UCL Université catholique de Louvain

LOUVAIN School of Management

Louvain School of Management

Analysis of Graphical User Interfaces Aesthetics based on Visual Metrics

Research Master's Thesis submitted by
Charlotte Botermans

With the view of getting the degree in
Master 120 crédits en sciences de gestion, à finalité spécialisée (à finalité approfondie)

Supervisor(s)
Jean Vanderdonckt
Mathieu Zen

Academic Year 2016-2017

I would like to express my sincere gratitude to everyone who have helped me and supported me, in any way, in the redaction of this thesis.

First of all, I would like to thank M Jean Vanderdonckt, professor at the Université catholique de Louvain, for accepting to be my thesis supervisor and to have provided me with all the material, knowledge, motivation and guidance that helped me in the production of this dissertation.

I would also like to thank M Mathieu Zen, Doctor of Philosophy in Management Sciences and Information System at the Université catholique de Louvain, for his availability, support and advice, which helped me to find solutions to move forward.

Finally, I thank my family and friends for their encouragement, constant support, devotion and involvement throughout the whole writing process of this thesis.

Table of content

List of Figures

List of Tables

1. Introduction

1.1. Problem statement

As the world is evolving at a fast pace, companies need to constantly adapt their strategies to their environment, so as to stay afloat and competitive on the market. During the last few decades, businesses have tried to keep up with the growing trend of doing business online. As the market grows it gives birth to new businesses, which can lead to a saturated market. Consequently, to disgorge the market, many companies decide to expand overseas using the internet to drive their activity. This is known as doing e-commerce. Doing business online has become crucial for many companies nowadays. Indeed, it allows organizations to sell their products and/or services worldwide, enabling them to increase their turnover, and broaden their target market. According to Queensland Government (2016), there are many benefits to doing business online; such as providing an improved client service, greater flexibility, global access, but also allowing cost savings. Moreover, doing business online offers many opportunities, such as communicating easier with customers, and thereby getting greater levels of feedback on business activities. It also enables companies to research their competitors, and to offer diverse online services to their clients.

Nonetheless, in order to succeed online, businesses need to develop their very own front-end, which is commonly known as Graphical User Interface (GUI) in computer sciences terms. When doing business online, the main goal of a GUI is to attract and enable customers, as well as potential customers, to buy and research about a company’s products and/or services online. As the trend is being adopted by many businesses around the world, it is fundamental to stand out and design an attractive and intuitive website. It is thus of prime importance to have an aesthetically pleasing, and user-friendly website so that it gets the attention of users.

That is the reason why, for many years, many studies have been conducted to determine to what extent a website can be considered as aesthetically pleasing. That being said, the aesthetics of GUIs remains subjective and hard to measure or quantify; hence the particular focus on this topic in our thesis.

Many studies reveal that aesthetics plays a significant role in the success of an online business. According to Ngo et al. (2000), improving the aesthetics of graphical user interfaces would enhance its usability, and thereby increase its attractiveness. Ngo et al. (2000) also suggest that users are often frustrated and confused when using interfaces, because they must deal with complex layouts, or face a large amount of information. Consequently, customers tend to leave the interface, and look for an easier and more user-friendly website that fits their needs or requests, which clearly shows the importance of aesthetics. Indeed, a complex and unusable graphical interface can hinder online business activity. Zen (2017) and other authors, also claim that aesthetics positively impact usability and user experience. As mention above, a company’s website is the key driver of any online activity; therefore, it is fundamental to provide users with a consistent and clear interface. Although aesthetics appears to be a major touchpoint for online businesses, it remains difficult to measure.

In this thesis, we aim to find a way to measure accurately the aesthetics of graphical user interfaces as this appears to be problematic for many companies to achieve. In fact, studying aesthetics has real importance from the point of view of both marketing as well as business. Consequently, conducting research on this topic aspires to determine the impact of badly designed interfaces and to find the best ways to achieve success when launching or doing business online. This brings us on to the next part of our thesis, namely, the research question in management.

1.1.1. Research question in management

How to measure the aesthetics of Graphical User Interfaces?

According to Hoegg (2010), the aesthetics, or design of any kind, affects the decision-making process of a customer, as visual aspects represent a competitive advantage. Companies understand this, and have therefore largely introduced this standard in their strategies. As aesthetics, design and pleasing visual aspects are more valuable than ever before, and are thus considered as sales assets, businesses have made them of prime importance. Reimann et al. (2010) assert that customers are more prone to buy aesthetically-pleasing products, than those with standardized packaging.

