



LOUVAIN
School of Management

EXCELLENCE & ETHICS IN BUSINESS

UNIVERSITE CATHOLIQUE DE LOUVAIN

LOUVAIN SCHOOL OF MANAGEMENT

**Analysis of the aesthetics of e-commerce web sites on mobile
devices**

Promotor: **Jean Vanderdonckt**

Co-promotor: **Nicolas Burny**

Research thesis presented by

Kirsch Antoine

As part of
the Master's degree
in Management.

ACADEMIC YEAR 2018-2019

Acknowledgements

Big thanks to all the people that follow me during the elaboration of this thesis. It was important to receive support in order to deliver a qualitative work.

First, M. Vanderdonckt help me to identify the purpose of the topic and guided me in my researches. His contribution was essential and I express my gratitude for his experience sharing. He also allows me to find the best methodology and the best tools to use.

I also thanks M. Burny for his help and especially the one concerning the experimental part. The elaboration of the experience was a key element for my work.

Finally, my family and friends also play an important role by giving my support and time. So thank you everybody.

Table of contents

Acknowledgements	
Table of contents	
List of figures	
1. Introduction	1
1.1 Problem statement	1
1.2 Research question in management	3
1.3 Research's motivation	4
1.3.1 Personal & academical motivations	4
1.3.2 Marketing motivations	5
1.3.3 Historical motivations	5
1.3.4 Management motivations & relationship with the economic environment.	5
1.3.5 Methodological motivations	6
1.4 Thesis structure	7
2. Literature review	8
2.1 Definition and evolution of aesthetics in history	8
2.1.1 Ergonomics and GUI	10
2.1.2 Aesthetics measurement.....	11
2.2 Graphical User Interface.....	14
2.2.1 Definition and history	14
2.2.2 Mobile device.....	15
2.2.3 Usability perception	17
2.3 Software quality product	18
2.3.1 ISO25000	18
2.3.2 Aesthetics-usability effect	19
2.4. Summary of the different used studies	20

3. Research's questions and hypothesis.	22
3.1. The context	22
3.2 Questions and hypothesis	24
4. Methodology	26
4.1. Research method.....	26
4.2. Data collection.....	27
5. Statistical analysis	38
5.1. Research Question 1: Do measures used by QuestIHM represent a good indicator of aesthetics of e-commerce websites on mobile device?.....	39
5.2. Research Question 2: Are aesthetics and popularity two linked concepts?	47
6. Results	56
7. Conclusions	57
7.1. Conclusions	57
7.2. Limits and problems	58
8. Bibliography.....	59

List of figures

Figure 1 - ISO/IEC 25010 – Software quality product	18
Figure 2 - Illustration of Whitefield Framework.....	22
Figure 3 - Illustration of some screenshots	27
Figure 4 - Illustration of the software QuestIHM.....	28
Figure 5 - Illustration of the zone definition in QuestIHM and metrics computation.....	29
Figure 6 – Balance formula	30
Figure 7 – Equilibrium formula	31
Figure 8 – Simplicity formula	31
Figure 9 – Density formula	32
Figure 10 – Horizontal balance formula	33
Figure 11 – Vertical balance formula.....	33
Figure 12 – Box plot of concentricity	38
Figure 13 – Table of coefficients of the multiple linear regression	40
Figure 14 – Table of adjusted R Square of the model.....	41
Figure 15 – Table of spread.....	42
Figure 16 – Effect size calculator illustration with saliency balance	43
Figure 17 - Table of the representation of the individual significance	45
Figure 18 - Table of the representation of the global significance.....	46
Figure 19 – Scatterplot graph of the mean score and popularity.....	47
Figure 20 – Descriptive analysis between Popularity and Score	48
Figure 21 – ANOVA test of popularity and score	49
Figure 22 – Mean plot of popularity and score	49
Figure 23 - Jonckheere-Terpstra test for score and popularity.....	50
Figure 24 – ANOVA test for every measure regarding popularity rank.....	52
Figure 25 – Example of mean plot (border density and simplicity regarding popularity rank).....	53
Figure 26 - Jonckheere-Terpstra test for border density, simplicity and popularity	55

1. Introduction

1.1 Problem statement

“In order to decide whether anything is beautiful or not, we refer the representation, not by the Understanding to the Object for cognition but, by the Imagination (perhaps in conjunction with the Understanding) to the subject, and its feeling of pleasure or pain. The judgement of taste is therefore not a judgement of cognition, and is consequently not logical but aesthetical, by which we understand that whose determining ground can be no other than subjective”.

This quote from Kant (1790)¹ implies that the judgment of taste, and therefore of aesthetics, is a phenomenon of a purely subjective character. It means that everyone's opinion is different about what is beautiful or what is not. However, the concept of aesthetics has become a key concept in the society and we see that businesses, as well as individuals, are eager to make every effort to please others and to draw the full attention to them. We can then ask ourselves the question of how to optimize the beauty or the design of someone or something if the factors of measurement of aesthetics are only purely subjective.

From an economic point of view, and more specifically marketing, companies use aesthetics to attract consumers. This strategy has always existed and, with the advent of the web and online shopping, this issue of beauty has kept its importance because the first goal when posting something on the internet is to attract attention. The web able companies to work on this component by offering the possibility to express their image and their vision via a website or e-shop, which is called Graphical User Interface (GUI) in technical terms. This is the subject on which we are interested in this thesis. Indeed, we will look at the aesthetic analysis of websites and more specifically e-commerce sites. This notion of aesthetics will therefore be fully related to the web and mobile devices and its ergonomic character, and not to the overall definition of aesthetics.

With the arrival of the web, companies have seen an opportunity for expansion in terms of profitability, customer database but also in terms of visibility around the world. E-commerce permits to intensify economic transactions, and this has been exponential since the beginning of

¹ Kant, I., (1790). *Critique of judgement*

the 2000s. Indeed, a chart proposed by Statista² shows us that the retail e-commerce sales in the world has a worth \$ 2304 billion in 2017 and could expect more than \$ 4878 billion in 2021.

In this sense, aesthetics is an important factor of potential competitive advantage for companies because the web is one of the main touchpoint between a company and the potential consumer. Indeed, it is an important component in terms of user's experience (UX) of a product. Aesthetics creates an attractiveness bias – it corresponds to the fact that in general, people are drawn to beautiful things. It also helps user to have a better usability. “Studies have found that visually appealing websites are rated as more usable than they actually are, because their attractiveness elicits pleasant emotions in users”³. We are also convinced that good aesthetics increase traffic and time on a site, decrease bounce rate and play on emotional perception.

According to Tractinsky and al.⁴, we can address the importance of aesthetics in the field of Human-Computer Interaction (HCI) from three different perspectives. The first one is the design perspective which implies that visual aesthetics is often related to other design aspects and that it is an important/integral part of any design discipline. The second one is the psychological perspective and is derived with three arguments : aesthetics is considered as an extension of the Self (people modify their devices, their screen as they want), it satisfies basic human needs and is a source of pleasure and, finally, aesthetics impressions are fast, enduring and consequential. We can add to this reflection of fast impressions a result coming from an experiment from Thomas Schmidt and Christian Wolf⁵. With this experiment, they conclude that intuitive aesthetics (impression after short presentation) and reflective aesthetics (impression after long presentation) are strongly correlate. That means that the quick judgement of people are often their definitive and right judgement. The last perspective linked to HCI is the practical one. From this, aesthetics is important because it allows to differentiate similar products and is already integrated into current socio-technical processes.

² Online on the website Statista. <https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/> (2018)

³ Online on the website Interaction design. <https://www.interaction-design.org/literature/topics/aesthetics> (2018)

⁴ Tractinsky and al. The Encyclopedia of Human-computer Interaction, 2nd Ed, visual aesthetics.

⁵ Schmidt, T., and Wolff, C. (2018). *The influence of user interface attributes on Aesthetics*.

In this thesis, we will measure the aesthetics of graphical user interfaces of a range of different popular e-commerce websites to test if the aesthetics plays a role in the popularity. We know that it is difficult for companies to optimize this component and that it represents a real challenge in term of marketing decision. That lead us to the next part which is called research question in management.

1.2. Research question in management

The Graphical User Interface has become a serious dilemma for marketers and companies globally. With the emergence of many researches about this topic, we are more certain than ever that it has a real impact on costumer perception, costumer pleasure and costumer decision making process. Visual aesthetics is one of the most important factors affecting consumer perception in many ways. It has a symbolic role that can affect evaluation of product (Mumcu and Kimzan, 2015, p.529)⁶. It is also a way to differentiate himself from other in a competitive market, because its practical perspective allows to differentiate similar products.

On the other hand, we know that the number of mobile devices is totally exploding all over the world. Indeed, more than 51% of the total web traffic is coming from mobile internet traffic⁷. Optimizing the GUI of its own website and its mobile-friendly side is definitely a challenge which can't be underestimated. It represents also a big opportunity in term of brand and corporate image. As Page and Herr said in their work, "Aesthetics can play a stronger role than brand name in terms of ratings and quality judgement" (Page and Herr, 2002, p.133-147)⁸.

From a management perspective, it seems important to ask the right questions to identify how to measure aesthetics on mobile device. Tractinsky, one of the main contributor to aesthetics researches, said that "What is beautiful is usable"⁹. This quote will be the starting point of this thesis. The problem is that the terms beautiful and usable in information systems have not a unique

⁶ Mumcu, Y., and Kimzan, H. (2015). *The effect of visual product aesthetics on consumer's price sensitivity*. Procedia Economics and Finance, 26, 528-534

⁷ Online on the website Statista. <https://www.statista.com/statistics/306528/share-of-mobile-internet-traffic-in-global-regions/> (2018)

⁸ Page, C., and Herr, P. M. (2002). *An investigation of the processes by which product design and brand strength interact to determine initial affect and quality judgment*. Journal of consumer psychology, 12, 133-147

⁹ Tractinsky N., Katz A.S., Ikar D (2000). *What is beautiful is usable*. Interacting with computers.

definition. Beautiful can obviously be interpreted in term of aesthetics, be also in term of preferences or attractivity. The usability of a system is evaluate regarding the effectiveness, the efficiency and the satisfaction that the system provides to an user.

In order to give answers and identify correlations between beautifulness and usability, we will analyze the aesthetics of popular e-commerce website based on software computation and user preferences rating. To do so, we will explain in the following sections of the thesis how those concepts (aesthetics, popularity and visual appealing rating) are evaluated and computed, and what can be their correlations if some links exist and can be justified.

These steps will allow to give a potential answer to the research question: **How to measure aesthetics of e-commerce web site on a mobile device?**

1.3. Research's motivation

1.3.1. Personal & academical motivations

Digital is becoming more and more influent in our society. It has taken a core role in a lot of disciplines and is not about to stop. One of the first to be touch by this revolution is the marketing. The way of doing marketing is not even the same as before and lots of technical notions and components must need to be taken into account, such as usability and more precisely aesthetics. Global user's interface is also becoming an interesting topic that must be mastered. It is important to develop this subject in order to lighten certain aspects and to permit to future marketers and human computer designers to better understand the issue. Indeed, not a lot of studies have been already done and it seems to me to be a good challenge to bring answers. Moreover, I am born during this digital revolution and feel very concerned for two reasons. First, everybody has a desktop and a mobile device. In 2015, there were more than 3,47 connected devices per person and this number could reach 6,58 in 2020¹⁰. Secondly, it seems logical that the world will become more and more digital and that it is necessary to do research in this direction.

¹⁰ Online on the website Statista. <https://www.statista.com/statistics/678739/forecast-on-connected-devices-per-person/> (2018)

1.3.2. Marketing motivations

As an economic and management student, I had cursus on a very large number of disciplines. The marketing side of my formation is the one I preferred and I really think that this topic related to Information systems can have implications in marketing. On the one hand, we know that the website, and in this case GUI, is one of the first tool used by company to share their value, their mission and to make their business. It is also a tool that give them a lot of visibility. Aesthetics can have a big impact on companies and on their strategies and I'm curious to know how. On the other hand, I'm convinced that this kind of study can give the opportunity to better understand what is important or not in term of website aesthetics. As Shaoolian G. said, "a website in itself is a marketing tool, and if it is designed and developed correctly it can be a very powerful one".¹¹

1.3.3. Historical motivations

For a long time, the concept of aesthetics have been associated to judgement of beauty and subjectivism. It is thus very difficult to assess values and evaluation to something which is perceived as not measurable. Indeed, subjectivity and human computer interaction are not linked concepts. With the arrival of the web, it has taken another dimension and its objective character has been developed. It is now possible to evaluate it, and by definition, evaluation is the computation of quantifiable results. Many studies and reports were implemented but in small quantities. Graphical User Interface and usability-aesthetics are emerging concepts that need to be develop.

1.3.4. Management motivations & relationship with the economic environment.

Nowadays, companies need to quickly adapt their position and strategy in a fast-growing market. The main reason of this fast growth is the digitalization, and an important element in the digitalization is the graphical use interface. The technology is growing and moving very quickly and companies need to stay on the cutting edge of technology if they want to stay competitive in the market. Moreover, in this case, GUI of website (first façade on a company) is a real challenge. On the counterpart, lots of company don't have the skills to understand and integrate technical skills in their business. That's why the elaboration of new studies seems to be interesting.

