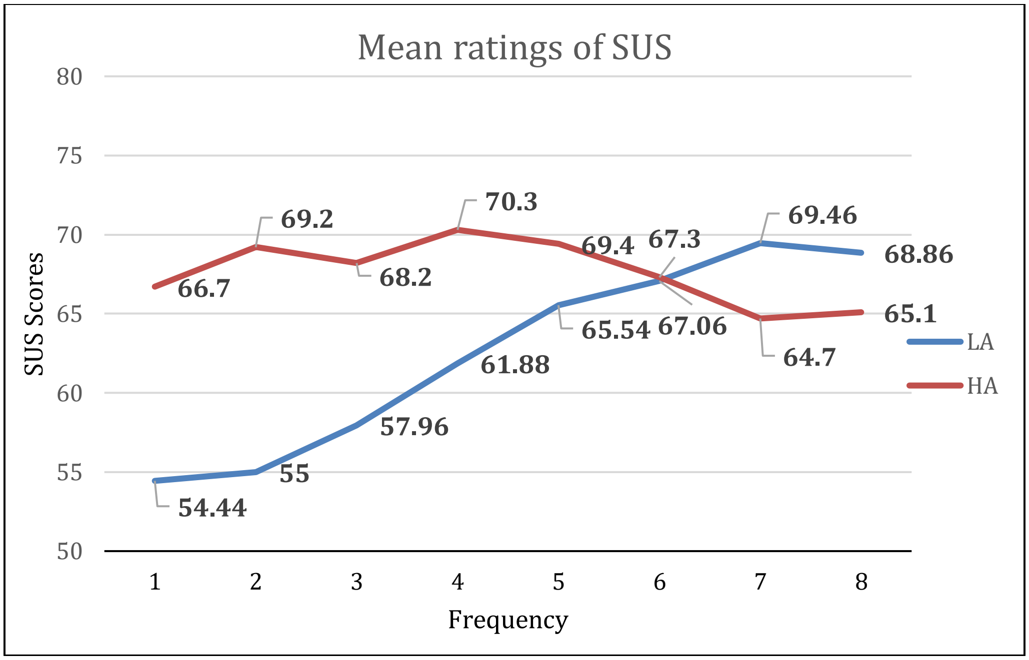

Figure 8. Mean ratings of SUS

Table 11. T-test for SUS of each time

	1st	2nd	3rd	4th	5th	6th	7th	8th
M	-12.26	-14.2	-10.24	-8.42	-3.86	-0.24	4.76	3.76
p	<.05	<.05	0.08	0.14	0.51	0.96	0.36	0.49

We can observe in Figure 8 that perceptions of usability increase steadily in the low-aesthetics group and decrease in the high-aesthetics group. During the first two tasks, SUS scores are significantly higher in the HA group than in the LA group. However, even the mean difference is still large during tasks 3 and 4, and the differences are not significant. At task 6, the mean SUS scores are very close (67.3 for HA and 67.06 for LA). Although the mean difference is not significant, the mean SUS scores in LA exceed HA during the last two tasks.

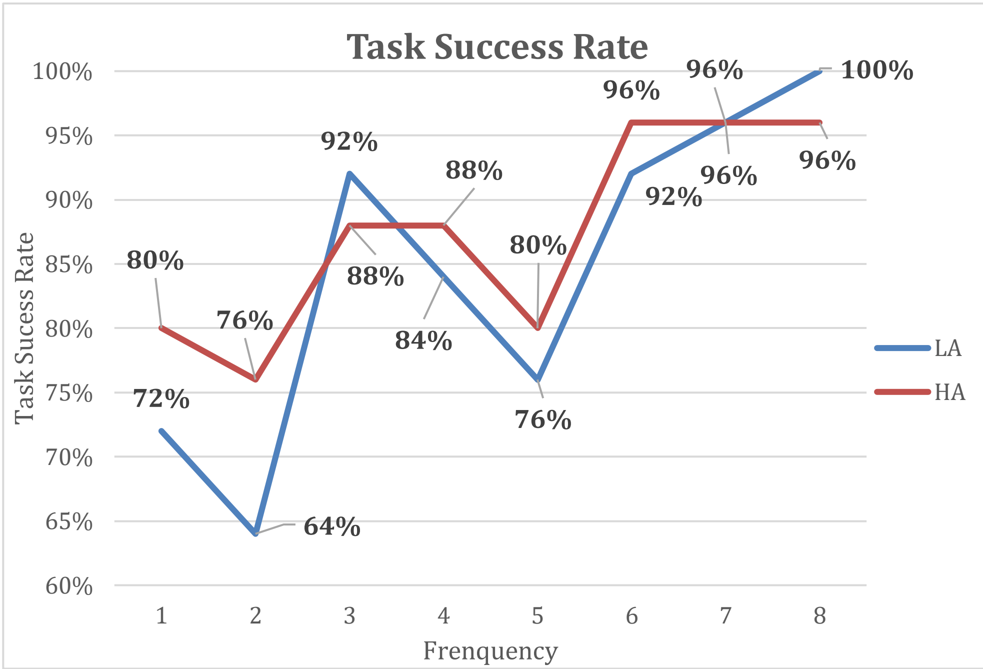

Figure 9. Mean task success rate

Table 12. T-test for task success rate of each time

	1st	2nd	3rd	4th	5th	6th	7th	8th
M	0.08	0.12	0.04	0.04	0.04	0.04	0	-0.04
p	0.51	0.35	0.64	0.68	0.73	0.55	1	0.31

Even with some fluctuations, the task success rate in both LA and HA retains the upward trend until the end of whole tasks. As task success rate is a binary variable, we performed a Wald test to compare the difference in proportions of the two groups. The result at each point is not significant, which is reasonable due to the fact that only 50 participants were enrolled in the experiment.

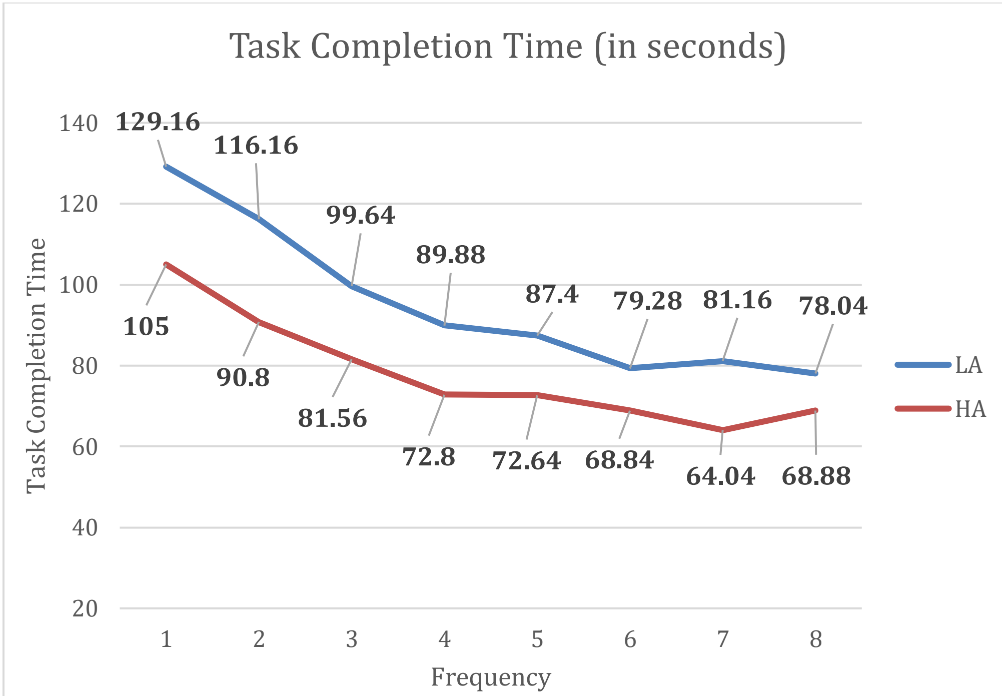

Figure 10. Mean task completion time

Table 13. T-test for task completion time of each task

	1st	2nd	3rd	4th	5th	6th	7th	8th
M	24.16	25.36	18.08	17.08	14.76	10.44	17.12	9.16
p	<.05	<.05	<.05	<.05	<.05	0.09	<.01	0.17

Task completion time in both groups show decreasing trends and flatten after the first several tasks. The mean difference between the two groups is significant - the LA group spent more time finishing the task than the HA group during the first four tasks. After task 5, the mean difference becomes insignificant, with the only exception of task 7 where the mean difference is significant (17.12, $p<0.05$).

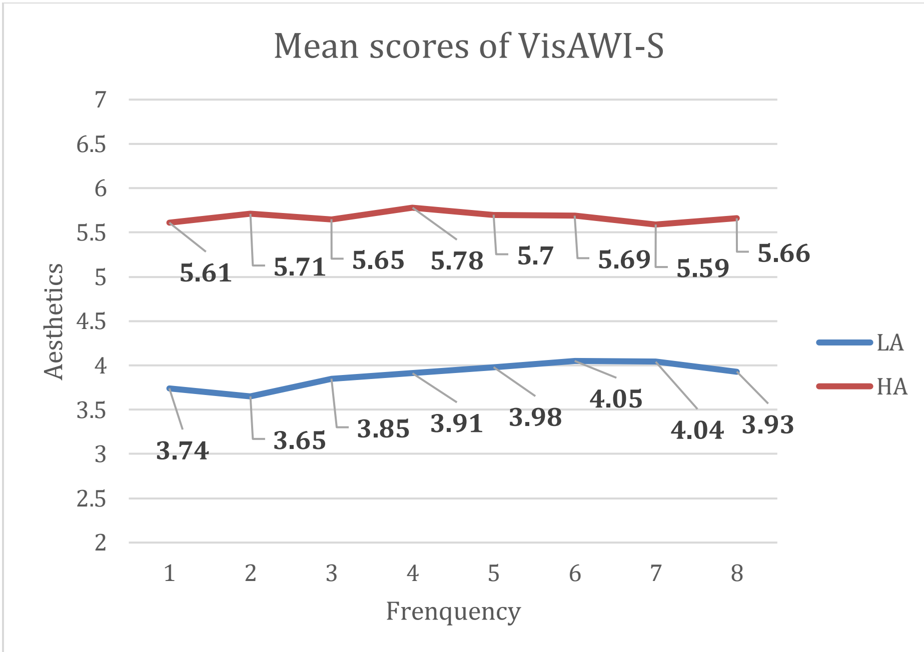

Figure 11. Mean scores of VisAWI-S

Table 14. T-test for task completion time of each time

	1st	2nd	3rd	4th	5th	6th	7th	8th
M	-1.87	-2.06	-1.8	-1.87	-1.72	-1.64	-1.55	-1.62
p	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

The perceptions of aesthetics of both groups remain stable and barely change during the whole experiment. The mean scores of VisAWI-S of the HA group are always higher than those of the LA group and the differences are always significant as well.