Therefore, this theory is also applicable to any GUI design. According to Cisco's white paper (2017), the number of GUIs keeps growing, and this trend is not about to disappear. In fact, the number of mobile devices per capita will reach 1,5 by 2021, which would overtake the earth's overall population. Nowadays, human beings constantly come across GUIs, therefore, it seems wise strategically to understand how they work and how there are built. Zen (2017) claims that GUIs intend to deliver messages to users. So, in this thesis, we aim to determine the best way to assess the aesthetics of websites (GUIs), and to understand how it affects a company's marketing strategy.

1.2. Research's motivations

1.2.1. Theoretical motivations

Although various studies have already been conducted on the topic of aesthetics, there is still a lack of theory on the topic. The field calls for further and more thorough research to determine precisely which factors really influence the perception of the aesthetics of an interface. As such, despite an abundance of scientific articles providing a broad range of research outcomes, there is very little literature available on this particular subject. In fact, surveys mainly suggest historical evolution or perceptions models but most importantly, they are mainly carried out in other disciplines, such as philosophy, social sciences, or art (Zen and Vanderdonckt, 2016). This implies a lack of theory regarding aesthetics in the domain of human computer interaction, because it continues to be a subjective topic and one which is hard to understand and master. Therefore, researching this topic would help human computer designers to determine what trends to follow, and what is considered as aesthetically pleasing in terms of human computer interaction.

1.2.2. Methodological motivations

There are few theories on aesthetics, and likewise few methodologies to use when analyzing this topic. Van Bremen et al. (1998) explain that computer support of industrial design is still immature, despite the relentless efforts made on trying to improve it. This is particularly notable for aesthetics design. They continue by clarifying that there is no strict methodology including aesthetics features, such as usability, conviviality and a pleasing display, in computer support (Van Bremen et al., 1998). It would be more than relevant to work on establishing a methodology that eases improving computer support of design for aesthetics. There is a strong need for a replicable methodology allowing to assess the perceived aesthetics.

1.2.3. Empirical motivations

Most of the inquiries in human computer interaction have progressively intended to combine traditional concerns and aesthetics aspects. While many attempts have been made on proposing empirical studies, such as market research or quantitative studies on the matter, they have all used different procedures and assessing tools, which do not allow clear conclusions to be drawn. Besides, Lavie and Tractinsky (2004) allege that there is a true lack of suitable concepts and measures regarding aesthetics, which might seriously limit future studies in this field. This is the reason why there is a need for carrying out research in order to establish an accurate tool to assess and measure perceived aesthetics of websites.

1.2.4. Observational motivations

There are clearly observational motivations as well. As mentioned earlier, most people are confronted with graphical interfaces on a daily basis, and this trend is not decreasing. Scientists have already surveyed the topic using an observational approach, and it appears that aesthetics is of great importance to many of us. Lavie and Tractinsky (2004) argue that “the visual aesthetics of computer interfaces is a strong determinant of users’ satisfaction and pleasure” (Lavie and Tractinsky, 2014, p.1). This illustrates the need for investigating and developing a measurement method of aesthetics in human computer interaction.

1.2.5. Management motivations

In the growing context of the digital era, GUIs keep increasing their added value for companies. In fact, companies must keep up with the newest trends, otherwise they could fall behind. Myers and Rossom (1992) assert that user interfaces account for a large part of software development; hence, the indispensability for businesses to enhance their software development skills. It is the responsibility of managers to ensure that their employees acquire the right experience, and get trained according to market trends. Myers and Rossom (1992) also claim that user interfaces are designed using prototype testing, which is a slow process. Therefore, boosting this process would enable designers to speed up software development, and to provide more sophisticated outputs. Thus, assessing the aesthetics of GUIs would definitely help improve the overall design process, and, as such, managers must be aware of this.