¹¹ Shaoolian, G. (2017). *The Value of a Website Strategy*. Online on the website Huffington post. https://www.huffingtonpost.com/gabriel-shaoolian/the-value-of-a-website-strategy_b_8807554.html?guccounter=1

1.3.5. Methodological motivations

Few researchers try to develop their own methodology in the domain of aesthetics evaluation. Therefore, there isn't a unique specific methodology to follow due to the limited number of studies in the area. "Methodology is the systematic, theoretical analysis of the methods applied to a field of study. It comprises the theoretical analysis of the body of methods and principles associated with a branch of knowledge"¹²

Developing a methodology of work in aesthetics seems to be a good challenge. I haven't enough skills to pretend to develop my own methodology but it could be interesting to see how I could learn and improve the existing ones.

¹² Igwenagu Ch. (2016). *Fundamentals of research methodology and data collection*.

1.4. Thesis structure

Before going further, it is essential to describe the structure of the thesis in order to have a better overview of the different parts. “The adoption of a well-defined structure contributes to the clarity of the document, allowing the reader to navigate naturally and navigate easily”¹³.

The first part is called “Introduction”. This part is dedicated to the setting of the context of the topic. It quickly explains what will the thesis speak about and express my motivations to write it. The goal is also to introduce the subject from an economic point of view and to attract the attention of the readers.

The second part is the literature review. In this one, every relevant content is gathered and summarized and we explain the different elements of the research question based on existing knowledges and references. It explains each component that is related to the research question (aesthetics, GUI, mobile device,...) in order to ease the reading and understanding of the thesis.

The next step (next part) is to explain research questions and the hypotheses. Those elements are the backbone of the statistical analysis.

The fourth part is dedicated to the explanation of the methodology that we have followed to perform the experience. This part will answer questions like “what is the audience? Which platform will they use? Which statistical program and methods will we use?”.

The following part explains the statistical analysis and will be followed by the sixth part that is dedicated to the results of the analysis.

Finally, the last part summarizes all the results, conclusions, as well as the limits of the study.

¹³ Padua A.(2003), *Structure d'un document scientifique*.

2. Literature review

Before entering in the practical and experimental part, it seems essential to explain the context and tell the story of the different key concepts on which we will work. To do so, we will make a literature review of many definitions and studies about specific subjects. Obviously, the first to be taken into account is the notion of aesthetics. We will see how this notion has evolving from its origin to today and has taken a different role depending of the context. The three main questions to be asked are: what does really mean aesthetics? how is it considered in term of ergonomics and GUI? and, how aesthetics is measured? That leads us to the second concept to define, which is the Graphical User Interface. We will also discuss of aesthetics in term of mobile device, usability and user experience as well as its role in the context of Software quality product.

2.1. Definition and evolution of aesthetics in history

The word aesthetics finds its origin in the 18th Century with the Greek word “aisthetikos”. It is related to “sense perception, perceptible, sensitive”¹⁴. But it has first be taken into account by Plato and Aristotle who felt that beautiful objects were aesthetically appealing. The word has then been used by Baumgarten in Germany which interpreted it as a “criticism of taste” (1750). Then, Kant refers to aesthetics as a “science which treats of the conditions of sensuous perception”. The English translations of Kant have permit to popularize the word. Those two philosophers see the concept from a distinct perspective. From the one hand, Baumgarten recognized it as something based on personal feeling. On the other hand, Kant claimed that aesthetics corresponds to every sensual perception, in his popular paper Critique of Judgement.

We can see that the first discipline where aesthetics was considered was the philosophy. It is even considered as the philosophy of art. In this context, it is associated to “the nature of art, beauty, and taste, with the creation and appreciation of beauty”¹⁵. Aesthetics is rapidly linked to the notion of subjectivity and to judgement of sentiment and taste¹⁶. For a long time, the notion is related to the perception of beauty and we know that everybody has a different judgement

¹⁴ Online of the website etymonline.com. <https://www.etymonline.com/word/aesthetic> (2018)

¹⁵ "Aesthetics". Merriam-Webster Dictionary. (2018)

¹⁶ Zangwill, Nick. (2003-2007). *Aesthetic Judgement*, Stanford Encyclopedia of Philosophy,. Retrieved 2008.

and perceptions about it. At this stage, aesthetics is considered as the study of beauty. In some others English studies, it is also called “standards of taste” or “judgements of taste” (David Hume – Scottish philosopher). Aesthetics has then taken lots of different forms. We can read topics about aesthetics in the field of post-modernity, psychoanalysis, mathematics, scientific and so on.

The definition of the term itself has also evolved throughout history. Here are some different definitions of aesthetics in chronological order:

In his paper *Metaphysics* (1739)¹⁷, Baumgarten associated aesthetics and taste, and defined it “as the ability to judge according to the senses, instead of according to the intellect, based on feelings of pleasure or displeasure”. He was the pioneer of the concept.

In the context of Art, Heidegger defined aesthetics as “the science of man's sensitive and emotional behavior and what determines him”.

Many others defined the concept. Clay (1908) claimed that aesthetics is the criterion by which beautiful things are to be judged. Van Damme (1996) said that it is a standard of beauty. Lavie and Tractinsky (2004) spoke about aesthetics user's perception and divided it into two dimensions: classical aesthetics and expressive aesthetics.

Nowadays, the common definition of the noun aesthetics is “a branch of philosophy concerned with the study of the idea of beauty”¹⁸.

The notion of aesthetics is concerning a lot of disciplines and is present in lots of domains. Obviously, it touches the painting, the sculpture and photography. It is also omnipresent in architecture and thus also in the indoor and outdoor design. With the arrival of internet, it has started to be taken into account in informatics and more specifically in Information Systems. Designers of information systems need to be more aware of that concept but also need to understand what it means. This new concept is maybe difficult to master for designers and some authors try to provide guidelines to help them. Haynes J. and Paradice D. (2007)¹⁹ wrote a paper concerning aesthetics and Information systems. They suggest that one of the first step to better understand the concept is to go from a passive approach (which is related to definitions and perceptions) to a “vital

¹⁷ Baumgarten, A. (1739). *Metaphysics a Critical Translation With Kant's Elucidations, Selected Notes and Related Materials*.

¹⁸ Online on the web site Collins Dictionary. <https://www.collinsdictionary.com/dictionary/english/aesthetics> (2018)

¹⁹ Haynes J., Paradice D. (June 2007). *Information technology and aesthetics : passive and active dimensions*. Australasian Journal of Information Systems, Volume 14, Number 2.

living experimental open thematic and conceived active approach” (oriented around comparison of previous and recent experiment). They also pretend that informatic designers saw aesthetics only as a need to improve “look & feel” on an intranet environment or in other words improve aesthetics design. But, according to their study and findings, “the real capacity for improving a “look and feel” in an IT environment comes from a thematic ability to connect who looks and feels with what is looked at and felt”.

The most relevant definition of aesthetics for us is the one related to design. In the field, aesthetics refers to the visual attractiveness of a product and have an impact on user experience in several ways: it creates attractiveness bias and solve some issues about usability.²⁰ In our context, the product is the GUI of the mobile device and the user experience is the relation between this device and the users. That definition brings us to the next part which concerns ergonomics and GUI.

2.1.1 Ergonomics and GUI

What make sense here is the notion of aesthetics in term of usability characteristics. The term is no longer simply associated with a judgment of what is beautiful or not, it comes closer to a factor that affects a user's browsing quality and the comfort he feels when consulting a web page, called User Experience (UX). The user experience is linked to usability and corresponds to the quality of interaction between a user and a product or a system with an interface. But UX and usability are different: “Usability is usually considered the ability of the user to use the thing to carry out a task successfully, whereas user experience takes a broader view, looking at the individual’s entire interaction with the thing, as well as the thoughts, feelings and perceptions that result from that interaction”²¹

In this thesis, it is essential to see aesthetics from an ergonomics point of view. Ergonomics refers to efficiency and usability of a system. In this case, the ergonomics of a website aims to effectively meet the expectations of users and provide them with a comfortable navigation. In this sense, aesthetics is closely linked to usability which correspond to ease-of -learning and ease-of-use. Scientifically, it is defined as “the effectiveness, efficiency and satisfaction with which a

²⁰ Online on the website Interaction design. <https://www.interaction-design.org/literature/topics/aesthetics> (2018)

²¹ Albert, W. and Tullis, T. (2013). *Measuring the User Experience 2nd edition: Collecting, Analyzing, and Presenting*.

product is used” (ISO 9241-11). However, the two concepts must have a different approach. While usability focuses on design, aesthetics is a tool to improve user experience, satisfaction and social acceptance.

Concerning the user interface, it is important to notice that the aesthetics must become a key element to focus on. It can be a tool to facilitate communication between someone and a device. More and more studies treat about this subject and it looks like a key aspect in the website designing for many reasons. (i.e. ease of use, ease of learning, ease of communication, user satisfaction and so on). Now that people are aware of the importance of the concept, it has become relevant to know how we can measure aesthetics of an information system and more precisely of a website.

2.1.2 Aesthetics measurement

Now that people are aware of the importance of the concept, it has become relevant to know how we can measure aesthetics of an information system and more precisely of a website. We'll see in the chapter what are the different possibilities to measure the aesthetics of a Graphical User Interface. As an integral part of the process of user experience and user satisfaction, it has been crucial to make research about this aspect. We know that the concept of aesthetics is confusing because it is not a purely objective element, but we will see that many researchers try to implement some systems to objectively evaluate it.

Firstly, two schools of thinking exist about the conception of aesthetics. The first one is the subjectivism, which claims that an object haven't any aesthetics characteristics. Every aesthetics aspects come from the perceiver and are related to his preference and his cultural background. The second way to see the concept is the objectivism. From this point of view, the aesthetics of an object come from this object directly. The evaluation of this one is thus based on its characteristics and on our knowledge and judgement.

In the field of user interface, the subjectivism approach is not relevant. Indeed, to evaluate something, it is required to have some criteria and characteristics. It is therefore difficult to know which factor we can take into account and how intense this factor will be in the aesthetic evaluation of something. The ultimate goal is to find a way to evaluate the aesthetics based on an objective quantitative method mixed with a human qualitative evaluation. Some researchers try to identify

the different components of aesthetics in order to evaluate and measure it. A good example is the experiment led by Tuch A. and al (2012)²². They have measured users' preferences of aesthetics of websites based on two components, the visual complexity and the prototypicality (which is "the amount to which an object is representative of a class of objects" (Leder et al., 2004, p. 496)²³). The results are clear: those elements play a crucial role in aesthetics assessment. Users prefer a website with low visual complexity and high prototypicality.

Several other researchers address the issue with a different approach and decide to build a framework comprising a set of components of aesthetics. One of the first, Ngo D., Teo L. and Byrne J. (2000)²⁴, try to propose a framework to analyze graphic displays. They explained and demonstrated a set of fourteen variables that can be used to measure objectively the aesthetics of a web interface. Those variables are: balance, equilibrium, symmetry, sequence, cohesion, unity, proportion, simplicity, density, regularity, economy, homogeneity, rhythm, and order and complexity. Based on those elements, they could try to define if a website is aesthetically appealing or not.

Later, Khalighy S. and al. (2014)²⁵ suggests dividing aesthetics in two categories: beauty and attractiveness. For them, "beauty is a certain balance between contrast, proportion, and pureness. And attractiveness is a certain balance between novelty and appropriateness". By putting together those qualities (contrast, proportion, pureness, novelty and appropriateness) and design elements²⁶ (form, detail, material, texture, color and composition), they have identified more than 30 measures that can be used in the evaluation of an interface's aesthetics.

More recently (2018), few Finnish researchers and students have implemented a new interface to test the usability and aesthetics of websites. They called it AIM. "AIM is an online service and an open code repository for computational evaluation of graphical user interface (GUI) designs"²⁷. The operation of the interface is easy: enter the webpage URL or upload a screenshot,

²² Tuch A., and Al. (2012). *The role of visual complexity and prototypicality regarding first impression of websites: Working towards understanding aesthetic judgments.*

²³ Leder and AL. (2004). A model of aesthetic appreciation and aesthetic judgments.

²⁴ Ngo D. C. L., Teo L. S., Byrne J. G. (2000). *A Mathematical Theory of Interface Aesthetics*. Visual Mathematics 10.8

²⁵ Khalighy S. and Al. (2014). *Measuring aesthetics in design*. International design conference.

²⁶ Kostellow, R. R. (2002). *Elements of design*, Princeton Architectural New York NY.

²⁷ Oulasvirta A. and Al. (2018). *Aalto Interface Measures (AIM): A Service and Codebase for Computational GUI Evaluation*. Aalto University.

select the desired measures and the program analyzes the page and then presents results via explanations and histograms.

In the practical part of this thesis, we use the software QUESTIHM. This software has been designed by Zen M. in order to compute the aesthetics of a GUI by using a set of relevant and validated variables. When realizing this program, he proposes to fragment the GUI in several regions and to evaluate them with a semi-automated process. To do so he identifies measures that could be relevant to measure aesthetics. A metric is “an attempt to quantify a visual technique characterizing the arrangement of the layout frame components in a UI”²⁸. Below, the existing definition of the different metrics.