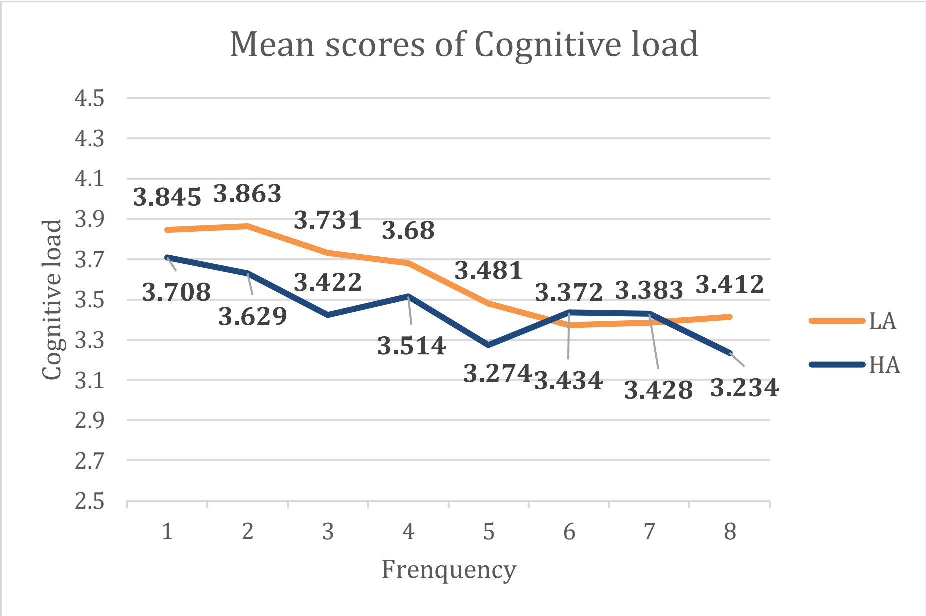

Figure 12. Mean scores of cognitive load

Table 15. T-test for cognitive load

	1st	2nd	3rd	4th	5th	6th	7th	8th
M	0.14	0.23	0.31	0.17	0.21	-0.06	-0.05	0.18
p	0.76	0.59	0.47	0.70	0.61	0.88	0.91	0.66

Perceived cognitive load in both groups dips with more interaction. The mean scores in the HA group are always higher than those in the LA group except after task 6 and task 7. However, the mean difference between two groups is not significant at all times.

5.3 Correlation test

As mentioned above, all variables were checked for linearity and normal distribution. Therefore, we adopted Pearson's correlation test to test whether the aesthetics-usability effect existed at first and then diminished. For the first step, we extracted a new metric, "task performance," from task completion time and task success rate by using principal component analysis to represent objective usability. This method was adopted from prior research (e.g, Miller, 2011; Tuch et al., 2012), which helped to build a comprehensive metric and made it easier to conduct quantitative analysis. The extracted metric explained 68% of the variance (see Appendix I for details). Table 16 shows the results of the correlation test.

Table 16. Results of Pearson's correlation test for perceived usability

	Perceived Usability		**Task Performance**	
	r	***p***	**r**	***p***
Aesthetics				
Task1	0.64	<.001	0.35	<0.05
Task2	0.66	<.001	0.32	<0.05
Task3	0.59	<.001	0.14	0.33
Task4	0.56	<.001	0.30	<0.05
Task5	0.31	<.05	0.31	<.05
Task6	0.17	0.23	0.45	<0.01
Task7	0.17	0.24	0.25	0.08
Task8	0.18	0.21	0.19	0.19

The correlations between perceived usability and visual aesthetics were significant during the first 5 tasks. However, the coefficient was reduced with continuous interactions with the system, thus it became insignificant from task 6 (p <0.05). This result supports Hypothesis 1; the impact of visual aesthetics on perceived usability gradually wanes and finally disappears.

The correlations between task performance and visual aesthetics were significant during the first 6 tasks except task 3, and became insignificant after task 7 and task 8. Therefore, we determined that the impact of visual aesthetics exists during task 1 to task 6 and disappears after task 7. This evidence supports Hypothesis 2; the impact of visual aesthetics on task performance diminishes after several interactions.

5.4 Moderation analysis

In order to check whether interaction frequency and cognitive load moderate the aesthetics-usability effect, linear mixed models were applied in our study. Compared with the traditional repeated measurements of ANOVA, the linear mixed model could better handle dependencies in repeated measures data (Kim et al., 2019). In our model, participants were treated as random factors to account for the group effect within interaction frequency.

We first built the model for perceived usability. To ensure that the model could best explain the variance of the dependent variable in the fixed effect part, we introduced interaction frequency, cognitive load and visual aesthetics as the main effects. Then, two-way interactions, aesthetics×frequency and aesthetics×cognitive load, were added to the model to check whether the moderating effects truly exist. Therefore, the question of testing the moderating effects was converted to calculating the covariance of aesthetics× frequency and aesthetics×cognitive load. In this model, 54.56% of the variance was explained by those predictors in the fixed effect part. In addition, to find the best fitted model for the random effect part, we applied the maximum likelihood method and compared AIC and BIC of nested models. Table 17 below indicates the results from regression of the mixed model.

Table 17. Mixed-effects regression of perceived usability

	Perceived Usability		
	B	**CI**	***P***
Intercept	47.45	28,93 to 65.98	<.001
Aesthetics	7.93	4.28 to 11.57	<.001
Frequency	5.17	3.79 to 6.55	<.001
Cognitive load	-9.63	-13.09 to -6.16	<.001
Aesthetics×Frequency	-0.99	-1.28 to -0.72	<.001
Aesthetics×Cognitive load	0.73	0.01 to 1.44	<.05
Random parts			
Model Type	Compound Symmetry		
AIC	2929.8		
BIC	2933.6		
p	0.67		
Observations	400		
χ^2	277.38		

Results indicate that the covariant of aesthetics and interaction frequency is -0.99 which is significant (p <0.05). The covariant is negative, which means that the influence of visual aesthetics gradually reduces with continuous interaction of the system. The result is reflected in the interaction chart below (Fig. 13); we can see that the slope of visual aesthetics to usability flattens with the increment of interaction frequency. Subsequently, the slope becomes insignificant after 6 tasks were finished. Therefore, combined with the correlation test, Hypotheses 1 is fully supported; interaction frequency plays the role of moderator in reducing the impact of visual aesthetics on perceived usability such that the aesthetics-usability effect diminishes when interaction frequency increases.

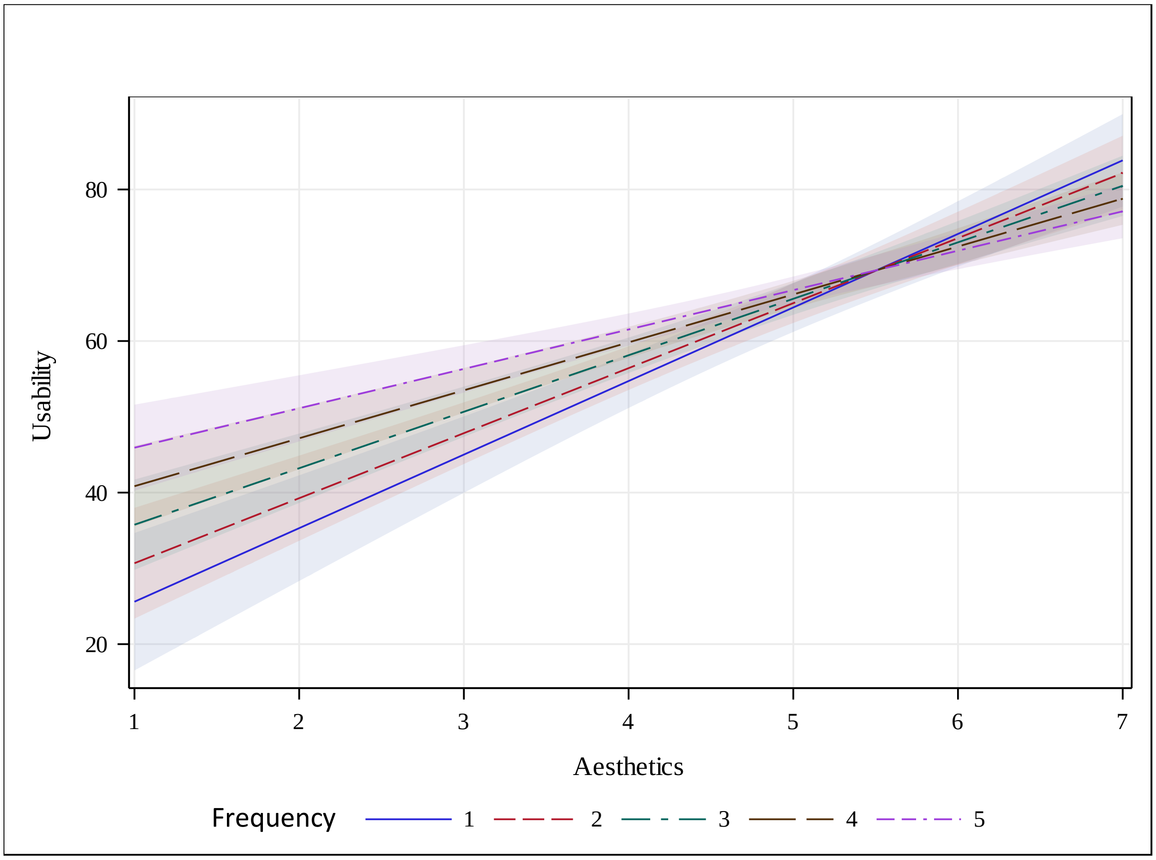

Figure 13. Interaction plot of aesthetics and interaction frequency on perceived usability

The covariant of aesthetics and cognitive load is 0.73, which is also significant (*p* <0.05). Figure 14 below reveals the moderating effect of cognitive load. As the mean scores of cognitive load decrease with interaction frequency increments, the positive moderating effect of cognitive load on the aesthetics-usability correlation gradually reduces, which is reflected in Figure 14. Therefore, Hypothesis 3 is fully supported; cognitive load also plays a role of moderator in the aesthetics-usability effect such that the impact of visual aesthetics on perceived usability gradually diminishes when cognitive load is reduced.