1.2.6. Marketing motivations

Along with management motivations, there are also marketing motivations. Indeed, marketing is a continuously evolving scientific matter and it remains quite difficult to determine a global marketing strategy, as every organization requires different needs, while the requests of consumers are becoming increasingly more precise, complex and tailored. Trying to understand how a company’s marketing strategy could be improved represents an everyday challenge. Thus, studying the aesthetics of GUIs will help in developing more efficient and visually pleasing websites for organizations doing business online, and thereby improving their overall online activity but most importantly their marketing impact. In fact, aesthetics, as described in the introduction, really matters in terms of the likes of marketing and product design. The visual appeal of a product tends to affect consumers’ opinions and attitudes, regardless of its quality. Moreover, various studies indicate that user satisfaction varies according to the visual aesthetics to which they are exposed when going online. Consequently, users’ impressions are crucial. Thus, well-designed user interfaces are more prone to satisfy better their users than standardized interfaces.

However, aesthetics remains a controversial concept. Indeed, many people question whether aesthetics can even be measured at all. Aesthetics is often considered to be highly subjective as it depends on perception. For many years, people have been using a single specific number, which is an irrational mathematical constant reaching approximately 1,61803398874989, also known as the golden ratio, to define the beautifulness of something.

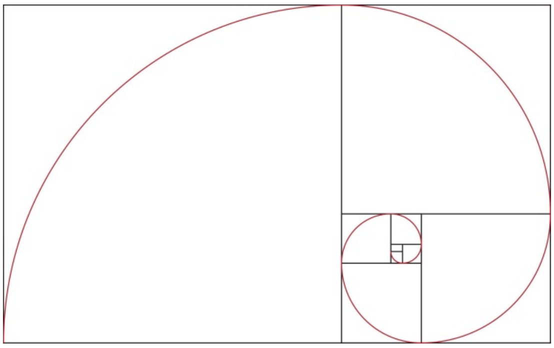


Figure 1 - Representation of the Golden ratio

Source: Wolfram Mathworld. (2017). *Golden ratio*. Available on: http://mathworld.wolfram.com/GoldenRatio.html

This ratio was used as a tool to determine the beauty of something. In fact, people believed that this ratio referred to pleasing things from an aesthetics point of view. However, scientifically speaking it seems interesting to participate in researching a new and accurate way to assess aesthetics, as it is widely relied upon to stand out in the market nowadays, regardless of the sector. On top of that, although the golden ratio has been used profusely as a reference over the years, it remains an irrational constant, and irrational numbers are incommensurable. Besides, there is a main empirical issue in using the golden ratio, it has never been demonstrated as being a guarantee of aesthetics. Thus, as the market and business both evolve in a very rational context, researching new and more rational assessment methods seems logical.

1.3. Contextualization of the subject

It is fundamental to define and determine the centering that we will operate for our research. Indeed, studying the impact of the aesthetics of a product, digital or not, or even of a service, is a highly complex topic because it is a multi-factorial phenomenon. Therefore, we must delimit the scope of our study. To do so, we will proceed to a series of funnels, starting from a general overview of the current market, before narrowing it down gradually. This will help us constrain the coverage of the issue under review, so as to then tackle our main focal point.

The global goods and services market grows at a very fast pace, as does competition therein. Consequently, selling, positioning and developing products and services worldwide are all crucial. In fact, according to Hofstede et al. (1999), “the globalization of the marketplace is arguably the most important challenge facing companies today” (Hofstede et al., 1999, para. 1). Thus, everything goes global and trespasses boundaries; there is a free flow of information on the Internet, while many advances in telecommunication technologies have been made, alongside the emergence of global media (Hofstede et al., 1999). Companies must differentiate themselves because the international environment has reached a high level of competitiveness. Moreover, this internationalization process strongly influences customers’ preferences, which have shifted toward greater interests in international products and services. This illustrates the fact that the market is being digitalized even more than ever before.

Computer systems play an increasingly important role in our society. In the space of about seventy years, computers have rapidly improved and democratized, as society today is increasingly dependent on computer systems. Therefore, companies intend to develop vast computer systems in order to automate the processing of information. Computer systems are based on two major components: a physical material, known as a computer and software, which precisely enables data to be processed (Bonaventure et al., 2015). Thus, in this digital era, companies seek to diffuse information technology across all types of businesses, suppliers, distributors, etc. The main concern is how to effectively implement communications and information technologies (Aldrich, 1999). This enhances the importance of online business, which enables companies to reach customers worldwide, provides them with goods and services regardless of their locations, and collects priceless information about their attitudes and desires. In this evolving digital economy, technology is regarded as the prevailing strength. In fact, as it is information-driven, information technology (computer system) has positioned itself as a key success factor in the enlarging market (Aldrich, 1999). Laudon and Laudon (2004) also claim that "in many industries survival and even existence without the extensive use of IT is inconceivable" (Laudon and Laudon, 2004, p.31).