- Balance: according to Ling Ngo, Seng Teo and G.Byrne²⁹, the balance refers to the distribution of optical weight. It is related to the perception of some objects in the picture. It takes into account the shape, color, size of the object.
- Equilibrium: according to Ling Ngo, Seng Teo and G.Byrne, the equilibrium refers to the distribution of the objects on the picture. According to the authors, the way to obtain the equilibrium is to centralize the layout, the disposition of the objects. It can be also defined as the difference between the center of mass of the object and the center of the screen.
- Simplicity: according to Ling Ngo, Seng Teo and G. Byrne³⁰, simplicity refers to the directness and the singleness of form. The simplicity can be increased by optimizing the number of objects on a screen and reduce the alignment.
- Symmetry: it analyzes the replication of one side of the screen to the other side. It allows us to know if one side of the screen is exactly replicated to the other side.
- Density: according to Ngo, Seng Teo and G.Byrne³¹, the density is defined as the number of objects covering the screen. Tully defined the measure of density; it is the percentage of character position on the frame containing data.
- Horizontal Balance: it is defined as the balanced disposition about same elements on a horizontal axis. The comparison is computed between the higher side and the lower side.

²⁸ Zen, Mathieu ; Vanderdonckt, Jean. *Towards an evaluation of graphical user interfaces aesthetics based on metrics*. IEEE Eighth International Conference on Research Challenges in Information Science (RCIS'2014).

²⁹ Chek Ling Ngo, D., Seng Teo, L., Byrne, G.A *Mathematical Theory of Interface Aesthetics*. (2000)

³⁰ Chek Ling Ngo, D., Seng Teo, L., Byrne, G.A *Mathematical Theory of Interface Aesthetics*. (2000)

³¹ Chek Ling Ngo, D., Seng Teo, L., Byrne, G.A *Mathematical Theory of Interface Aesthetics*. (2000)

- Vertical Balance: it refers to the balanced disposition about same objects on a vertical axis. The analysis of the replication is computed from the left to the right or the opposite.
- Center Alignment: it is defined as the way the elements are dispatched around the center of the screen. More the object will be centralized more the center alignment rating will be high.
- External Alignment: it refers to the distance of the disposition of the elements from the border of the screen. Less element near to the center, more the external alignment rating will be high.
- Saliency Balance: according to Shen and Zhao (2014, quoted in M. Zen and J. Vanderdonckt lecture)³², this variable refers to the computation of the balance according to the saliency. The saliency refers to something which is prominent.

In this thesis, the term “measure” will replace “metric” because metric can be interpret in different ways. To understand the measurement of the aesthetics of a GUI, we have first to understand what does GUI mean.

2.2 Graphical User Interface

2.2.1 Definition and history

This work mentioned several times the expression Graphical User Interface, which a is fundamental concept in our analysis. A simple definition of the GUI is all means, all possibilities than someone have to interact with an electronic device.

From a professional and recent point of view, Zen M. and Vanderdonckt explain it as “A visual platform representing a way for humans - i.e. users - to interact with a device. It consists generally on a set of graphical items (windows, buttons, menus, etc.) with which users interact through the use of one or several modalities such as mouse, keyboard, touchscreen, voice recognition system, etc.”³³. But it is not the only definition that we can find about GUI. We can have in some articles definition such as “means by which people and computers communicate with each other’s (Bonsiepe, 1990) or “mechanisms by which users can interact with computer program”

³² Vanderdonckt, J., Zen, M. (2016) *Assessing User Interface Aesthetics based on the Intersubjectivity of Judgment*, 12.

³³ Zen, Mathieu ; Vanderdonckt, Jean. *Towards an evaluation of graphical user interfaces aesthetics based on metrics*. IEEE Eighth International Conference on Research Challenges in Information Science (RCIS'2014).

(Van Hoff, 1998). The words utilized are not the same, but the purpose of the definitions remain nearly the same.

Historically, the first apparition of a form of GUI was in 1962 when Douglas Engelbart made research on human computer interaction. His first experiment was composed of a television (screen) and a cursor (mouse) which were called together On Line system³⁴. But it wasn't yet considered as GUI. The first application with a GUI was designed in 1977 at the Xerox Palo Alto Research Center. After that, Steve Jobs had the intention of developing its own GUI, called Apple Lisa, with the help of other designers. Similarly, to Xerox Star, Apple Lisa has encountered difficulties and can't be commercialized. We must wait the year 1984 to see the apparition of the Apple Macintosh which has become a worldwide reference in term of GUI.³⁵ The Macintosh interface is nowadays worldwide known in the development of desktops, but it is also one of the main reason of the development and proliferation of mobile device in the world.

2.2.2 Mobile device

The total number of mobile devices has totally exploded since last decade. According to Statista³⁶, more than 63.4% of the worldwide population will use their mobile phone on internet and 17.8% will use a tablet in 2019. It seems essential to know and learn how GUI can be applied to mobile system. There are not a lot of studies about this topic but by commuting few of them we can have a good overview about the GUI on mobile device and its potential impact on consumer.

Firstly, it is important to know what aesthetics means in term of mobile device. Aesthetics could be associated to the physical aspect of a mobile. In this sense, it corresponds to features such as the size of the screen, the quality of materials used for the screen and the casings, but also the different forms of themes. On the other side, and what is relevant in this work, is the aesthetics related to the interface of a mobile. Here, it is more technical and make reference to the organization of the interface. By organization, we mean every component that form the graphical interface, such

³⁴ Doug Engelbart Institute. *The Dawn of Interactive Computing*. Online on the website Doug Engelbart. <http://dougengelbart.org/firsts/interactive-computing.html> (2018)

³⁵ Jansen, B. J. (1998). The Graphical User Interface: An Introduction . SIGCHI Bulletin. 30(2), 22-26

³⁶ Online on the website Statista. <https://www.statista.com/statistics/219909/global-tablet-penetration-forecast/> & <https://www.statista.com/statistics/284202/mobile-phone-internet-user-penetration-worldwide/> (2018)

as the colors, the density of information available on the screen, the visual appealing and so on. In other words, it corresponds to any features that make an interface pleasant and satisfying. Color, for example, is a key factor. It plays an affective role and influence users regarding their preferences, but it has also a functional role by being considered as “an organizer of information and contributor to a clearer overall impression of the UI”³⁷.

Soui M. and Al. (2017)³⁸, researchers in computer sciences, identify that the user interface of mobile application, called Mobile User Interfaces (MUIs), is an important factor concerning user satisfaction and effectiveness. They also affirm that there is a need to develop an evaluator of MUIs to decrease the maintenance costs of mobile applications and to enhance the quality of MUIs by identifying usability defects. The first step in their work was to note that MUIs has a more limited number of characteristics compared to desktop GUI. So, they thought to design a tool that combines the use of existing measures and that take into account the characteristics of mobile interface. This tool is called “PLAIN (PLugin for predicting the usAbility of mobile user INterface). It is an Eclipse plug-in that detects the quality problems of MUIs in early stages of software development, based on the above-defined usability measures.” To make it works, they include measures divided in two categories: guidance and coherence. “User guidance refers to the means available to advise, orient, inform, interact, and guide the users throughout their interaction with the computer. Coherence supplies how good the interaction between users and mobile user interfaces is and secures the efficient use of the MUI.” They finally test the effectiveness of the plugin by evaluating four android applications and the results show that it is effective.

We also know that people judge in the same way the GUI of a widescreen desktop and the GUI of a mobile application thanks to the study of Miniukovich A. and De Angeli A. (2014)³⁹. The impressions and judgement are made very quickly and reliably. Moreover, the people associate the aesthetics to the visual appearance. It induces that we can reliably predict the visual quality of a mobile app GUIs.

With those studies, people are aware that the aesthetics on mobile device is also a key concept. It represents a real challenge and it could be very interesting to know in which way it

³⁷ Silvennoinen J, Vogel M and Kujala S. (2014). *Experiencing Visual Usability and Aesthetics in Two Mobile Application Contexts*. Journal of usability studies, vol. 10, pp 46-62.

³⁸ Soui M., Chouchane M., Gasmi I. and Mkaouer M. (2017). PLAIN: PLugin for predicting the usAbility of Mobile User INterface. In Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 1.

³⁹ Miniukovich A. and De Angeli A. (2014), *Visual impressions of mobile app interfaces*. Conference: NordiCHI.

impacts the consumers. The challenge is to interpret the impressions of people when using a website and to know how those impressions can be translated into website value. One important study about that question is the one driven by Ming Li and Yung-Shao Yeh (2010)⁴⁰. They test few hypotheses in order to know if the aesthetics can have an impact on the perception of trust with customers and ease of use of website. The results were pretty relevant and show that good aesthetics increases the usability of e-commerce website and the trust that people have against the mobile website. That kind of study highlights the importance of the GUI and the perception that people have thought it. This perception is called Usability perception and is the topic of the next part. It could be interesting to perform again this study with a larger panel of website in order to identify if this conclusion can be generalized.

2.2.3 Usability perception

With those experiments, we know that the usability and aesthetics are two linked concepts. For casual consumers, a beautiful and elegant webpage is easy and nice to use. This phenomenon is called the usability perception. According to Tractinsky and Al.⁴¹, it exists a strong correlation between a system's perceived aesthetics and perceived usability. It induces that people associate aesthetics and usability by judging the physical appearance of a computer interface without using it. It shows that the physical attractiveness can give other value perception to something. On the basis of this study, other researchers⁴² have tested several hypotheses concerning the impact of aesthetics, and more precisely the color harmony, on people perception of a webpage quality. The results have shown that the evaluation of pragmatic qualities (effectiveness and efficiency) is lower when the colors of the site are disharmonic and there is no difference of evaluation between the two designs in terms of perceived usability. That being said, there should also be a relationship between aesthetics and the real usability, because aesthetics, from a technical perspective, is more than just appearance. Indeed, aesthetics has been recognized as a quality factor by the ISO standard, which is the subject of the next section.

⁴⁰ Yung-Ming, L., Yung-Shao, Y. (2010). Increasing trust in mobile commerce through design aesthetics. *Computers in Human Behavior*, 26 (4), 673-684.

⁴¹ Tractinsky N., Katz A.S., Ikar D (2000). *What is beautiful is usable*. *Interacting with computers*, 127-145.

⁴² Venny J. (2017). *Étude de l'effet de l'esthétique sur l'utilisabilité d'une interface lors d'une tâche de recherche d'informations sur un site Internet*.

2.3 Software quality product

2.3.1 ISO25000

Aesthetics has been recognized as a quality factor by an ISO standard. But, what does it mean? Product quality and process quality became very important aspects of software development. In this way, there was a need to develop a system which comprises a series of standards that will guide the software developers. That's why the standards ISO 25000 has been developed. This model is just the evolution of previous standards which were used to evaluate software quality. The first version (ISO/IEC 9126) was made in 1991, and has been updated in 2001, to finally become ISO 25000 in 2011. It gathers information and gives specifications about the quality requirements and evaluation of a system. From this, the product quality model has been designed. It is composed of eight characteristics, divided in sub-categories, that is used to evaluate the quality of a system: functional suitability, performance efficiency, computability, usability, reliability, security, maintainability and portability.



Figure 1 - ISO/IEC 25010 – Software quality product

What make sense in this work is that aesthetics is taken into account in the standard of quality evaluation. It is classified in the characteristic “Usability”. That means that aesthetics is a factor influencing the real usability and not only the perceived usability of a graphical user interface. User interface aesthetics is defined as “the degree to which a user interface enables pleasing and satisfying interaction for the user”.⁴³

⁴³ Online on the website Iso25000.com. <https://iso25000.com/index.php/en/iso-25000-standards/iso-25010?limit=3&start=3> (2018)

2.3.2 Aesthetics-usability effect

To make the link between usability and aesthetics, it is important to know in which way the aesthetics impacts the usability perception of users. We can deduce that it increases the visual appealing of an interface, and therefore gives a pleasant aspect to the navigation. One could also imagine that a relatively aesthetic interface facilitates consumer choice by an impression of ease of use.

Those hypotheses have been tested in several experiments, and globally people are less aware of usability problems of a site if they find this one visually appealing. This impression is called aesthetic-usability effect. It is defining as “users’ tendency to perceive attractive products as more usable. People tend to believe that things that look better will work better — even if they aren’t actually more effective or efficient”⁴⁴. It is not a new concept, this effect has been first studied in 1995 and the results were very similar. It is also a reference to the perceived usability identified by Tractinsky. All that things being said, we need to remind that this perception is true when usability problems are small. If anybody encounter lots of difficulties of navigation on a website, a pleasant interface wouldn’t solve any issues. De Angeli A. and al. (2006)⁴⁵ conducted a study to identify how preference of users are influenced by interaction, usability and aesthetics. The experiment revealed that “perception of information quality is affected by the interaction style implemented in the interface, in a manner resembling the halo effect in person perception”.

Now that theoretical concepts have been defined, we can start to answer to the main question in the next section called “research’s questions and hypothesis”.

⁴⁴ Moran K. (2017), *The aesthetics-usability effect*. Nielsen Norman group article.

⁴⁵ De Angeli A. and Al. (2006). *Interaction, usability and aesthetics: what influences users' preferences?* DIS '06 Proceedings of the 6th conference on Designing Interactive systems, pages 271-280

2.4. Summary of the different used studies

	Year	Participants	Stimuli	Dependent variable	Independent variable	Results
Khalighy S. and Al.	2014	50	1 group of image with pictures alphabetically labelled and 1 group of image with a single product in each image	Image preference	The interaction of beauty and attractiveness with each other and reaction of people to the different applications of aesthetic qualities	Revealed the potential of measuring the qualities of aesthetic via an accurate and high performance tool.
Soui M., and Al.	2017	User rating extracted from the Google Play Store	4 mobile applications (Duolingo, Accuweather, loan calculator and HandicraftWomen)	Quality of the applications evaluated by the PLAIN application	User rating	The PLAIN tool can be used to predict the defects of MUIs in early stages of software development and evaluate quality of applications.
Tractinsky N., Katz A.S., Ikar D	2000	132 third year Industrial Engineering students	9 ATM layouts	Post experimental rating (aesthetics, ease of use and amount of information on the screen + user' satisfaction)	Pre experimental rating (aesthetics, ease of use and amount of information on the screen)	It exists a strong correlation between a system's perceived aesthetics and perceived usability.
Yung-Ming, L., Yung-Shao,Y.	2010	200 responses	A virtual m-vendor (camera task)	Trust, perceived usefulness, customization, design aesthetics, perceived ease-of-use	Website browsing and decision making process	Good aesthetics increases the usability of e-commerce website and the trust that people have against the mobile website.