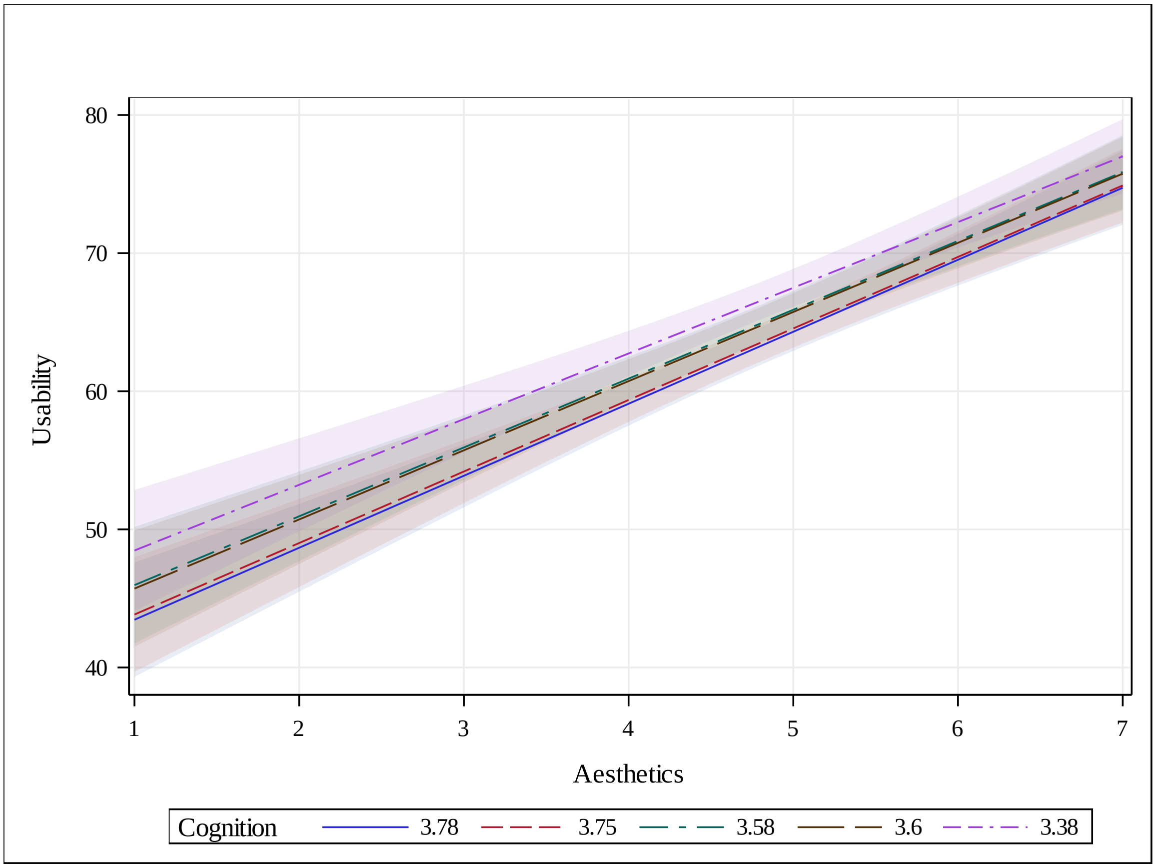

Figure 14. Interaction plot of aesthetics and cognitive load on perceived usability

The same method was applied for task performance. In this model, 24.44% of the variance is explained by all the predictors in the fixed effect part. Table 18 below indicates the results of the mixed regression model. The results indicate that the covariant of aesthetics and interaction frequency is -0.04 ($p < 0.05$), which means that the impact of visual aesthetics on task performance gradually reduces with increased interaction frequency. Figure 15 shows the result as the slope of visual aesthetics to task performance that flattens with interaction frequency increments. Therefore, combined with the result of the correlation test, the evidence fully supports Hypothesis 3. However, the covariant of visual aesthetics and cognitive load is not significant. Consequently, Hypothesis 4 is not supported as the result of insufficient evidence.

Table 18. Mixed-effects regression of task performance

	Task Performance		
	B	**CI**	***P***
Intercept	-1.48	-3.17 to 0.20	0.08
Aesthetics	0.24	-0.08 to 0.56	0.15
Frequency	0.36	0.20 to 0.52	<.001
Cognitive loads	-0.26	-0.59 to 0.07	0.12
Aesthetics × Frequency	-0.04	-0.07 to -0.01	<.05
Aesthetics × Cognitive load	0.03	-0.04 to 0.09	0.43
Random parts			
Model Type	Compound Symmetry		
AIC	1148.1		
BIC	1155.8		
p	0.18		
Observations	400		
χ^2	27.55		

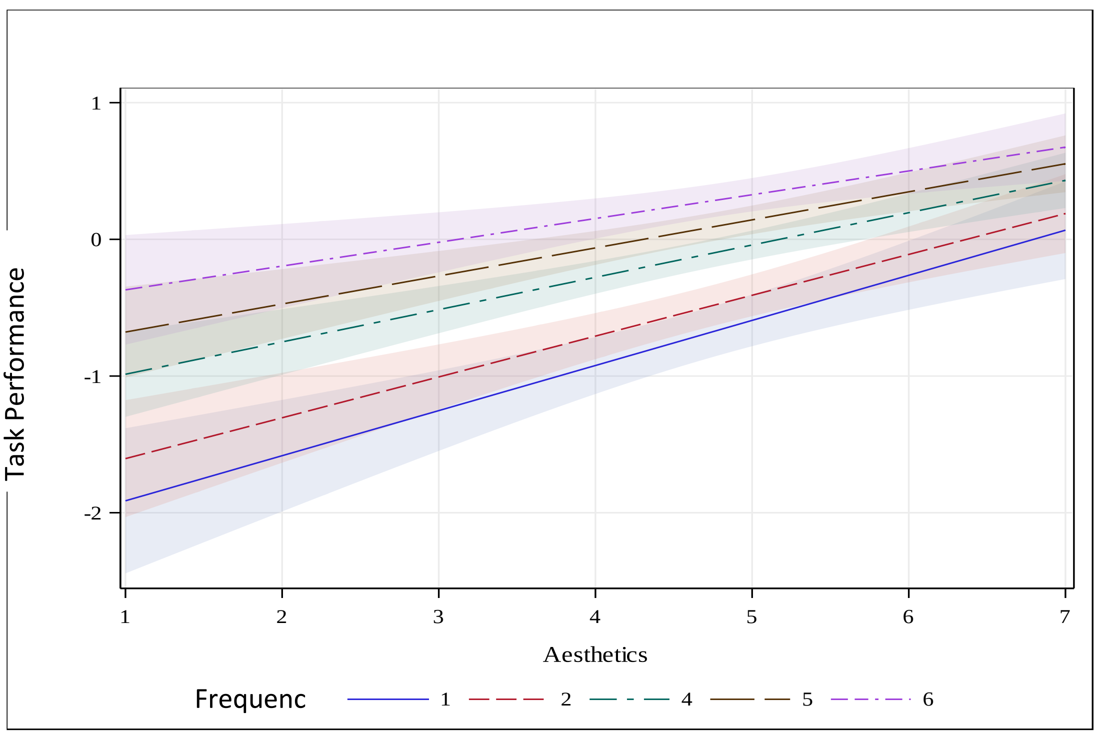

Figure 15. Interaction plot of aesthetics and interaction frequency on task performance

Chapter 6 Discussion

6.1 General discussion

In this thesis, we propose and explore two BCs of the aesthetics-usability effect, namely, the different contexts of interaction frequency and cognitive load.

First, the results from the analyses reveal that there is a positive impact of visual aesthetics on both perceived usability and task performance at first, and that this impact gradually diminishes with the increment of interaction frequency. More specifically, Pearson's correlation tests indicate that there is a significantly positive correlation between visual aesthetics and perceived usability for the first 5 tasks. This means that when the visual aesthetics of the interface is higher, the user's perception of usability is higher. Then, the correlation between visual aesthetics and perceived usability becomes insignificant from task 6, which means that the influence of visual aesthetics on the user's perception of usability diminishes after several interactions with the system. Regarding objective usability, there is also a significant correlation between visual aesthetics and task performance for the first 6 tasks and the correlation diminishes after that.

The results of the mixed model indicate that interaction frequency has a moderating effect on both perceived usability and task performance. More specifically, interaction frequency reduces the influence of visual aesthetics on usability. In other words, the more interactions participants have with the website, the weaker the influence of visual aesthetics on usability becomes. The results are reflected by the negative covariant of aesthetics × frequency as the influence of visual aesthetics to perceived usability are reduced with interaction frequency increments. Similarly, the moderating effect of interaction frequency also weakens the influence of visual aesthetics on task performance.

Second, our results from the mixed model suggest that cognitive load weakens the impact of visual aesthetics on perceived usability. The covariant of aesthetics × cognitive load is positive, which means cognitive load increases the impact of visual aesthetics on perceived usability. The reason may be that, at first serval interactions when users spend a lot of mental effort on the interface, the aesthetics design helps them have a better perception of usability. However, as cognitive load reduced with more interactions, the influence of visual aesthetics on perceived usability weakens.

6.2 Theoretical contributions

Considering these results, our study provides several theoretical contributions. First, as previous studies emphasize the importance of exploring BCs of the aesthetics-usability effect, especially under temporal conditions, this thesis explores the changes of the aesthetics-usability effect in a short time period (orientation phase). Our study reveals that the impact of visual aesthetics on both perceived usability and task performance diminishes in a short time period. Thus, we find that a possible answer to the divergence in the aesthetics-usability effect lies in temporal conditions because the aesthetics-usability effect only exists for a short time.

Second, both interaction frequency and cognitive load are identified as moderators to the aesthetics-usability effect. Even though prior research has already confirmed that temporal conditions are BCs that can change the aesthetics-usability effect, this thesis, this thesis treats interaction frequency as a moderator to further increase the accuracy of the applicability of the aesthetics-usability effect in terms of temporal conditions. Additionally, our study identifies cognitive load as another moderator that increases the impact of visual aesthetics to perceived usability based on the findings concerning interaction frequency. As cognitive load reduces with more interactions, its moderating effect to the impact of aesthetics on perceived usability decreases. Thus, two BCs, interaction frequency and cognitive load, have moderating effects changing the relationship between visual aesthetics and usability.

Third, based on calls from previous researchers to find more complete ways of assessing usability, this thesis investigates both objective and subjective usability. Our study finds similar results for both perceived usability and task performance; the influence of visual aesthetics on both dimensions of usability gradually diminishes with more interactions. Thus, in the context of utilitarian use, both the objective and subjective measures of usability lead to similar conclusions regarding the usability of the interface and the aesthetics-usability effect. Rather than investigating one dimension of usability, our study contributes by having a more balanced focus on both the objective and subjective measures of usability; researching usability not only improves UX but also improves objective performance.

6.3 Practical implications

The results of our study also provide useful insights for digital product development. Aesthetics design can be applied to providing users with a good first impression as well as reducing difficulty when starting to use a product. As noted earlier, users make the decision to stay and purchase a product within a short time period. Given that the aesthetics-usability effect diminishes very quickly, it is important to effectively use aesthetics design at the earliest stage of product use to outshine other fiercely competing products. That is to say, we should apply aesthetic design to improve both objective and subjective usability. For objective usability and task performance, rather than using conventional tasks that require a long time to perform, we propose that the ideal task should be less complex, requiring less time to complete. As aesthetics design helps in having higher perceived usability during this shorter time period, improvement of both objective and subjective usability helps to increase user retention and engagement. It is worth noting that some examples of viral applications already exist, for example, the success of TikTok (watching 15 seconds video clips) and Instagram (browsing photos) could partially be ascribed to their application of aesthetics design and less complex tasks that are worthy of references.

However, the suggestions we propose above do not imply that many resources should be allocated to UX/UI design work. The dramatically growing cost of digital product development should be avoided because it could result in failure as well. Regardless of the scale, all projects could apply traditional designs that are consistent with the digital products most users are familiar with. Some cost-effective methods such as ergonomic evaluation could also be used to check minor usability issues. Ultimately, less workload is required for users to interact with digital products so that they can get started and become familiar with them much faster. This corresponds to that famous notion, “What is beautiful is usable.”

6.4 Limitations and further research

The findings of our study are limited in several ways. First, as preliminary research that explores the changes of the aesthetics-usability effect over a short time period, interaction frequency was regarded as the moderator. However, we propose that there could possibly be some deeper rationale, that of cognitive load. We used self-report measurement to measure perceived cognitive load. But self-report measurement has its own shortcomings because it is not an objective measurement (Klepsch et al., 2017). Thus, we highly recommend that future studies apply neurophysiological measurements such as EEG eye-tracking to measure physiological cognitive load data to examine the moderating effect of cognitive load in the aesthetics-usability correlation.

Second, there are other task performance metrics that could be included in measuring objective usability. Due to the fact that we did not have technical methods to collect other metrics like task errors and page clicks in our experiment, this has clearly had influence on the accuracy of the results. Consequently, we recommend further research to collect more task performance metrics to comprehensively evaluate objective usability.

Finally, our study has a goal-oriented task focus. Other research has shown/proven that task type has certain influence while users are evaluating a system (van Schaik & Ling, 2009). Typically, tasks can be divided into two categories: goal-oriented tasks which have a clear target that leads users to focus on the usability aspect while judging the system; hedonic tasks which do not have a clear target and that focus more on the usability aspect of the system. We recommend that future research considers task type as another factor while exploring the aesthetics-usability effect.

Chapter 7 Conclusion

This thesis was first inspired by the research finding that users tend to visit an online retailing website several times in short sessions before they decide to make a purchase and become loyal customers. Prior research suggests that usability plays an important role in users' decision-making during this short time period and that visual aesthetics have positive impact on usability; this is called the "aesthetic-usability effect."

From a theoretical perspective, there is controversy about whether the correlation between aesthetics and usability truly exists. Our research suggests that this divergence results from previous studies that did not take into consideration strong temporal conditions. For example, Sonderegger et al. (2012) found that the impact of aesthetics on usability wanes after a long period of interaction with systems (two weeks). Whereas Sonderegger et al. (2012)'s study found that the aesthetics-usability effect only exists at the first orientation phase (first day), our study is dedicated to exploring whether the aesthetics-usability effect diminishes in this short time period. Thus, we propose that there are two BCs in the aesthetics-usability effect, namely, interaction frequency (how many times tasks have been finished) and cognitive load, which both have moderating effects in changing the relationship between aesthetics and usability.

Following a literature review, a usability testing experiment was undertaken to solve our research questions. The findings from our 50 subjects are significant: the influence of visual aesthetics on both perceived usability and task performance diminishes after the first several uses; interaction frequency weakens the correlation between aesthetics and usability (both objective and subjective) while cognitive load moderates the impact of aesthetics to perceived usability.

The results provide useful insights into business implications: digital product development should apply aesthetics design for tasks that require less time to complete because users tend to make decisions after only a very short interaction time. This strategy improves the objective and subjective usability of their products, allowing their products to stand out amid fierce competition and retaining users. From a theorical perspective, the results also improve the accuracy of applicability of the aesthetics-usability effect in terms of empirical contexts.

In summary, significant results were found as the aesthetics-usability effect wanes in short time periods. However, limited evidence was found supporting the moderating effect of cognitive loads to the impact of aesthetics on objective usability. Thus, future research could be conducted by using neurophysiological measurements to improve our results.

Bibliography

Altaboli, A. (2013). Investigating the effects of font styles on perceived visual aesthetics of website interface design. International Conference on Human-Computer Interaction,

Bamberger, P. (2008). From the Editors Beyond Contextualization: Using Context Theories to Narrow the Micro-Macro Gap in Management Research. *Academy of Management Journal*, *51*(5), 839-846. https://doi.org/10.5465/amj.2008.34789630

Bauerly, M., & Liu, Y. (2008). Effects of symmetry and number of compositional elements on interface and design aesthetics. *Intl. Journal of Human–Computer Interaction*, *24*(3), 275-287.

Ben-Bassat, T., Meyer, J., & Tractinsky, N. (2006). Economic and Subjective Measures of the Perceived Value of Aesthetics and Usability. *ACM TRANSACTIONS ON COMPUTER HUMAN INTERACTION*, *13*(2), 210-234.

Bridgeman, B., Lennon, M. L., & Jackenthal, A. (2003). Effects of screen size, screen resolution, and display rate on computer-based test performance. *Applied Measurement in Education*, *16*(3), 191-205.

Busse, C., Kach, A. P., & Wagner, S. M. (2017). Boundary Conditions: What They Are, How to Explore Them, Why We Need Them, and When to Consider Them. *Organizational Research Methods*, *20*(4), 574-609. https://doi.org/10.1177/1094428116641191

Chawda, B., Craft, B., Cairns, P., Heesch, D., & Rüger, S. (2005). Do "attractive things work better"? An exploration of search tool visualisations.

Coursaris, C. K., & Kripintris, K. (2012). Web aesthetics and usability: An empirical study of the effects of white space. *International Journal of E-Business Research (IJEBR)*, *8*(1), 35-53.

Coursaris, C. K., Swierenga, S. J., & Watrall, E. (2008). An empirical investigation of color temperature and gender effects on web aesthetics. *Journal of usability studies*, *3*(3), 103-117.

Cyr, D., Head, M., & Ivanov, A. (2006). Design aesthetics leading to m-loyalty in mobile commerce. *Information & management*, *43*(8), 950-963.

De Angeli, A., Lynch, P., & Johnson, G. I. (2002). CHAPTER SEVEN Pleasure versus Efficiency inUser Interfaces: Towards an Involvement Framework. *Pleasure with products: Beyond usability*, 94.

De Angeli, A., Sutcliffe, A., & Hartmann, J. (2006). Interaction, usability and aesthetics: what influences users' preferences? Proceedings of the 6th conference on Designing Interactive systems,

Douneva, M., Haines, R., & Thielsch, M. T. (2015). Effects of interface aesthetics on team performance in a virtual task.

Flavián, C., Guinalíu, M., & Gurrea, R. (2006). The role played by perceived usability, satisfaction and consumer trust on website loyalty. *Information & management*, *43*(1), 1-14.

Forlizzi, J., & Battarbee, K. (2004). Understanding experience in interactive systems. Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques,

Gould, J. D., & Lewis, C. (1985). Designing for usability: key principles and what designers think. *Communications of the ACM*, *28*(3), 300-311.