	Year	Participants	Stimuli	Dependent variable	Independent variable	Results
Tuch A., and Al.	2012	59	119 screenshots of real websites with different presentation time	Perceived beauty	Visual complexity+high prototypicality and presentation time	Users prefer low visual complexity and high prototypicality.
Miniukovich A. and De Angeli A.	2014	17 students	3 screenshots for 17 free mobile apps (51 screenshots) with interface controls and that cover the entire screen	Score rated by participants	The exposure duration as a within-subject independent variable (50ms VS 4s)	The immediate-impression principles from the Web domain also function in the mobile app domain. Participants did not alter their evaluations between the 50ms and 4s conditions.
De Angeli A. and Al.	2006	28 undergraduate and postgraduate students (23 male and 5 female)	A live website	The usability evaluation, aesthetic appreciation, information quality, engagement, the memorability test	Post experiment participants' overall preference	The link between aesthetics and usability reported by Tractinsky's studies, is more complex than "what is usable is beautiful". Participants evaluated an interface as more attractive on the expressive aesthetics dimension. Perception of information quality is affected by the interaction style implemented in the interface.

3. Research's questions and hypothesis.

3.1. The context

Before going further, it is essential to explain the context in which we will work. We will make the use of the classification framework of Whitefield. This framework allows to differentiate experiences (human factors evaluation) regarding the representation (real or representational) of the computer and the representation (real or representational) of the user. “Applied to the computer component of the system, *real* here means the physical presence of the computer or an approximation of it. On the other hand, specifications and notational models are *representational* computer presences, as are users' mental representations of the computer. Applied to the user component of the system, *real* similarly means actual users or approximations of them. In contrast, the presence of *representational* users means (explicit or implicit) descriptions or models of the users.”⁴⁶

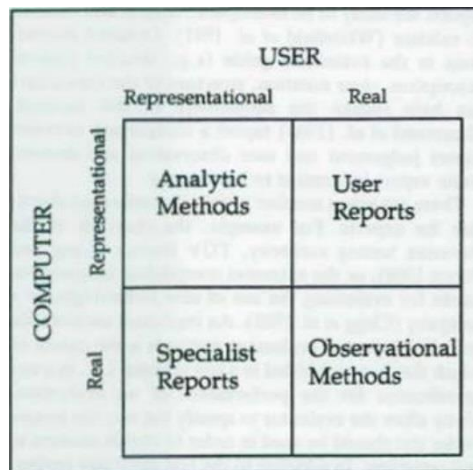


Figure 2 - Illustration of Whitefield Framework

⁴⁶ Whitefield A. and al. (1991). A framework for human factors evaluation, Behaviour & Information Technology, 10:1, 65-79, DOI: 10.1080/01449299108924272

In this work, the computer component of the system is always represented by screenshots, which are not real representation of the system. It is therefore classified as representational computer. Concerning the user component of the system, it is represented by a popularity ranking based on software computation (Alexa) and by a aesthetics evaluation represented by algorithms computation (aesthetics evaluation experience with QuestIHM). Thus, we will use the “Analytic methods” framework. We also need to take into account the “User reports” framework when we will use the aesthetics rating given by participants (real user) in our experience.

The **popularity** ranking of Alexa is a one month rank. “It is calculated using a combination of average daily visitors and pageviews over the past month”⁴⁷. This ranking can be discussed because it is computed via software which analyze the user behavior on internet but not directly via the user. It can represents a limit in the analysis. For the experience, we choose the best ranked website for each e-commerce sector.

The aesthetics evaluation of screenshots is divided in two parts. First, it is done via QuestIHM, which compute **aesthetics** thanks to several measures and algorithms formula (representational user). Then, it is done via real users thanks to a **rating evaluation** (from 1 to 5) of screenshots (real user).

In this thesis, we try to analyze correlation between those three elements. We will first check how the computed aesthetics can influence the score given by the participants. We will then analyze the potential correlations between aesthetics and popularity of the websites. Finally, we will explore the possibility of a link between the score of participants and popularity of the chosen websites.

⁴⁷ Online on the website Alexa.com (2019) <https://www.alexa.com/topsites>

3.2 Questions and hypothesis

Research question: How to measure aesthetics of e-commerce web site on a mobile device ?

To run a scientific and relevant study, we need to ask a main research question and propose some sub-questions and hypothesis. Those hypothesis represent potential answers to the questions. To justify “what is beautiful is usable”, we will test different association between **aesthetics**, **popularity** and **score** of participants in order to identify potential correlations. We hypothesize that sites are popular (combination of average daily visitors and pageviews) because they are (potentially) usable and that the beauty can be evaluate by measures of aesthetics and rating of users in the experience. Those questions and hypotheses will guide us during the statistical analysis.

The objective of this part is to set up some questions that can give answers or contribution to the main question and justify the sentence of Tractinsky “What is beautiful is usable”. To do so, some restrictions need to be put in place because it is a complex question to answer. We will analyze the aesthetics of popular e-commerce websites screenshots on mobile device thanks to a software and people score rating. The popularity will be computed by Alexa ranking⁴⁸, and participants of the experience won’t be aware of this ranking.

The starting point to identify how to measure aesthetics of e-commerce websites on mobile device is to analyze how the different measures of aesthetics computed by a software can influence the score given by the participants. It will give a good overview of the impacts (positive or negative) and role of measures of aesthetics, and it will show which measures is relevant to predict to perceived beauty by users. This part is summarize in the research question 1 below.

⁴⁸ Online on the website Alexa. <https://www.alexa.com/topsites/category/Top/Shopping> (2018)

Research Question 1: Do measures used by QuestIHM represent a good indicator of aesthetics of e-commerce websites on mobile device?

Hypothesis 1: All measures are predictors of aesthetics of e-commerce websites.

After analyzing the impact and the relevance of the different visual techniques (measures) on the score, we will test if the popularity can be taken into account in the measure of aesthetics. This sub-part corresponds to the analysis of correlation between visual techniques (measures) and popularity as well as the correlation between the score (rating) and popularity. The part is summarize in the research question 2 below.

Research Question 2: Are aesthetics and popularity two linked concepts?

Hypothesis 2: Popular websites are beautiful for participants.

Hypothesis 3: Popular websites are aesthetics.

By answering those two research questions, we will have a clearer view on how to measure aesthetics of e-commerce website on mobile device. We will identify what are the most important measures in term of aesthetics measurement on mobile device as well as the correlation of those variables with the popularity of the different e-commerce websites.

To perform this analysis, we have to choose a methodology, which is explained in the next section.

4. Methodology

The methodology represents a key part in the understanding of the statistical analysis. The methodology is an important part in the reading materials, and a good reading material always provide a better reading comprehension.

4.1. Research method

Two types of methodology research exist: the qualitative and the quantitative.⁴⁹ In this work, we use the quantitative one. This method is divided in different steps. First, it is necessary to identify and clarify the theory of the topic based on documentary researches. Then, it is essential to identify a problem and ask questions that can solve the problem. Finally, few assumptions have to be set up. Those assumptions are potential results to the questions. Afterwards, a large amount of data has to be collected and then analyzed thanks to statistical tools. The purpose is to analyze the results based on the assumptions and to verify them.

The experience will follow this methodology because aesthetics remains relatively a subjective topic from a human perspective. By computing a large amount of data, we could probably highlight a general trend. Indeed, a premise needs to be considered when assessing aesthetics studies: “by collecting different (subjective) opinions on the aesthetics of an object, a work or any exposure, it is possible to get an answer that is collectively objective”⁵⁰. Moreover, the quantitative research is the most used in aesthetics studies.

It is important to align measures computed by a software and the representation of the measures that people have in mind. Two approaches exist to tackle this issue: either one derives a measure of a human judgment, or one derives a human judgment from measurements. In this work, we will conduct only one experience and we will directly compare human results (rating) and software results (aesthetics) and try to derive those human judgement from measures. We will also test if the aesthetics is linked to the popularity of websites (ranking) and if there is a link between the rating of participants and popularity of websites.

⁴⁹ Avenier, M-J., Gravard-Perret, M-L. (2012). *Méthodologie de la recherche en sciences de gestion*. Pearson France.

⁵⁰ Zen M., Vanderdonckt J. (2016). *Assessing User Interface Aesthetics based on the Intersubjectivity of Judgment*

4.2. Data collection

The data collection is the most important part of the experience and analysis. It represents the first step that allow to start the experience and the evaluation of aesthetics of screenshots (stimuli) via people beauty judgment expressed in term of a rating. To collect the participants, no discrimination criteria was set up. Everybody could take part to the experience. More people mean more accurate and relevant results. The reason of including everybody is simple. Consumers of e-commerce websites correspond to everybody, everywhere, at any time.

For the experience, 180 screenshots coming from 90 different e-commerce websites were gathered. To select the websites, we used the Alexa data base⁵¹. First, we made a percentage distribution of all e-commerce sites by sector. A percentage number was then selected for each sector in order to have a better market representativeness. For each sector, the most popular sites were selected based on the Alexa rankings (cf. section 3.1.). The screenshots corresponds to a picture of the landing page of each website. Those screenshots represent the stimuli of our experience, where participants have to rate them on a score between 1 and 5.

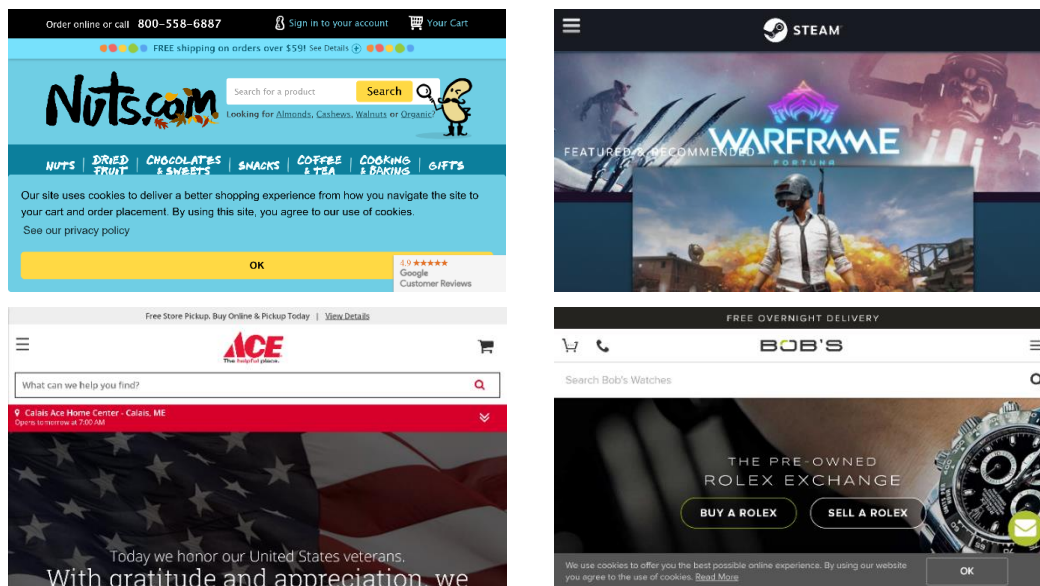


Figure 3 - Illustration of some screenshots

⁵¹ Online on the website Alexa. <https://www.alexa.com/topsites/category/Top/Shopping> (2018)

The next step was to gather a panel of person without any restriction of selection and to ask them to rate all the screenshots website. This step was feasible thanks to the web experience developed by Nicolas Burny⁵². The principle is simple, applicants must evaluate the aesthetic and visual beauty of the 180 pictures on a scale of 1 to 5. 1 means that the screenshot has the minimum level of aesthetics and 5 means that the screenshot have the highest level of aesthetics – in other words 5 means that the screen shot is visually appealing for the person. For each website, the participant has to evaluate its visual appealing in portrait and in landscape format. Obviously, two screenshots coming from the same website are not presented simultaneously, in order to not bias the results. The collected results were collected in a .json format that need to be transformed into .csv format. It allows to create an excel spreadsheet to reference all the results.



Figure 4 - Illustration of the software QuestIHM

The third step was to analyze the 180 screenshots in the software QuestIHM⁵³. This software has been created by Mathieu Zen from the Louvain School of Management. Its purpose was to analyze the aesthetics of picture or screenshot based on measures (called metrics in his work). To do so, a large panel of measures are available, and we have to define which measures we want to analyze by drawing the region (represented by grey area on pictures below) of the screenshot that have to be analyze. There is also an option called “automatic measures computing”. By choosing this method, QuestIHM will analyze by default a set of 6 measures.