Grishin, J. (2018). Exploring the Boundary Conditions of the Effect of Aesthetics on Perceived Usability.

Haier, R. J., Siegel Jr, B. V., MacLachlan, A., Soderling, E., Lottenberg, S., & Buchsbaum, M. S. (1992). Regional glucose metabolic changes after learning a complex visuospatial/motor task: a positron emission tomographic study. *Brain research*, *570*(1-2), 134-143.

Hamborg, K.-C., Hülsmann, J., & Kaspar, K. (2014). The interplay between usability and aesthetics: More evidence for the "what is usable is beautiful" notion. *Advances in Human-Computer Interaction*, *2014*.

Hartmann, J., Sutcliffe, A., & De Angeli, A. (2007). Investigating attractiveness in web user interfaces. Proceedings of the SIGCHI conference on Human factors in computing systems,

Hasbullah, N. A., Osman, A., Abdullah, S., Salahuddin, S. N., Ramlee, N. F., & Soha, H. M. (2016). The relationship of attitude, subjective norm and website usability on consumer intention to purchase online: An evidence of Malaysian youth. *Procedia*

Economics and Finance, *35*, 493-502.

Hassenzahl, M. (2004). The Interplay of Beauty, Goodness, and Usability in Interactive Products. *Human-Computer Interaction*, *19*(4), 319-349. https://doi.org/10.1207/s15327051hci1904_2

Hoffmann, R., & Krauss, K. (2004). A critical evaluation of literature on visual aesthetics for the web. Proceedings of the 2004 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries,

Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human - Computer Studies*, *64*(2), 79-102. https://doi.org/10.1016/j.ijhcs.2005.06.002

Johnson, P.-A. (1994). *The Theory of Architecture: Concepts Themes & Practices*. John Wiley & Sons.

Karapanos, E., Hassenzahl, M., & Martens, J.-B. (2008). User experience over time. In *CHI'08 extended abstracts on Human factors in computing systems* (pp. 3561-3566).

Katz, A. (2010). Aesthetics, Usefulness and Performance in User--Search-Engine Interaction. *Journal of Applied Quantitative Methods*, *5*(3), 424-445.

Kim, J. Y.-J., Block, C. J., & Nguyen, D. (2019). What's visible is my race, what's invisible is my contribution: Understanding the effects of race and color-blind racial attitudes on the perceived impact of microaggressions toward Asians in the workplace. *Journal of Vocational Behavior*, *113*, 75-87.

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in psychology*, *8*, 1997.

Kurosu, M., & Kashimura, K. (1995). *Apparent usability vs. inherent usability: experimental analysis on the determinants of the apparent usability* Conference Companion on Human Factors in Computing Systems, Denver, Colorado, USA. https://doi.org/10.1145/223355.223680

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An Introduction to Latent Semantic Analysis. *Discourse Processes*, *25*(2-3), 259-284.

Lavie, T., Oron-Gilad, T., & Meyer, J. (2011). Aesthetics and usability of in-vehicle navigation displays. *International Journal of Human-Computer Studies*, *69*(1-2), 80-99.

Lavie, T., & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-Computer Studies*, *60*(3), 269-298.

Lee, S., & Ha, T. (2019). Changes in perceived usability and aesthetics with repetitive use in the first use session. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *29*(6), 517-528. https://doi.org/10.1002/hfm.20814

Lindgaard, G., Dudek, C., Sen, D., Sumegi, L., & Noonan, P. (2011). An exploration of relations between visual appeal, trustworthiness and perceived usability of homepages. *ACM Transactions on Computer-Human Interaction*, *18*(1). https://doi.org/10.1145/1959022.1959023

Longo, L., Rusconi, F., Noce, L., & Barrett, S. (2012). The Importance of Human Mental Workload in Web Design. WEBIST,

Marcus, A. (1992). *Graphic design for electronic documents and user interfaces*.

Miller, C. (2011). Aesthetics and e-assessment: The interplay of emotional design and learner performance. *Distance Education*, *32*(3), 307-337. https://doi.org/10.1080/01587919.2011.610291

Miller, R. B. (1971). *Human ease of use criteria and their tradeoffs*. IBM, Systems Development Division, Poughkeepsie Lab.

Moe, W. (2003). Buying, Searching, or Browsing: Differentiating Between Online Shoppers Using In-Store Navigational Clickstream. *JOURNAL OF CONSUMER PSYCHOLOGY*, *13*(1/2), 29-39.

Moshagen, M., Musch, J., & Göritz, A. S. (2009). A blessing, not a curse: Experimental evidence for beneficial effects of visual aesthetics on performance. *Ergonomics*, *52*(10), 1311-1320.

Moshagen, M., & Thielsch, M. (2013). A short version of the visual aesthetics of websites inventory. *BEHAVIOUR AND INFORMATION TECHNOLOGY*, *32*(12), 1305-1311.

Nasar, J. L. (2016). 10 Perception and Evaluation of Residential Street Scenes. *Directions in Person-Environment Research and Practice (Routledge Revivals)*, 229.

Nicholls, J. (1979). Programming the end user. *Man–Computer Communication, Infotech State of the Art Report*, *2*.

Norman, D. A. (2004). *Emotional design: Why we love (or hate) everyday things*. Basic Civitas Books.

Oyibo, K., & Vassileva, J. (2017a). The Interplay of Aesthetics, Usability and Credibility in Mobile Website Design and the Effect of Gender. *SBC Journal on Interactive Systems*, *8*(2), 4-19.

Oyibo, K., & Vassileva, J. (2017b). What drives perceived usability in mobile Web design: classical or expressive aesthetics? International Conference of Design, User Experience, and Usability,

Parizotto-Ribeiro, R., Hammond, N., Mansano, J., & Cziulik, C. (2004). Aesthetics and perceived usability of VLEs: preliminary results. *Proceedings of Human Computer Interaction*, 217-221.

Peng, X., Peak, D., Prybutok, V., & Xu, C. (2017). The effect of product aesthetics information on website appeal in online shopping. *Nankai Business Review International*.

Phillips, B. K., Prybutok, V., & Peak, D. (2014). Determinants of task performance in a visual decision-making process. *Journal of Decision Systems*, *23*(4), 373-387.

Quinn, J. M., & Tran, T. Q. (2010). Attractive phones don't have to work better: independent effects of attractiveness, effectiveness, and efficiency on perceived usability. Proceedings of the SIGCHI conference on human factors in computing systems,

Raptis, D., Tselios, N., Kjeldskov, J., & Skov, M. B. (2013). Does size matter? investigating the impact of mobile phone screen size on users' perceived usability, effectiveness and efficiency. Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services,

Reppa, I., & McDougall, S. (2015). When the going gets tough the beautiful get going: aesthetic appeal facilitates task performance. *Psychonomic bulletin & review*, *22*(5), 1243-1254.

Reppa, I., McDougall, S., Sonderegger, A., & Schmidt, W. C. (2021). Mood moderates the effect of aesthetic appeal on performance. *Cognition and Emotion*, *35*(1), 15-29.

Sánchez-Franco, M. J., & Martín-Velicia, F. A. (2011). The interaction effects of ego involvement on the relationships between aesthetics, usability and commitment: an application to electronic banking services and virtual travel communities. *Online Information Review*.

Satpute, A. B., & Lieberman, M. D. (2006). Integrating automatic and controlled processes into neurocognitive models of social cognition. *Brain research*, *1079*(1), 86-97.

Sauer, J., & Sonderegger, A. (2009). The influence of prototype fidelity and aesthetics of design in usability tests: Effects on user behaviour, subjective evaluation and emotion. *Applied ergonomics*, *40*(4), 670-677. https://doi.org/10.1016/j.apergo.2008.06.006

Schmidt, K. E., Liu, Y., & Sridharan, S. (2009). Webpage aesthetics, performance and usability: Design variables and their effects. *Ergonomics*, *52*(6), 631-643.

Sénécal, S., Fredette, M., Léger, P.-M., Courtemanche, F., & Riedl, R. (2015). Consumers' Cognitive Lock-in on Websites: Evidence from a Neurophysiological Study. *Journal of Internet Commerce*, *14*(3), 277-293. https://doi.org/10.1080/15332861.2015.1028249

Seo, K.-K., Lee, S., & Chung, B. D. (2016). Effects of perceived usability and aesthetics on emotional responses in different contexts of use. *International Journal of Human-Computer Interaction*, *32*(6), 445-459.

Shackel, B. (1984). The concept of usability. *Visual display terminals: Usability issues and health concerns*, 45-87.

Shackel, B. (2009). Usability – Context, framework, definition, design and evaluation. *Interacting with computers*, *21*(5-6), 339-346. https://doi.org/10.1016/j.intcom.2009.04.007

Sonderegger, A., & Sauer, J. (2010). The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. *Applied ergonomics*, *41*(3), 403-410.

Sonderegger, A., Zbinden, G., Uebelbacher, A., & Sauer, J. (2012). The influence of product aesthetics and usability over the course of time: a longitudinal field experiment. *Ergonomics*, *55*(7), 713-730. https://doi.org/10.1080/00140139.2012.672658

Sweller, J. (2011). Cognitive load theory. In *Psychology of learning and motivation* (Vol. 55, pp. 37-76). Elsevier.

Thielsch, M. T., Haines, R., & Flacke, L. (2019). Experimental investigation on the effects of website aesthetics on user performance in different virtual tasks. *PeerJ*. https://doi.org/http://dx.doi.org/10.7717/peerj.6516

Thüring, M., & Mahlke, S. (2007). Usability, aesthetics and emotions in human-technology interaction. *International Journal of Psychology*, *42*(4), 253-264. https://doi.org/10.1080/00207590701396674

Tractinsky, N., Katz, A. S., & Ikar, D. (2000). What is beautiful is usable. *Interacting with computers*, *13*(2), 127-145.