⁵² irim-chilab.herokuapp.com/experiments/mobile_aesthetics

⁵³ questimapp.appspot.com

For this experience, the first method was chosen to gather all the measures. In order to obtain results based on the entire screenshot, the region selection covered each region of the screenshots (as on the picture below). This selection allows to get the most accurate measures computation of the balance, horizontal balance, vertical balance, equilibrium, density, center and external alignment, concentricity, simplicity and symmetry.

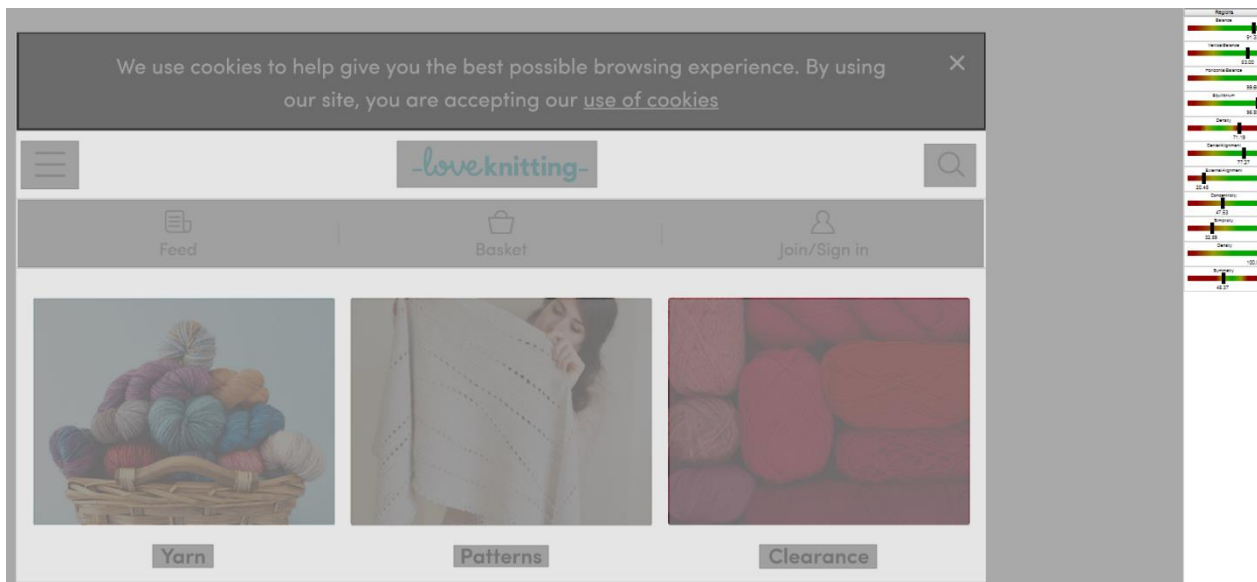


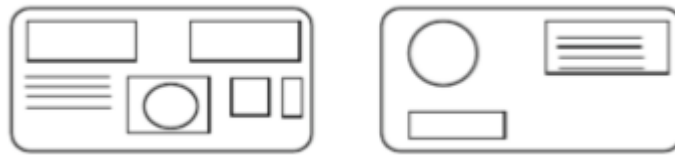
Figure 5 - Illustration of the zone definition in QuestIHM and metrics computation

Before going deeper in the analysis, it is important to understand the meaning of each variables used by QuestIHM. I'll define each of them. The variables that I'll use in this work are the same than Rolin Nelson⁵⁴ used in his thesis, so I'll use exactly the same definitions. Obviously, some of them will have more impact than others in our analysis. Those variables are called measures. A Measure (metric) is an attempt to quantify a visual technique characterizing the arrangement of the layout frame components in a User Interface (e.g. Balance). It basically takes a set of regions as an input, runs a computation through a pre-conceived formula (each one uses a different set of region attributes) and returns a numeric value between 0 and 1. This value is then assumed to quantitatively represent the UI score for a visual property. To end up, it is important to notice that QuestIHM works by using a software-based approach and compute measures via mathematical algorithms. It justifies that the user is representational in the Whitefield framework.

⁵⁴ Rolin N., *Analysis of the aesthetics of web sites on a mobile phone device*. (2018)

Balance

According to Ling Ngo, Seng Teo and G.Byrne⁵⁵, the balance refers to the distribution of optical weight. It is related to the perception of some objects in the picture. It takes into account the shape, color, size of the object.



The balance is computed as the difference between the total weighting of components on each side of the horizontal and vertical axis and is given by this formula:

$$BM = 1 - \frac{|BM_{\text{vertical}}| + |BM_{\text{horizontal}}|}{2} \in [0, 1].$$

Figure 6 – Balance formula

The balance is an independent variable qualified as scale variable between 0 and 1.

Equilibrium

According to Ling Ngo, Seng Teo and G.Byrne, the equilibrium refers to the distribution of the objects on the picture. According to the authors, the way to obtain the equilibrium is to centralize the layout, the disposition of the objects. It can be also defined as the difference between the center of mass of the object and the center of the screen.



⁵⁵ Chek Ling Ngo, D., Seng Teo, L., Byrne, G.A *Mathematical Theory of Interface Aesthetics*. (2000)

The equilibrium is computed as the difference between the mass center of the displayed elements and the physical center of the screen, and is given by

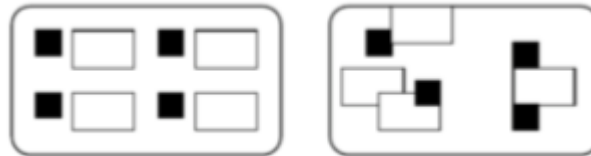
$$EM = 1 - \frac{|EM_x| + |EM_y|}{2} \in [0, 1].$$

Figure 7 – Equilibrium formula

The equilibrium is an independent variable qualified as scale variable between 0 and 1.

Simplicity

According to Ling Ngo, Seng Teo and G. Byrne⁵⁶, simplicity refers to the directness and the singleness of form. The simplicity can be increased by optimizing the number of objects on a screen and reduce the alignment.



Simplicity is computed thanks to the formula

$$SMM = \frac{3}{n_{vap} + n_{hap} + n} \in [0, 1],$$

Figure 8 – Simplicity formula

Where nvap and nhap are the numbers of vertical and horizontal alignment points.

The balance is an independent variable qualified as scale variable between 0 and 1.

⁵⁶ Chek Ling Ngo, D., Seng Teo, L., Byrne, G.A *Mathematical Theory of Interface Aesthetics*. (2000)

Symmetry

Symmetry analyzes the replication of one side of the screen to the other side. It allows us to know if one side of the screen is exactly replicated to the other side.



The symmetry is an independent variable qualified as scale variable between 0 and 1.

Density

According to Ngo, Seng Teo and G.Byrne⁵⁷, the density is defined as the number of objects covering the screen. Tully defined the measure of density; it is the percentage of character position on the frame containing data.



The density is computed via this formula where a_i and a_{frame} are the areas of object i and the frame, respectively, and n is the number of objects on the frame.

$$\text{DM} = 1 - 2 \left| 0.5 - \frac{\sum_i^n a_i}{a_{\text{frame}}} \right| \in [0, 1],$$

Figure 9 – Density formula

The density is an independent variable qualified as scale variable between 0 and 1.

⁵⁷ Chek Ling Ngo, D., Seng Teo, L., Byrne, G.A *Mathematical Theory of Interface Aesthetics*. (2000)

Horizontal Balance

The horizontal balance is defined as the balanced disposition about same elements on a horizontal axis. The comparison is computed between the higher side and the lower side.

$$\text{BM}_{\text{horizontal}} = \frac{w_T - w_B}{\max(|w_T|, |w_B|)},$$

Figure 10 – Horizontal balance formula

The horizontal balance is an independent variable qualified as scale variable.

Vertical Balance

Vertical balance refers to the balanced disposition about same objects on a vertical axis. The analysis of the replication is computed from the left to the right or the opposite.

$$\text{BM}_{\text{vertical}} = \frac{w_L - w_R}{\max(|w_L|, |w_R|)},$$

Figure 11 – Vertical balance formula

The vertical balance is an independent variable qualified as scale variable.

Center Alignment

Center alignment is defined as the way the elements are dispatched around the center of the screen. More the object will be centralized more the center alignment rating will be high.

The center alignment is an independent variable qualified as scale variable between 0 and 1.

External Alignment

External alignment refers to the distance of the disposition of the elements from the border of the screen. Less element near to the center, more the external alignment rating will be high.

The external alignment is an independent variable qualified as scale variable between 0 and 1.

Saliency Balance

According to Shen and Zhao (2014, quoted in M. Zen and J. Vanderdonckt lecture)⁵⁸, this variable refers to the computation of the balance according to the saliency. The saliency refers to something which is prominent.

The saliency balance is an independent variable qualified as scale variable between 0 and 1.

For the next variables, we didn't find any right definitions. As consequences, we tried to extrapolate the meaning without any guarantee of defining correctly the term.

Concentricity

According to Mathieu Zen (2017)⁵⁹, the concentricity measures the concentration of the element in the centre of the picture.



The concentricity is an independent variable qualified as scale variable between 0 and 1.

⁵⁸ Vanderdonckt, J., Zen, M. (2016) *Assessing User Interface Aesthetics based on the Intersubjectivity of Judgment*, 12.

⁵⁹ Zen, M. (2017). *A methodology for assessing aesthetics of a graphical user interface of an information system : visual measures-based automated evaluation*. Dial. 135

Border Density

Border density is defined as the number of elements covering the border of the screen.

The border density is an independent variable qualified as scale variable between 0 and 1.

Color Density

The color density refers to the intensity of color covering the screen.

The color density is an independent variable qualified as scale variable between 0 and 1.

Colorfulness

The colorfulness is the capacity of the image of being colorful



The balance is an independent variable qualified as scale variable.

Border Balance

We can define the border balance as the distribution of the optical weigh in the border of the picture.

The balance is an independent variable qualified as scale variable.

Compression Complexity

According to Tuch et al. (2012)⁶⁰, the mechanism of the compression complexity variable is to measure visual complexity according to the compression level of a picture. Complexity is defined as the sophistication of the image. It includes the number of elements, the disposition...The compression level refers to the compressed picture.

The balance is an independent variable qualified as scale variable.

Two other measures have to be defined for the statistical analysis: the score and the popularity.

Score

The score refers to the rating given by the applicants based on the aesthetics of the screenshot. The score is based on a scale from 1 to 5.

The score is a dependent variable qualified as scale variable.

Popularity

The popularity ranking of Alexa is a one month rank. “It is calculated using a combination of average daily visitors and pageviews over the past month”⁶¹. In this work, we have selected the most popular sites in each sector in the e-commerce (shopping) category. Websites are rank from 1 to 90. 1 is the best ranked website. We have to notice that this rank is a dynamic data, it changes and it updates at any time.

The popularity is an independent variable qualified as ordinal variable.

⁶⁰ Tuch, A. N., E. E. Presslauer, M. Stöcklin, K. Opwis, and J. a. Bargas-Avila (2012). *The role of visual complexity and prototypicality regarding first impression of websites: Working towards understanding aesthetic judgments*.70 (11), 794–811.

⁶¹ Online on the website Alexa.com (2019) <https://www.alexa.com/topsites>

Before starting the statistical analysis, an accurate excel sheet with all the results has to be prepared. The data from the experience were collected under a file format .json (JavaScript Object Notation), so the data have to be transferred into .xls (eXceL Spreadsheet) format. Then, every data has been sorted in order to have each screenshot associated with its computed measures, its score coming for human scoring and its popularity ranking. A distribution analysis of each measures has also been done (cf. table below). We also assume that every variable is numeric. Once the sheet was ready, the analysis could start thanks' to "SPSS". It is a complete statistical software that able to perform a large number of statistical tests.

5. Statistical analysis

This section is oriented around the question “how measure aesthetics of e-commerce website on mobile device” and “What is beautiful is usable”. This part is dedicated to the testing of our hypothesis. To do so, we will use SPSS. It is used a complete software used for interactive, or batched, statistical analysis designed by IBM. To get accurate results, we had to specify the type of data for each variable. Popularity is a numeric and ordinal data, score is a numeric and scale data between 1 and 5, and measures from QuestIHM are numeric and scale data between 0 and 1.

The first step is to identify the distribution of those variables. This distribution is summarized in appendix 1. We can see that the variables are very volatile. The range is relatively high for each measure, as well as the standard deviation (from 3% to 27%). Moreover, we see that the Skewness value of the variables are not equal to 0, it means that the distribution is not symmetric. To prove it, we perform a Shapiro-Wilk normality test (appendix 2) and the results prove that the data significantly deviate from a normal distribution ($\text{Sig} < 0,05$). We can also use the box-plot distribution of the different variables (cf. appendix 3) to analyze basic descriptive data. From the figure 12 (box-plot of concentricity) below, we see the minimum (around 0,3), the first quartile (around 0,5), the median (around 0,58), the third quartile (around 0,65) and the maximum (around 0,8). We also see that the data concentricity have some outliers (data number 1529, 1530, 1531 and 1532), that correspond to value that fall far outside the other values of the data set.