Travis, C., & Murano, P. (2014). A comparative study of the usability of touch-based and mouse-based interaction. *International Journal of Pervasive Computing and Communications*, *10*(1), 115-134.

Tuch, A. N., Roth, S. P., Hornbæk, K., Opwis, K., & Bargas-Avila, J. A. (2012). Is beautiful really usable? Toward understanding the relation between usability, aesthetics, and affect in HCI. *COMPUTERS IN HUMAN BEHAVIOR*, *28*(5), 1596-1607.

van Schaik, P., & Ling, J. (2003). The effect of link colour on information retrieval in educational intranet use. *COMPUTERS IN HUMAN BEHAVIOR*, *19*(5), 553-564.

van Schaik, P., & Ling, J. (2009). The role of context in perceptions of the aesthetics of web pages over time. *International Journal of Human-Computer Studies*, *67*(1), 79-89. https://doi.org/https://doi.org/10.1016/j.ijhcs.2008.09.012

Wood, J. R., & Wood, L. E. (2008). Card sorting: current practices and beyond. *Journal of usability studies*, *4*(1), 1-6.

Zettl, H. (2013). *Sight, sound, motion: Applied media aesthetics*. Nelson Education.

Zhuang, M., Toms, E. G., & Demartini, G. (2016). The relationship between user perception and user behaviour in interactive information retrieval evaluation. European Conference on Information Retrieval,

Appendix A: Demographic Information of official test

Table 19. Demographic information of official test

	Options	**Frequency**
Gender	Male	33
	Female	17
Age	18-29	18
	30-39	18
	40-49	11
	50-	3
Education Background	Highschool	15
	Bachelor's Degree	20
	Graduate Degree	6
	Post-Graduate Degree	8
	Others	1
Employment Status	Employed	42
	Studying	12
	Retired	4

Appendix B: Electronic Consent

User's judgement of website after multiple interactions

Dear participant,

I am a master's student majoring in User Experience at HEC Montreal. Currently, I am working on my master's thesis about user's judgement of website after multiple interactions. As such, I would like to invite you to participate in an experiment. The experiment will use a prototype of online shopping website and requires no previous experience.

There is total **8** tasks in our experiment and time limitation for each task is **3** minutes. Therefore, we have estimated that it should take about **30** minutes to finish all the tasks. We ask you to finish the questionnaire each time after you have finished one task. Since your first impressions best reflect your true opinions, we would ask that you please answer the questions included in this questionnaire without any hesitation. There is no time limit for completing the questionnaire, although we have estimated that it should take about **3** minutes to finish it each time. As you will have to repeat this process for **8** times, the estimated time for completing the whole experiment and questionnaire would be around **1** hour.

POTENTIAL RISKS AND DISCOMFORTS

There are no foreseeable physical, psychological, emotional, financial or social risks associated with this study. If for any reason, you feel uncomfortable moving forward with the study, you may at any point in time ask to withdraw from the study.

COMPENSATION FOR PARTICIPATION

There will be $ 15 cash compensation for your participation.

CONFIDENTIALITY

The information collected will be anonymous and will remain strictly confidential. It will be used solely for the advancement of knowledge and the dissemination of the overall results in academic or professional forums.

The online data collection provider agrees to refrain from disclosing any personal information (or any other information concerning participants in this study) to any other users or to any third party, unless the respondent expressly agrees to such disclosure or unless such disclosure is required by law.

PARTICIPATION AND WITHDRAWAL

You are free to refuse to participate in this project and you may decide to stop answering the questions at any time. By completing this questionnaire, you will be considered as having given your consent to participate in our research project and to the potential use of data collected from this questionnaire in future research.

Appendix C: Instructions

Dear Participant:

Thank you for your participation in our experiment. This experiment is mainly about user's perception of website after multiple interaction. This experiment will take about **60** minutes. The whole procedure is:

1. Fill in the demographic questionnaire (3 minutes)

2. Finish the task in an online shopping website (3 minutes)

3. Complete the post-questionnaire (3 minutes)

Then repeat step **2** and **3** for **8** times. The task is not difficult, you are required to find the targeted product and add it into shopping cart in **3** minutes every time. If you can't find the product in designated time or find the wrong product, the task will be regarded as a failure but still finished as one time. There will be a pop out window with a timer when you start the task, if you can't find the product in **3** minutes, you can click 'give up'. Because there are multiple questionnaires so please write the last three digits of your worker's ID in every questionnaire for my data collection work.

*Note: 1. You may see the same product picture for several times at the instruction page every time you start the task.

2. This is just a prototype so be aware that not every function and link is usable.

3. Only if all **8** responses have been collected then the compensation will be allocated to you.

Appendix D: Demographic Questionnaire

1.Please write down your age _____

2.What is your gender?

□Male □Female □Prefer not to answer

3.What is the highest level of education you have completed or are currently enrolled in?

□Highschool □Bachelor's Degree □Graduate Degree □Post-Graduate Degree

□Others

4.Are you currently…?

□Employed □A student □Retired

Appendix E: SUS Questionnaire

(1) I think that I would like to use this website frequently

Strongly Disagree 1 2 3 4 5 Strongly Agree

(2) I found this website unnecessarily complex

Strongly Disagree 1 2 3 4 5 Strongly Agree

(3) I thought the website was easy to use

Strongly Disagree 1 2 3 4 5 Strongly Agree

(4) I think that I would need the support of a technical person to be able to use this website

Strongly Disagree 1 2 3 4 5 Strongly Agree

(5) I found the various functions in this website were well integrated

Strongly Disagree 1 2 3 4 5 Strongly Agree

(6) I thought there was too much inconsistency in this website

Strongly Disagree 1 2 3 4 5 Strongly Agree

(7) I would imagine that most people would learn to use this website very quickly

Strongly Disagree 1 2 3 4 5 Strongly Agree

(8) I found the website very cumbersome to use

Strongly Disagree 1 2 3 4 5 Strongly Agree

(9) I felt very confident using the website

Strongly Disagree 1 2 3 4 5 Strongly Agree

(10) I needed to learn a lot of things before I could get going with this website

Strongly Disagree 1 2 3 4 5 Strongly Agree

The odd items of SUS questionnaire are phrased in positive tone while the even items of SUS questionnaire are phrased with negative tone. The alternating tone is intended to reduce acquiescence and extreme response biases. Below is the calculation formula of SUS scores:

[(Q1-1) +(5-Q2) +(Q3-1) +(5-Q4) +(Q5-1) +(5-Q6) +(Q7-1) +(5-Q8) +(Q9-1) +(5-Q10)] *2.5.

Appendix F: VisAWI-S Questionnaire

(1) Everything goes together on this website

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

(2) The layout is pleasantly varied on this website

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

(3) The color composition is attractive on this website

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

(4) The layout on this website appears professionally designed

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

Appendix G: Perceived Cognitive Load Questionnaire

(1) For this task, many things needed to be kept in mind simultaneously.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

(2) The task was very complex.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

(3) For the task, I had to highly engage myself.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

(4) For this task, I had to think intensively what things meant.

Strongly Disagree 1 2 3 4 5 Strongly Agree

(5) During this task, it was exhausting to find the important information.

Strongly Disagree 1 2 3 4 5 Strongly Agree

(6) The design of this task was very inconvenient for learning.

Strongly Disagree 1 2 3 4 5 Strongly Agree

(7) During this task, it was difficult to recognize and link the crucial information.