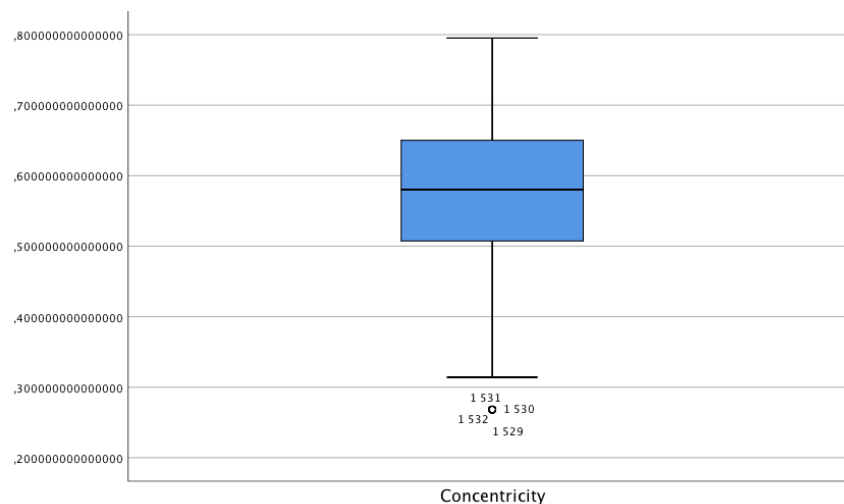


Figure 12 – Box plot of concentricity

5.1. Research Question 1: Do measures used by QuestIHM represent a good indicator of aesthetics of e-commerce websites on mobile device?

The aim of this question is to identify in which way the measures computed by QuestIHM can predict the score (e.g. the beauty judgement of users). This analyze is located in “User report” field in the framework of Whitefield. This first step is essential to test the potential correlation between popularity and aesthetics. In order to get a clearer view of the roles of each measure and their impact on score, we will use the multiple linear regression method. It will allow us to understand a large number of information such the effect of each variable, their relationship with the dependent variable, the significance, ... In this case, the dependent variable is the score (rate of screenshots by participants) and the independents variables represents all the variables (16) computed via formula by QuestIHM.

Hypothesis 1: All measures are predictors of aesthetics of e-commerce websites.

The first step to test this hypothesis is to perform a multiple linear regression of the variables. Score is the dependent variable and the measures provided by QuestIHM are the independent variables. Starting from this process, we can analyze the coefficient of regression of each variable. It is a first representation of the influence that the variables could have on the score. SPSS gives us a table where the statistic factor (e.g. beta, error standard,) are on the top of the tables and the list of coefficients are in the left column.

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Standard error	Bêta		
1	(Constant)	6,160	,849		7,252	,000
	saliencyBalance	,256	,200	,033	1,279	,201
	borderBalance	-,113	,167	-,016	-,678	,498
	borderDensity	3,765	,902	,127	4,173	,000
	colorDensity	,547	,205	,076	2,671	,008
	Colorfulness	-,417	,115	-,100	-3,634	,000

	CompressionComplexity	-2,100	,406	-,155	-5,174	,000
	VerticalBalance	-,202	,118	-,043	-1,712	,087
	HorizontalBalance	,500	,145	,093	3,446	,001
	Equilibrium	-,450	,578	-,020	-,780	,436
	Density	-1,600	,337	-,275	-4,746	,000
	CenterAlignment	,546	,236	,076	2,311	,021
	ExternalAlignment	,307	,411	,020	,747	,455
	Concentricity	-1,276	,308	-,114	-4,149	,000
	Simplicity	-3,695	,592	-,377	-6,245	,000
	Symmetry	-1,169	,980	-,029	-1,193	,233

a Dependent variable : Score

Figure 13 – Table of coefficients of the multiple linear regression

This table shows the influence of each variables on the score. The variables that have the most positive impact are border density (3,765), color density (0,547), horizontal balance (0,5) and center alignment (0,546). Others variables such as compression complexity (-2.100), concentricity (-1,276) or simplicity (-3,695) have a negative impact on score. The result concerning the complexity is align with the results coming from the experience of Tuch A. and al (2012)⁶², users prefer interface with a lower complexity. The interpretation of the different numbers is easy: if the border density increase by 0,1 the score will increase by $0,1 \times 3,765 = 0,376$.

From this multiple linear regression, we can also analyze the R square. It is a relevant factor to analyze because it demonstrates the proportion of variance that is explained by the model. The R square is express by a number between 0 and 1. In our case, this number represents the percentage of variation that the independents variables have on the score. To get a more accurate prediction, we will use the adjusted R square, which take more predictors into account and give a better interpretation on the residual standard error.

⁶² Tuch A., and Al. (2012). *The role of visual complexity and prototypicality regarding first impression of websites: Working towards understanding aesthetic judgments.*

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,317	,100	,093	1,067

Figure 14 – Table of adjusted R Square of the model

We can see that only 9,3% of the score variance is explained by the variation of the different independents variables. From a statistical point of view, it is relatively low.

An important parameter to analyze is the Residual Standard Error. We can first analyze this component for the linear regression and then analyze it for each variables individually. The residual standard error explains the deviation of the score for each screenshot. In this case, it is equal to 1,067. From this statistic, we can compute the Percentage Error by dividing the RSE by the mean of score: $1,067/2,64=40\%$. To analyze the RSE of each variables, another table is needed.

	N	Mean	Standard Deviation	Std Error Mean
saliencyBalance	1788	,721	,146	,003
borderBalance	1788	,705	,164	,004
borderDensity	1788	,063	,038	,001
colorDensity	1788	,373	,155	,003
Colorfulness	1788	,491	,267	,006
CompressionComplexity	1788	,249	,083	,002
Balance	1788	,752	,169	,003
VerticalBalance	1788	,643	,241	,005
HorizontalBalance	1788	,860	,209	,004
Equilibrium	1788	,950	,049	,001
Density	1788	,550	,193	,004
CenterAlignment	1788	,373	,156	,003
ExternalAlignment	1788	,074	,075	,002
Concentricity	1788	,579	,099	,002
Simplicity	1788	,339	,114	,002
Density	1788	1,00	,000	,000
Symmetry	1788	,459	,028	,001

Figure 15 – Table of spread

The Standard Error Mean are quite small and don't have any effect. The figure 15 can also be used to compute another significative data, the effect size coefficient. It analyzes the effect size by checking the impact of one variable on the other ones. The relevant effect size coefficient here concerns the impact of each variable on the score. To do so, we need to collect the mean and standard deviation of each variable. The mean and standard deviation of the score are respectively 2,64 and 1,116. Now that we have all the data, we'll use the effect size calculator provided by the University of Colorado⁶³ where group 1 = the score and group 2 = other variables.

⁶³ <https://www.uccs.edu/lbecker/>

Group 1	Group 2
M_1	M_2
2.64	0.721
SD_1	SD_2
1.116	0.146
Compute	
Reset	
Cohen's d	effect-size r
2.411241695227954	0.7696899285876836

Figure 16 – Effect size calculator illustration with saliency balance

To use the calculator, we have to use the mean and standard deviation of the score (group 1) and the other variables (group 2) one by one. The figure 15 shows an example with the score and the saliency balance.

The results collected are:

- | | |
|--------------------------------|----------------------------|
| • Saliency balance: 0,77 | • Horizontal balance: 0,74 |
| • Border balance: 0,77 | • Equilibrium: 0,73 |
| • Border density: 0,85 | • Density: 0,79 |
| • Color density: 0,82 | • Center alignment: 0,82 |
| • Colorfulness: 0,8 | • External alignment: 0,85 |
| • Compression complexity: 0,83 | • Concentricity: 0,79 |
| • Balance: 0,76 | • Simplicity: 0,82 |
| • Vertical balance: 0,78 | • Symmetry: 0,81 |

The effect size interpretation is described as: between 0,2 and 0,5 the effect is small, between 0,5 and 0,8 the effect is medium and above 0,8 the effect is high. We can see that from the Cohen kappa coefficient and effect size computation, each variable have an important effect on the score.

To conclude this part, we can test the significance of the model and the variables. To do so, we have to analyze the p-value statistic. “The P value means the probability, for a given statistical model that, when the null hypothesis is true, the statistical summary would be equal to or more extreme than the actual observed results. Therefore, P values only indicate how incompatible the data are with a specific statistical model (usually with a null-hypothesis). The smaller the P value, the greater statistical incompatibility of the data with the null hypothesis.”⁶⁴

Hypothesis test: $H_0: \beta = 0$ and $H_1: \beta \neq 0$. The rule of the test are the same for each variable and corresponds to: we reject the null hypothesis if $\text{Sig.} < 0.05$. The null hypothesis (H_0) is “all measures are not predictors of aesthetics”, and H_1 is “all measures are predictors of aesthetics”.

⁶⁴ Nahm Sahngun F. (2017). *What the P value really tell us*. Korean J. Pain. 30(4): 241–242.

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Standard error	Bêta		
1	(Constant)	6,160	,849		7,252	,000
	saliencyBalance	,256	,200	,033	1,279	,201
	borderBalance	-,113	,167	-,016	-,678	,498
	borderDensity	3,765	,902	,127	4,173	,000
	colorDensity	,547	,205	,076	2,671	,008
	Colorfulness	-,417	,115	-,100	-3,634	,000
	CompressionComplexity	-2,100	,406	-,155	-5,174	,000
	VerticalBalance	-,202	,118	-,043	-1,712	,087
	HorizontalBalance	,500	,145	,093	3,446	,001
	Equilibrium	-,450	,578	-,020	-,780	,436
	Density	-1,600	,337	-,275	-4,746	,000
	CenterAlignment	,546	,236	,076	2,311	,021
	ExternalAlignment	,307	,411	,020	,747	,455
	Concentricity	-1,276	,308	-,114	-4,149	,000
	Simplicity	-3,695	,592	-,377	-6,245	,000
	Symmetry	-1,169	,980	-,029	-1,193	,233

Figure 17 - Table of the representation of the individual significance

We can reject the null hypothesis for border density, colorfulness, compression complexity, horizontal balance, density, concentricity and simplicity. Those variables are statically significant for the model. Other variables are not.

We can do exactly the same for the entire model, by keeping the same rules and hypothesis: $H_0: \beta = 0$ and $H_1: \beta \neq 0$. We reject the null hypothesis if $\text{Sig.} < 0.05$. The null hypothesis (H_0) is “all measures are not predictors of aesthetics”, and H_1 is “all measures are predictors of aesthetics”. In the model, we reject the null hypothesis and so the model is significant (figure 18).

ANOVA

Model		Sum of Square	ddl	Mean Square	F	Sig.
1	Regression	224,858	15	14,991	13,158	,000b
	de Student	2018,809	1772	1,139		
	Total	2243,666	1787			

Figure 18 - Table of the representation of the global significance

a Dependent variable : Score

b Predictors : (Constant), Symmetry, Colorfulness, VerticalBalance, ExternalAlignment, HorizontalBalance, borderDensity, Concentricity, borderBalance, Equilibrium, saliencyBalance, Simplicity, colorDensity, CompressionComplexity, CenterAlignment, Density

By gathering those analysis, we can give a potential answer to the first hypothesis: yes, all measures of aesthetics computed by QuestIHM are predictors of aesthetics. The effect size coefficient ($>0,8$) and global variance significance test (Sig. $<0,05$) provide the positive conclusion that all measures of are predictors of aesthetics. However, some elements must be taken into account: only 7 measures are individually significant for the model (border density, colorfulness, compression complexity, horizontal balance, density, concentricity, simplicity). Moreover, only 9,3% (adjusted R square) of the variance of the score is explained by the model, which is a low percentage and the percentage error is 40%.

We can draw a new model by taking into account those significant variables:

$$\text{Score} = 3,765 * \text{borderDensity} - 0,417 * \text{Colorfulness} - 2,1 * \text{CompressionComplexity} + 0,5 * \text{HorizontalBalance} - 1,6 * \text{Density} - 1,276 * \text{Concentricity} - 3,695 * \text{Simplicity}.$$

5.2. Research Question 2: Are aesthetics and popularity two linked concepts?

Now that we identified how the measures impact the score, we want to test if the aesthetics and the score can have a link with the popularity of the different e-commerce websites. As a reminder, we choose 90 websites that cover each sector of the category “e-commerce” and that are the most popular in their respective category. We take that rank into account and classify the screenshots from 1 to 90 by popularity order. We will analyze if there is a link between popularity rank and aesthetics. This analyze is located in “Analytic methods” field in the framework of Whitefield.

Hypothesis 2: Popular websites are beautiful for participants.

To give answer to this hypothesis, we could have done a Spearman’s test or Kendall's tau-b (τ_b) correlation coefficient. Two assumptions need to be verified to perform this kind of test. The variables must be measured as ordinal, interval or ratio scale, which is verified in this case. Then, a monotonic relationship between the two variables is required. It means that when a variable increase or decrease, the other variable follows the same trend. To test it, we have performed a scatterplot graph with our ordinal variable popularity and the mean of score for each rank of popularity. Thanks to the graph below, we see that there is no monotonic relationship and so the Spearman’s test won’t be relevant.

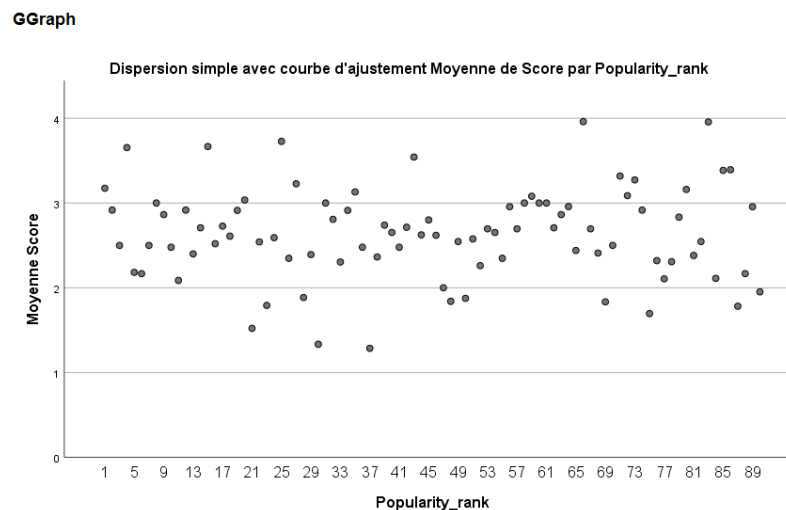


Figure 19 – Scatterplot graph of the mean score and popularity

This first ascertainment shows that the two variables are not well correlated, but we'll perform more relevant statistical test to identify their link (if existent). The test that we will perform to test correlation is a complete One-way ANOVA test. First, we have a look at the descriptive statistics to get picture of the data.

Score											
Popularity_rank	Moyenne	N	Ecart type								
1	3,17	23	1,072	31	3,00	22	1,155	61	3,00	22	,816
2	2,92	24	1,139	32	2,81	26	1,201	62	2,71	24	,999
3	2,50	24	,978	33	2,30	23	,926	63	2,86	22	,774
4	3,65	26	1,129	34	2,91	23	1,125	64	2,96	24	1,197
5	2,18	22	,907	35	3,13	23	,920	65	2,44	25	,917
6	2,17	24	,963	36	2,48	23	1,039	66	3,96	25	1,306
7	2,50	24	1,103	37	1,29	21	,463	67	2,70	23	,926
8	3,00	23	1,168	38	2,36	22	1,093	68	2,41	22	,734
9	2,86	22	1,424	39	2,74	23	1,096	69	1,83	24	,868
10	2,48	23	,947	40	2,65	23	1,112	70	2,50	20	,761
11	2,09	23	,949	41	2,48	23	,994	71	3,32	22	1,287
12	2,92	24	1,381	42	2,71	21	1,007	72	3,09	23	1,041
13	2,40	25	,866	43	3,54	24	,779	73	3,27	22	1,120
14	2,71	24	,806	44	2,63	24	1,056	74	2,92	24	,881
15	3,67	24	1,007	45	2,80	20	1,105	75	1,70	23	,703
16	2,52	25	1,046	46	2,62	21	,973	76	2,32	25	,852
17	2,73	22	,827	47	2,00	25	,764	77	2,11	19	,809
18	2,61	23	,891	48	1,84	25	,800	78	2,31	26	1,050
19	2,91	23	,949	49	2,55	22	1,184	79	2,83	24	,917
20	3,04	28	1,105	50	1,88	24	,850	80	3,16	25	1,028
21	1,52	23	,846	51	2,58	26	,945	81	2,38	21	1,024
22	2,54	24	1,062	52	2,26	23	1,176	82	2,55	22	,800
23	1,79	24	,779	53	2,70	23	1,146	83	3,96	23	,706
24	2,59	22	,908	54	2,65	23	,832	84	2,11	27	,892
25	3,73	22	,827	55	2,35	23	,935	85	3,38	26	,941
26	2,35	23	1,027	56	2,96	23	,928	86	3,39	28	1,100
27	3,23	22	1,152	57	2,70	23	1,105	87	1,78	23	,850
28	1,88	26	,816	58	3,00	23	1,168	88	2,17	24	1,049
29	2,39	23	1,234	59	3,08	25	1,152	89	2,96	23	1,065
30	1,33	24	,565	60	3,00	24	1,180	90	1,95	21	,973
								Total	2,64	2105	1,116

Figure 20 – Descriptive analysis between Popularity and Score

By analyzing this table, we can see that each popularity rank have a different mean but the means do not follow a specific trend. Examples of trend could be that the means decrease rank by rank (meaning that more popular the website is, more beautiful it is judged) or the contrary increase rank by rank. We can now perform that ANOVA test. The output is given below.

ANOVA

Model		Sum of Square	ddl	Mean Square	F	Sig.
1	Regression	224,858	15	14,991	13,158	,000b
	de Student	2018,809	1772	1,139		
	Total	2243,666	1787			

Figure 21 – ANOVA test of popularity and score

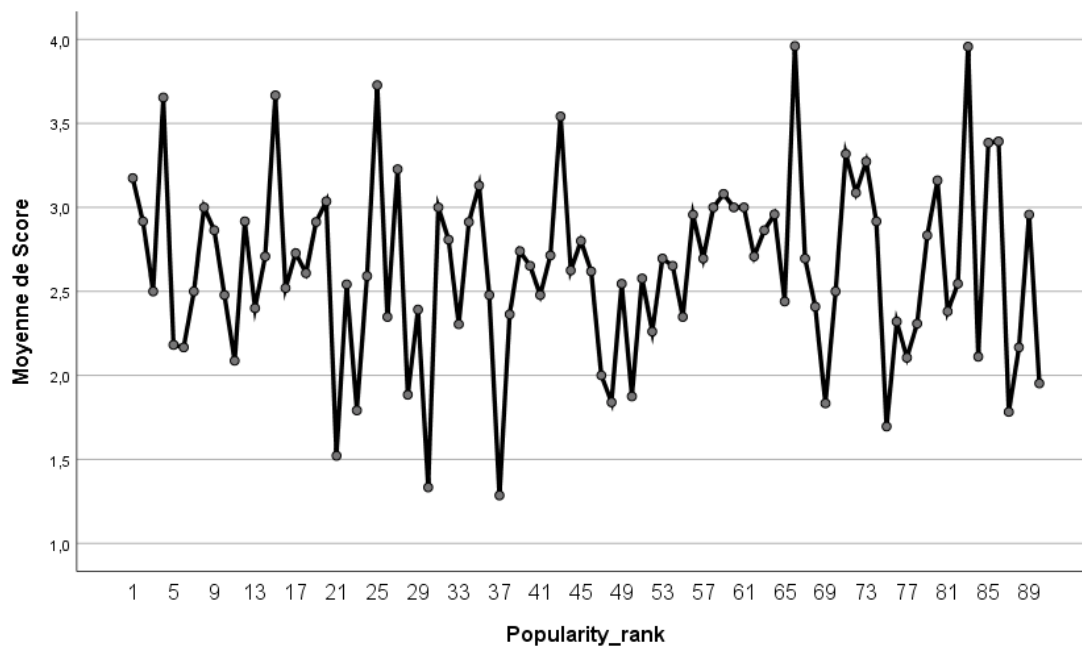


Figure 22 – Mean plot of popularity and score

We can see that the mean score is significantly different for the popularity rank groups ($F_{89, 2015} = 6,910$, $p < 0.001$) but the differences between the mean are not linear. The website ranked 1 does not have the higher value and the website ranked 90 does not have the lower value (or vice versa). It is the reason why we couldn't perform a bivariate correlation analysis. The null hypothesis is not reject, so popular websites are not necessarily beautiful (in other words, good rank from Alexa does not mean good rate from user). To confirm this conclusion, we analyze the data

extracted from Jonckheere-Terpstra test. “The Jonckheere-Terpstra test is a rank-based nonparametric test that can be used to determine if there is a statistically significant trend between an ordinal independent variable and a continuous or ordinal dependent variable.”⁶⁵

	Null hypothesis	Test	Sig.	Decision
1	The distribution of Score is the same accross categories of Popularity_rank.	Independent-Samples Jonckheere-Terpstra Test for Ordered Alternatives	,537	Retain the null hypothesis

Figure 23 - Jonckheere-Terpstra test for score and popularity

H0: $\beta = 0$ and H1: $\beta \neq 0$

The Jonckheere-Terpstra test for ordered alternatives showed that the alternative hypothesis is rejected, and that **there is not a statistically significant trend of higher median scores with higher rank popularity**, $p > 0.05$.

⁶⁵ Online on the website statistics.laerd.com, <https://statistics.laerd.com/spss-tutorials/jonckheere-terpstra-test-using-spss-statistics.php>

Hypothesis 3: Popular websites are aesthetics.

To test this third hypothesis, we will compare if there is a link between aesthetics measured by formula via QuestIHM and the rank popularity. To do so, we will analyze how correlate are the visual techniques (measures) with the popularity ranking of the websites via ANOVA and Jonckheere-Terpstra test.

ANOVA

		Sum of square	ddl	Mean Square	F	Sig.
borderDensity	Intergroup	1,964	86	,023	66,410	,000
	Intragroup	,585	1701	,000		
	Total	2,549	1787			
saliencyBalance	Intergroup	28,571	89	,321	38,121	,000
	Intragroup	16,968	2015	,008		
	Total	45,539	2104			
borderBalance	Intergroup	28,906	86	,336	30,254	,000
	Intragroup	18,898	1701	,011		
	Total	47,804	1787			
colorDensity	Intergroup	36,920	89	,415	78,069	,000
	Intragroup	10,707	2015	,005		
	Total	47,627	2104			
Colorfulness	Intergroup	135,329	89	1,521	219,439	,000
	Intragroup	13,963	2015	,007		
	Total	149,292	2104			
CompressionComplexity	Intergroup	12,837	89	,144	146,088	,000
	Intragroup	1,989	2015	,001		
	Total	14,827	2104			
Balance	Intergroup	36,317	89	,408	36,737	,000
	Intragroup	22,382	2015	,011		
	Total	58,699	2104			
VerticalBalance	Intergroup	66,162	89	,743	24,729	,000

	Intragroup	60,574	2015	,030		
	Total	126,736	2104			
HorizontalBalance	Intergroup	60,797	89	,683	61,840	,000
	Intragroup	22,258	2015	,011		
	Total	83,055	2104			
Equilibrium	Intergroup	3,084	89	,035	32,266	,000
	Intragroup	2,164	2015	,001		
	Total	5,249	2104			
Density	Intergroup	42,248	89	,475	29,823	,000
	Intragroup	32,073	2015	,016		
	Total	74,321	2104			
CenterAlignment	Intergroup	27,245	89	,306	22,280	,000
	Intragroup	27,686	2015	,014		
	Total	54,931	2104			
ExternalAlignment	Intergroup	8,696	89	,098	40,795	,000
	Intragroup	4,826	2015	,002		
	Total	13,522	2104			
Concentricity	Intergroup	9,756	89	,110	20,157	,000
	Intragroup	10,958	2015	,005		
	Total	20,713	2104			
Simplicity	Intergroup	12,060	89	,136	19,797	,000
	Intragroup	13,792	2015	,007		
	Total	25,853	2104			
Symmetry	Intergroup	,947	89	,011	32,294	,000
	Intragroup	,664	2015	,000		
	Total	1,610	2104			

Figure 24 – ANOVA test for every measure regarding popularity rank

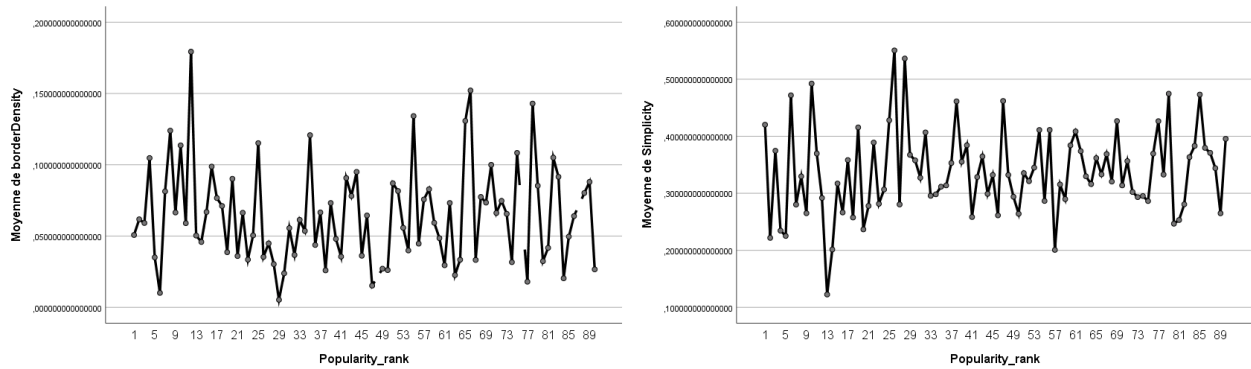


Figure 25 – Example of mean plot (border density and simplicity regarding popularity rank)

For each visual techniques, we can see that the mean is significantly different for the popularity rank groups ($p < 0.05$). But they are not linearly different, as for the score (hypothesis2).