Strongly Disagree 1 2 3 4 5 Strongly Agree

Appendix H: Graph check of normal distribution and linearity

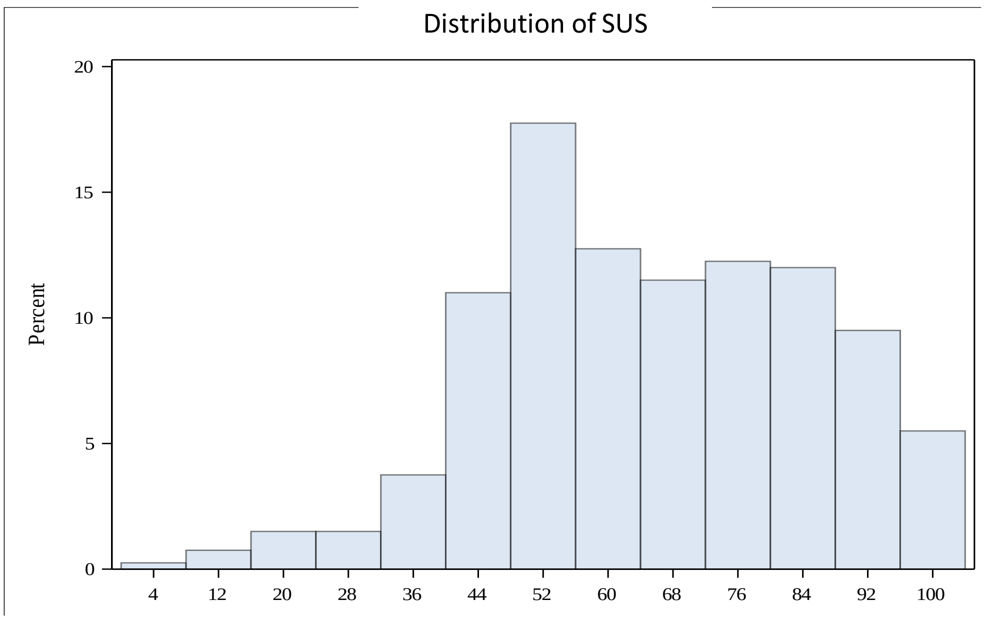

Figure 16. Distribution Plot of SUS

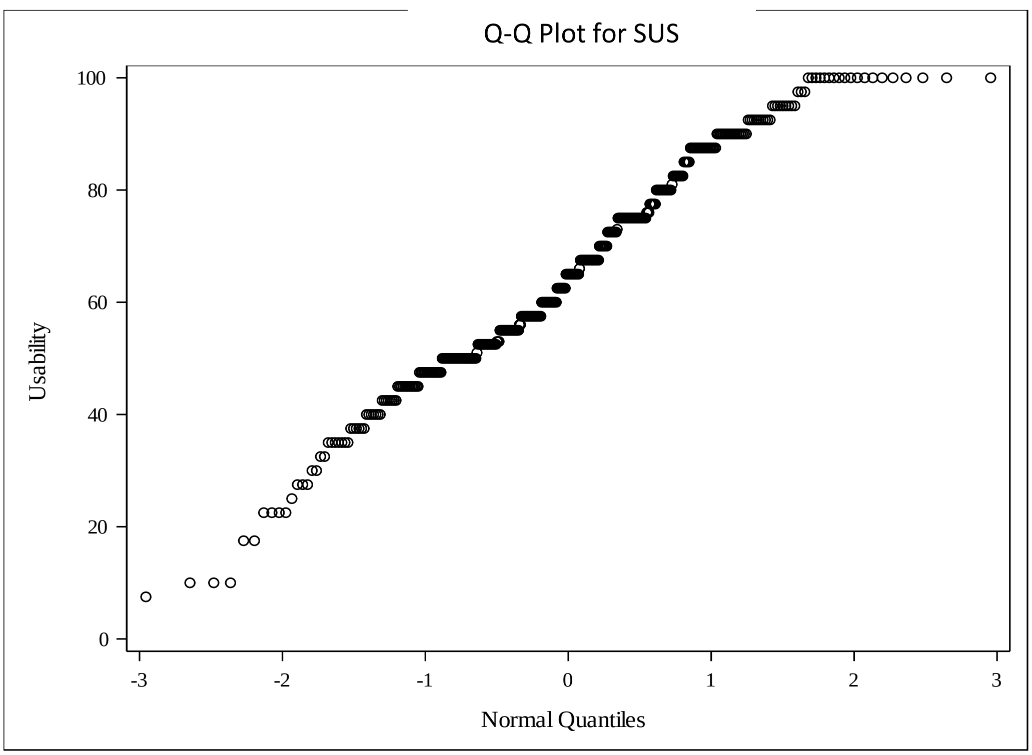

Figure 17. Q-Q Plot for SUS

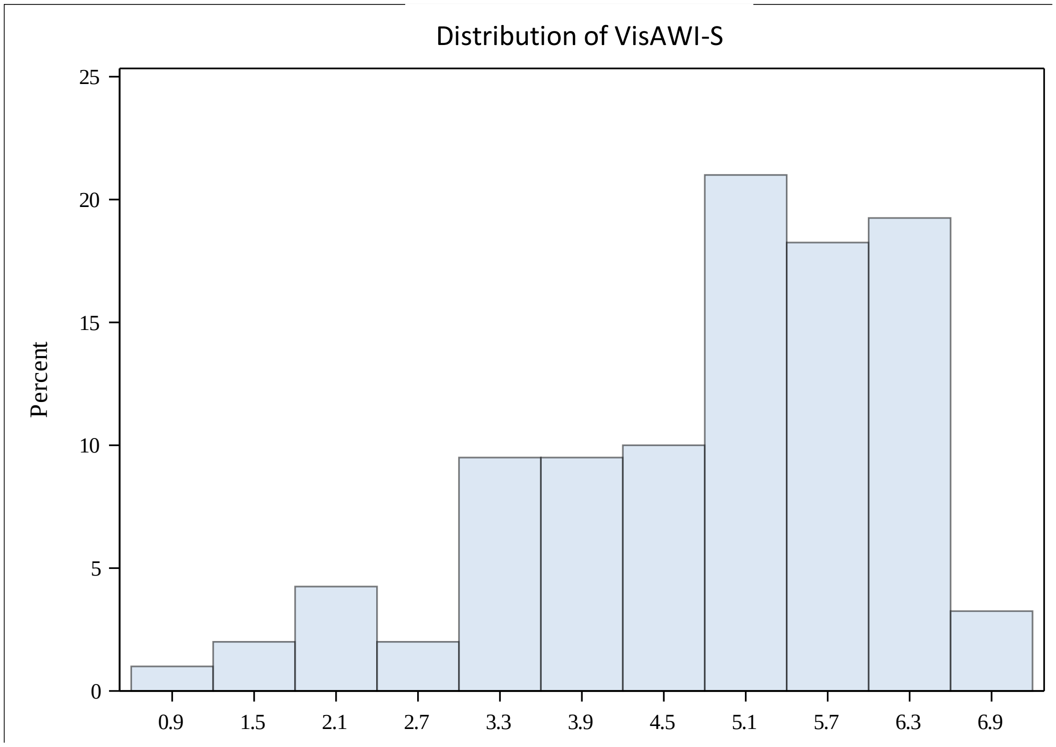

Figure 18. Distribution of VisAWI-S

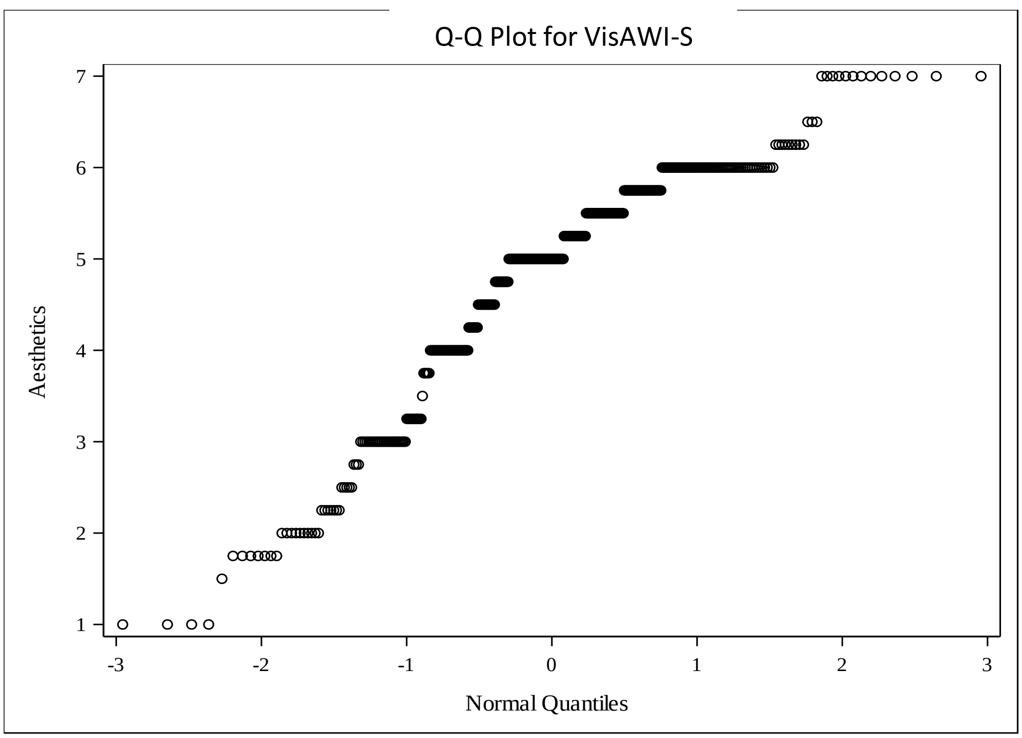

Figure 19. Q-Q Plot for VisAWI-S

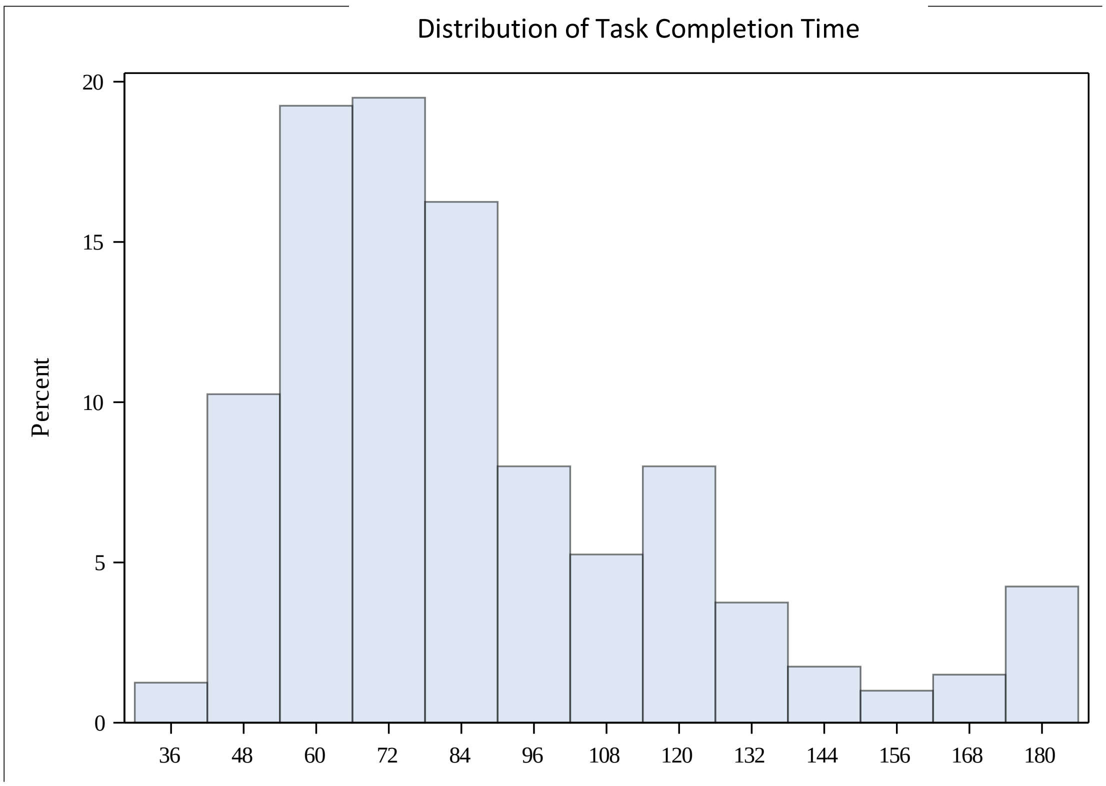

Figure 20. Distribution Plot of task completion time

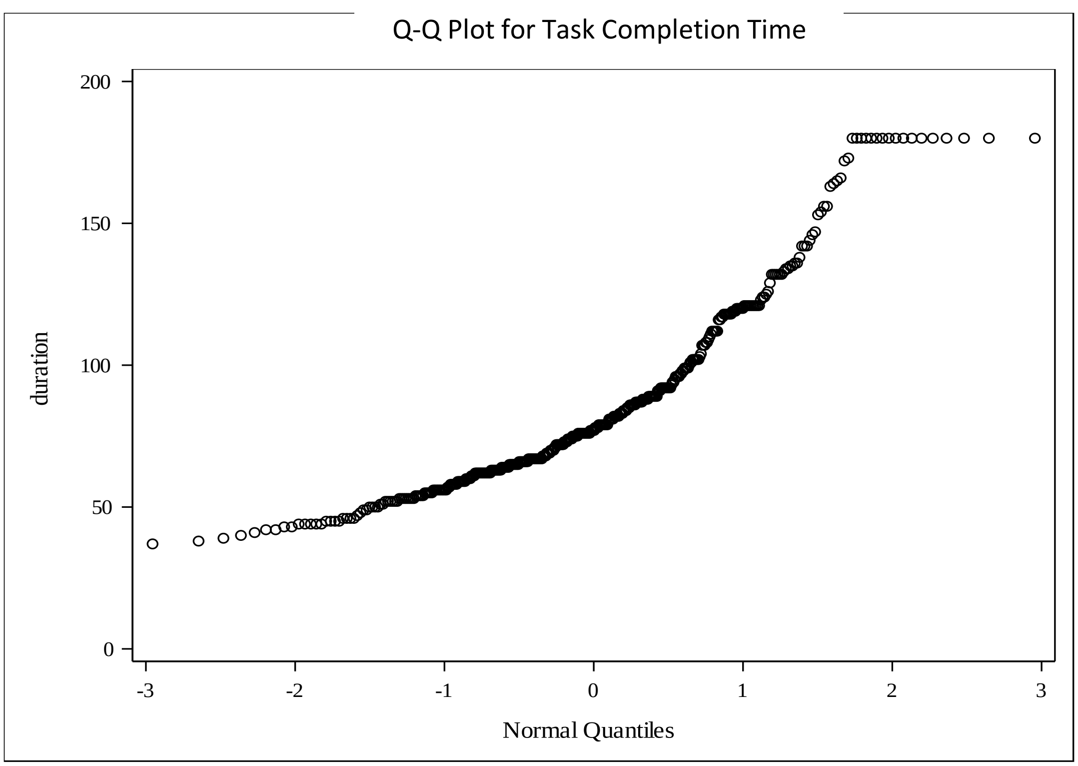

Figure 21. Q-Q Plot for task completion time

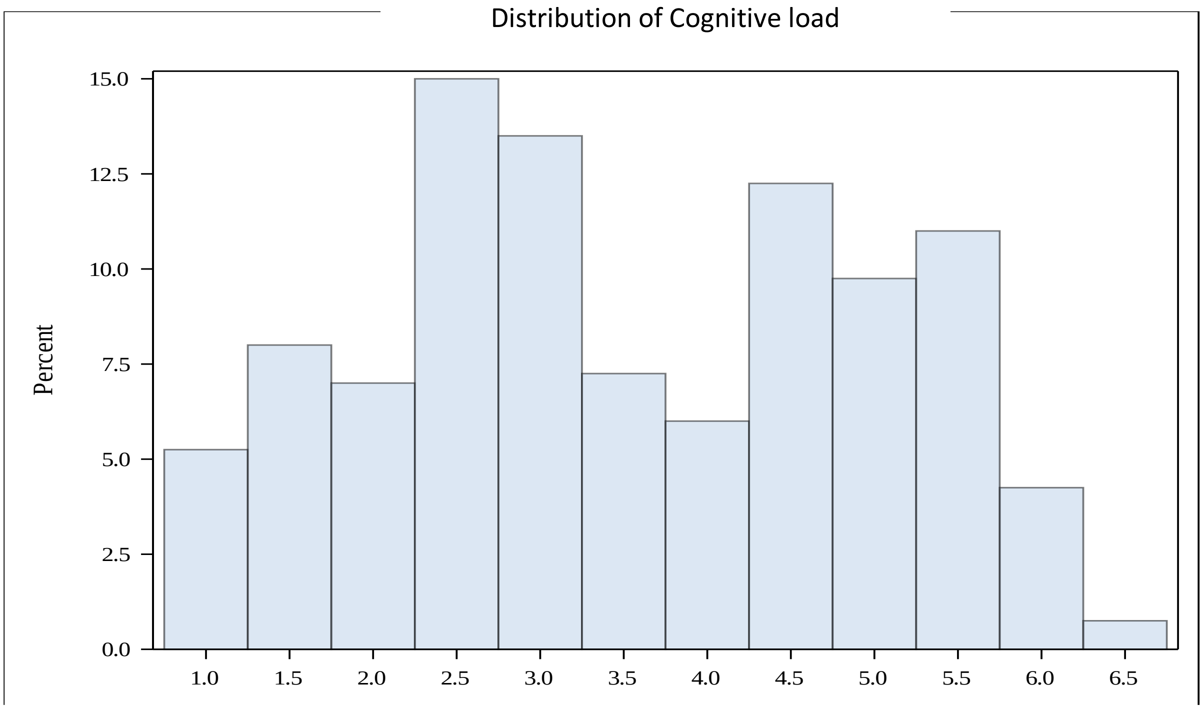

Figure 22. Distribution Plot of cognitive load

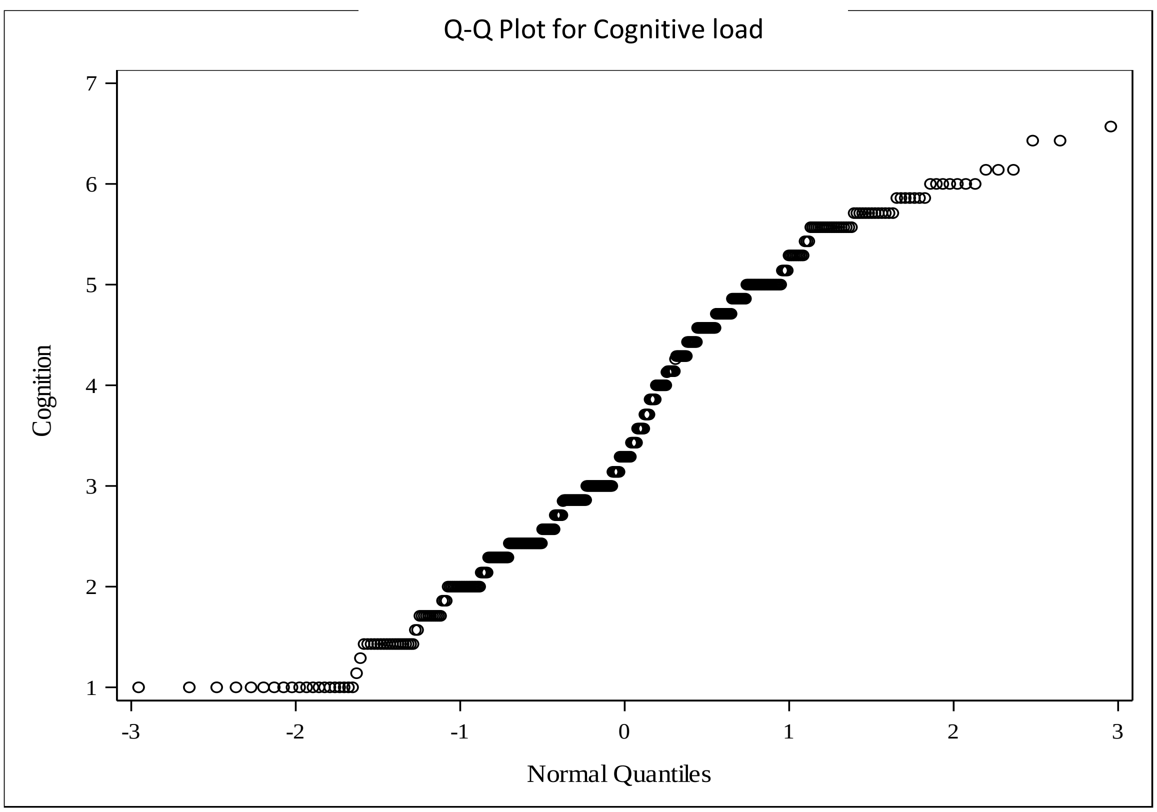

Figure 23. Q-Q Plot for Cognitive load

Appendix I: Principal Component Analysis of Task Performance

Table 20. Simple Statistics of task performance metrics

	Task Success Rate	Task Completion Time
M	0.86	86.58
SD	0.35	33.75

Table 21. Correlation matrix of task performance metrics

		Task Success Rate	Task Completion Time
M	Task Success Rate	0.86	86.58
SD	Task Completion Time	0.35	33.75

Table 22. Eigenvalues of task performance correlation matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	1.3583		0.6792	0.6792

Table 23. Eigenvectors of task performance

		Task Performance
Task Success Rate	Task Success Rate	0.707
Task Completion Time	Task Completion Time	-0.707

Appendix J: Demographic Information of Pre-test

Table 24. Demographic information of pre-test

	Options	Frequency
Gender	Male	6
	Female	4
Age	18-29	9
	30-39	0
	40-49	1
	50-	0
Education Background	Highschool	0
	Bachelor’s Degree	0
	Graduate Degree	9
	Post-Graduate Degree	1
	Others	0
Employment Status	Employed	0
	Studying	10
	Retired	0