Summary of the Jonckheere hypothesis test

	Null hypothesis	Test	Sig.	Decision
1	The distribution of saliencyBalance is the same across categories of Popularity_rank.	Independent-Samples Jonckheere-Terpstra Test for Ordered Alternatives	,000	Reject the null hypothesis
2	The distribution of borderBalance is the same across categories of Popularity_rank.	Independent-Samples Jonckheere-Terpstra Test for Ordered Alternatives	,020	Reject the null hypothesis
3	The distribution of borderDensity is the same across categories of Popularity_rank.	Independent-Samples Jonckheere-Terpstra Test for Ordered Alternatives	,093	Retain the null hypothesis
4	The distribution of colorDensity is the same across categories of Popularity_rank.	Independent-Samples Jonckheere-Terpstra Test for Ordered Alternatives	,015	Reject the null hypothesis

5	The distribution of Colorfulness is the same across categories of Popularity_rank.	Independent-Samples Jonckheere-Terpstra Test for Ordered Alternatives	,570	Retain the null hypothesis
6	The distribution of CompressionComplexity is the same across categories of Popularity_rank.	Independent-Samples Jonckheere-Terpstra Test for Ordered Alternatives	,006	Reject the null hypothesis
7	The distribution of Balance is the same across categories of Popularity_rank.	Independent-Samples Jonckheere-Terpstra Test for Ordered Alternatives	,595	Retain the null hypothesis
8	The distribution of VerticalBalance is the same across categories of Popularity_rank.	Independent-Samples Jonckheere-Terpstra Test for Ordered Alternatives	,977	Retain the null hypothesis
9	The distribution of HorizontalBalance is the same across categories of Popularity_rank.	Independent-Samples Jonckheere-Terpstra Test for Ordered Alternatives	,884	Retain the null hypothesis
10	The distribution of Equilibrium is the same across categories of Popularity_rank.	Independent-Samples Jonckheere-Terpstra Test for Ordered Alternatives	,000	Reject the null hypothesis
11	The distribution of Density is the same across categories of Popularity_rank.	Independent-Samples Jonckheere-Terpstra Test for Ordered Alternatives	,000	Reject the null hypothesis
12	The distribution of CenterAlignment is the	Independent-Samples Jonckheere-Terpstra Test for Ordered Alternatives	,000	Reject the null hypothesis

	same across categories of Popularity_rank.			
13	The distribution of ExternalAlignment is the same accross categories of Popularity_rank.	Independent-Samples Jonckheere-Terpstra Test for Ordered Alternatives	,352	Retain the null hypothesis
14	The distribution of Concentricity is the same across categories of Popularity_rank.	Independent-Samples Jonckheere-Terpstra Test for Ordered Alternatives	,230	Retain the null hypothesis
15	The distribution of Simplicity is the same accross categories of Popularity_rank.	Independent-Samples Jonckheere-Terpstra Test for Ordered Alternatives	,001	Reject the null hypothesis
16	The distribution of Symmetry is the same across categories of Popularity_rank	Independent-Samples Jonckheere-Terpstra Test for Ordered Alternatives	,000	Reject the null hypothesis

Figure 26 - Jonckheere-Terpstra test for border density, simplicity and popularity

$H_0: \beta = 0$ and $H_1: \beta \neq 0$

The Jonckheere-Terpstra test for ordered alternatives showed that we can reject the alternative hypothesis if $p > 0.05$, and so that there is not a statistically significant trend of higher median of the measures with higher rank popularity.(border density, colorfulness, balance, external alignment, concentricity). Otherwise, we can reject the null hypothesis when $p < 0.05$ and that there is a statistically significant trend of higher median of the other measures with higher rank popularity.

6. Results

Thanks to SPSS, we have performed several statistical tests to explore the potential links between aesthetics computed by QuestIHM, the score of people and the popularity of the e-commerce websites computed by Alexa. The data that we collected show that the measures of aesthetics computed by QuestIHM are good predictors of aesthetics. But some remarks must be taken into account when assessing this result. The measures explains only 9,3% of the variance of the model and only 7 measures are individually significant in the prediction of the score. The negative influence of compression complexity (-2,100) is similar to the conclusion of Tuch A. and al (2012)⁶⁶. From those conclusions, we can draw an accurate model of prediction for the score.

$$\text{Score} = 3,765 * \text{borderDensity} - 0,417 * \text{Colorfulness} - 2,1000 * \text{CompressionComplexity} + 0,500 * \text{HorizontalBalance} - 1,6 * \text{Density} - 1,276 * \text{Concentricity} - 3,695 * \text{Simplicity}.$$

Then, we perform several statistical tests to identify in which way the popularity of the different websites can have a correlation with the aesthetics. Overall, the ascertainment is that the two concepts are not linked. Thanks to the mean plot, and then the Jonckheere-Terpstra test, we saw that the relation between the popularity and the aesthetics is not monotonic. That means that the aesthetics level does not follow the popularity rank.

However, some measures have a statistically significant trend of higher median with higher rank popularity. It is the case for saliency balance, border balance, color density, compression complexity, equilibrium, density, center alignment, simplicity and symmetry. We can therefore suggest that those measures are predictors of the level of popularity of e-commerce websites.

⁶⁶ Tuch A., and Al. (2012). *The role of visual complexity and prototypicality regarding first impression of websites: Working towards understanding aesthetic judgments.*

7. Conclusions

7.1. Conclusions

Despite the fact that the topic is not yet totally developed, we quickly noticed that the aesthetics and the GUI are an important dilemma for companies and information systems designers. The existing studies bring some answers to aesthetics measurement questions, thanks to several frameworks, measures and formula. Our goal was also to bring a part of answer to the question “how to measure aesthetics?”. That is why we choose to focus on a specific topic of the discipline, and analyze the aesthetics of 1) e-commerce websites, 2) which are ranked as popular and 3) on mobile device. The reason of this choice is because the electronic commerce is totally exploding as well as the use of the mobile device. It seems essential to lighten to topic.

The answer to the first research question **“Do measures used by QuestIHM represent a good indicator of aesthetics of e-commerce websites on mobile device?”** is yes, but not for every measures. Some of them are not relevant to predict aesthetics. The measures to retain as predictor of aesthetics of e-commerce website on mobile device are border density, colorfulness, compression complexity, horizontal balance, density, concentricity and simplicity. We also have to noticed that all measures only predict 9,3% of the variance of the model and that the standard error is 40%.

The second question **“Are aesthetics and popularity two linked concepts?”** is clearer to answer. No, a popular website is not necessarily aesthetics, and an aesthetics website is not necessarily popular. The aesthetics results are significantly different from a rank to another but the relation is not linear. However, we can see the saliency balance, border balance, color density, compression complexity, equilibrium, density, center alignment, simplicity and symmetry measures have a statistically significant trend of higher median with higher rank popularity. Those results could be interesting to test on a larger scale.

7.2. Limits and problems

Several limits must be considered in this work. Concerning the regression, we know that the significance of the model and the impacts of the variables can change with the sample size. Moreover, due to the large number of variables, we could encounter multicollinearity problems.

Another limit in our analysis is the fact that we used the Alexa ranking to evaluate the popularity of the chosen e-commerce websites. Other tools exist to evaluate popularity of websites (such as WooRank⁶⁷ for example). Our results would probably be different if another tool was used. It can be interesting to perform those tests with another popularity ranking tool to compare the results. Another remark concerning the Alexa ranking: the popularity rank is computed via several measures concerning the behavior of users on different websites. Those results can be inaccurate because it does not take into account a panel of measures of judgement, such as questionnaire and personal opinion of users for example. Moreover, the ranks given by Alexa are dynamic, they change every day.

We also hope that the participants give sincere answers and perform the experience with accuracy. Rating 180 screenshots takes time and some of them could have been bored to finish it. Unfortunately, we can't verify this hypothetical statement.

⁶⁷ <https://www.woorank.com/fr/>

8. Bibliography

- “Aesthetics” definition. Merriam-Webster Dictionary. (2018)
- Albert, W. and Tullis, T. (2013). Measuring the User Experience 2nd edition: Collecting, Analyzing, and Presenting.
- Avenier, M-J., Gravard-Perret, M-L. (2012). Méthodologie de la recherche en sciences de gestion. Pearson France.
- Baumgarten, A. (1739). Metaphysics a Critical Translation With Kant’s Elucidations, Selected Notes and Related Materials.
- Chek Ling Ngo, D., Seng Teo, L., Byrne, G.A Mathematical Theory of Interface Aesthetics. (2000)
- De Angeli A. and Al. (2006). Interaction, usability and aesthetics: what influences users' preferences? DIS '06 Proceedings of the 6th conference on Designing Interactive systems, pages 271-280
- Doug Engelbart Institute. The Dawn of Interactive Computing. Online on the website Doug Engelbart. <http://dougengelbart.org/firsts/interactive-computing.html> (2018)
- Haynes J., Paradice D. (June 2007). Information technology and aesthetics : passive and active dimensions. Australasian Journal of Information Systems, Volume 14, Number 2.
- Igwenagu Ch. (2016). Fundamentals of research methodology and data collection.
- Jansen, B. J. (1998). The Graphical User Interface: An Introduction . SIGCHI Bulletin. 30(2), 22-26
- Kant, I., (1790). Critique of judgement
- Khalighy S. and Al. (2014). Measuring aesthetics in design. International design conference.
- Kostellow, R. R. (2002). Elements of design, Princeton Architectural New York NY.
- Leder and AL. (2004). A model of aesthetic appreciation and aesthetic judgments.
- Miniukovich A. and De Angeli A. (2014), Visual impressions of mobile app interfaces. Conference: NordiCHI.
- Moran K. (2017), The aesthetics-usability effect. Nielsen Norman group article.
- Mumcu, Y., and Kimzan, H. (2015). The effect of visual product aesthetics on consumer’s price sensitivity. Procedia Economics and Finance, 26, 528-534
- Nahm Sahngun F. (2017). What the P value really tell us. Korean J. Pain. 30(4): 241–242.

- Ngo D. C. L., Teo L. S., Byrne J. G. (2000). A Mathematical Theory of Interface Aesthetics. Visual Mathematics 10.8
- Oulasvirta A. and Al. (2018). Aalto Interface Measures (AIM): A Service and Codebase for Computational GUI Evaluation. Aalto University.
- Padua A.(2003), Structure d'un document scientifique.
- Page, C., and Herr,P. M. (2002) . An investigation of the processes by which product design and brand strength interact to determine initial affect and quality judgment. Journal of consumer psychology,12, 133-147
- Rolin N., (2018). Analysis of the aesthetics of web sites on a mobile phone device.
- Schmidt, T., and Wolff, C. (2018). The influence of user interface attributes on Aesthetics.
- Shaoolian, G. (2017). The Value of a Website Strategy. Online on the website Huffington post. https://www.huffingtonpost.com/gabriel-shaoolian/the-value-of-a-website-strategy_b_8807554.html?guccounter=1.t
- Silvennoinen J, Vogel M and Kujala S. (2014). Experiencing Visual Usability and Aesthetics in Two Mobile Application Contexts. Journal of usability studies, vol. 10, pp 46-62.
- Soui M., Chouchane M., Gasmi I. and Mkaouer M. (2017). PLAIN: PPlugin for predicting the usAbility of Mobile User INterface.In Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 1.
- Tractinsky and al. The Encyclopedia of Human-computer Interaction, 2nd Ed, visual aesthetics.
- Tractinsky N., Katz A.S., Ikar D (2000). What is beautiful is usable. Interacting with computers.
- Tuch A., and Al. (2012). The role of visual complexity and prototypicality regarding first impression of websites: Working towards understanding aesthetic judgments.
- Venny J. (2017). Étude de l'effet de l'esthétique sur l'utilisabilité d'une interface lors d'une tâche de recherche d'informations sur un site Internet.
- Whitefield A. and al. (1991). A framework for human factors evaluation, Behaviour & Information Technology, 10:1, 65-79, DOI: 10.1080/01449299108924272
- Yung-Ming, L., Yung- Shao,Y. (2010). Increasing trust in mobile commerce through design aesthetics. Computers in Human Behavior, 26 (4),673-684.
- Zangwill, Nick. (2003-2007). Aesthetic Judgment, Stanford Encyclopedia of Philosophy,. Retrieved 2008.

- Zen M., Vanderdonckt J. (2016). Assessing User Interface Aesthetics based on the Intersubjectivity of Judgment
- Zen, M. (2017). A methodology for assessing aesthetics of a graphical user interface of an information system : visual measures-based automated evaluation. Dial. 135
- Zen, Mathieu ; Vanderdonckt, Jean. Towards an evaluation of graphical user interfaces aesthetics based on metrics. IEEE Eighth International Conference on Research Challenges in Information Science (RCIS'2014).
- <http://questimapp.appspot.com>
- http://irim-chilab.herokuapp.com/experiments/mobile_aesthetics
- <https://www.woorank.com/fr/>
- Online on the website uccs.edu. <https://www.uccs.edu/lbecker/> (2019)
- Online on the website Alexa. <https://www.alexa.com/topsites/category/Top/Shopping> (2018)
- Online on the website Interaction design. <https://www.interaction-design.org/literature/topics/aesthetics> (2018)
- Online on the website Interaction design. <https://www.interaction-design.org/literature/topics/aesthetics> (2018)
- Online on the website Statista. <https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/> (2018)
- Online on the website Statista. <https://www.statista.com/statistics/306528/share-of-mobile-internet-traffic-in-global-regions/> (2018)
- Online on the website Statista. <https://www.statista.com/statistics/678739/forecast-on-connected-devices-per-person/> (2018)
- Online of the website etymonline.com. <https://www.etymonline.com/word/aesthetic> (2018)
- Online on the web site Collins Dictionary. <https://www.collinsdictionary.com/dictionary/english/aesthetics> (2018)
- Online on the website Alexa.com (2019) <https://www.alexa.com/topsites>
- Online on the website Alexa.com (2019) <https://www.alexa.com/topsites>
- Online on the website Iso25000.com. <https://iso25000.com/index.php/en/iso-25000-standards/iso-25010?limit=3&start=3> (2018)

- Online on the website Statista. <https://www.statista.com/statistics/219909/global-tablet-penetration-forecast/> & <https://www.statista.com/statistics/284202/mobile-phone-internet-user-penetration-worldwide/> (2018)
- Online on the website statistics.laerd.com, <https://statistics.laerd.com/spss-tutorials/jonckheere-terpstra-test-using-spss-statistics.php> (2019)