This leads us to another fundamental concept for businesses nowadays: the information system, which stems from computer systems. Information systems are considered essential for organizations to function. Laudon and Laudon (2004) argue, that "Information systems knowledge is essential for creating successful, competitive firms, managing global corporations, adding business value and providing useful products and service to customers" (Laudon and Laudon, 2004, p.XIX). Servigne (2017) defines it as a set of resources, such as personnel, software, processes, data, computer and telecommunication equipment that enables gathering, stocking, structuring, modeling, exchanging and diffusing information, such as texts, images, sounds, or videos, within an organization. Among those resources, which include hardware and software, we can find database management systems, customer management tools, Internet, Intranets and Extranets, but also application or presentation servers, such as websites. Information systems are complex and can vary along with companies' needs, which result in different system architectures.

As mentioned above, since information systems are becoming essential, the importance of websites is thus too. Indeed, having a web presence is now considered as key for success. Developing a business online or having a web presence requires designing a website (Sano, 1997). Rosen and Purinton (2004) explain that web content includes "texts, pictures, graphics, layout, sounds, motions and, someday, even smell" (Rosen and Purinton, 2004, p.1). This illustrates the complexity of choosing the right content and the amount of information needed to produce an effective web interface design. Moreover, although some may think there is a mainstream type of website customization represents in fact a major aspect (Thongpapanl and Ashraf, 2011). Rosen and Purinton also argue that a balanced and usable website enhances customer satisfaction along with their purchasing behavior. Chen and Ryu (2013) explain that the conception of websites has always been complex, because it requires finding the right balance between a pleasing, user-friendly and effective architecture. It should be noted that there are different types of websites, such as graphical user interface websites and virtual reality websites; however, we will focus our analysis on graphical user interface of websites only.

A website relies on, and functions through a web interface, which is also known as a graphical user interface. Galitz (2007) argues that the key component of any computer system remains the user interface. The main purpose of an interface is to ease users' interaction with computers. Therefore, it is of prime importance to produce a good design because the interface is the visible part of a computer and its software. Galitz (2007) continues, in that "the user interface has essentially two components: input and output" (Galitz, 2007, p.4). Output refers to the ability of a computer to transmit its results of computation to the user, while input refers to the way the needs and desires of the users are delivered to the computer. According to Galitz (2007), great interfaces go unnoticed in the eyes of users. This emphasizes the significance of well-designed interfaces, which obviously include aesthetics.

This analysis brings us to the main issue on which our thesis is based: measuring the aesthetics of these graphical user interfaces. In order to soak up and understand the topic and all its concepts but also to identify pathways, we consulted previous studies. We thus came across Reinecke's study (Reinecke et al., 2013), which looked at the aesthetics of a sample of 430 graphical user interfaces, using human reviews. We also discovered that other studies had been conducted using other samples and techniques but we believed that using Reinecke's sample was more appropriate and relevant in our case. Besides, instead of taking Reinecke's corpus, we could also have constituted our own corpus of screens based on various criteria, or even constitute a corpus based on an existing base. For example, we could have considered the first 100 top websites from http://www.alexa.com/topsites, but It would have been complex to control and verify the representativeness of the sample. Reinecke's corpus has the advantage of being empirically validated and giving a strong reference.

We must keep in mind that the current matter is to make the internet and its interfaces user-friendly, and by this we mean satisfying users' needs and requests, as well as easing their interactions with computers. This highlights the importance of the qualities of the aesthetics of graphical interfaces, but also of the computer software. There are various norms, such as the ISO/IEC 25000, also known as SQuaRE (System and Software Quality Requirements and Evaluation) that aim to create scope for assessing the quality of a software product. The ISO 25000 is in fact the improved version of other previous standards, and gathers five different sections of new standards, one of which is the ISO/IEC 2501n quality model division. This suggests specific standards for assessing computer systems, software products, quality in use, and data (ISO25000, 2017). The ISO/IEC 25010 consists of eight criteria which are the following: functional suitability, performance efficiency, compatibility, usability, reliability, security, maintainability, and portability. The figure below illustrates the ISO/IEC 25010 model.

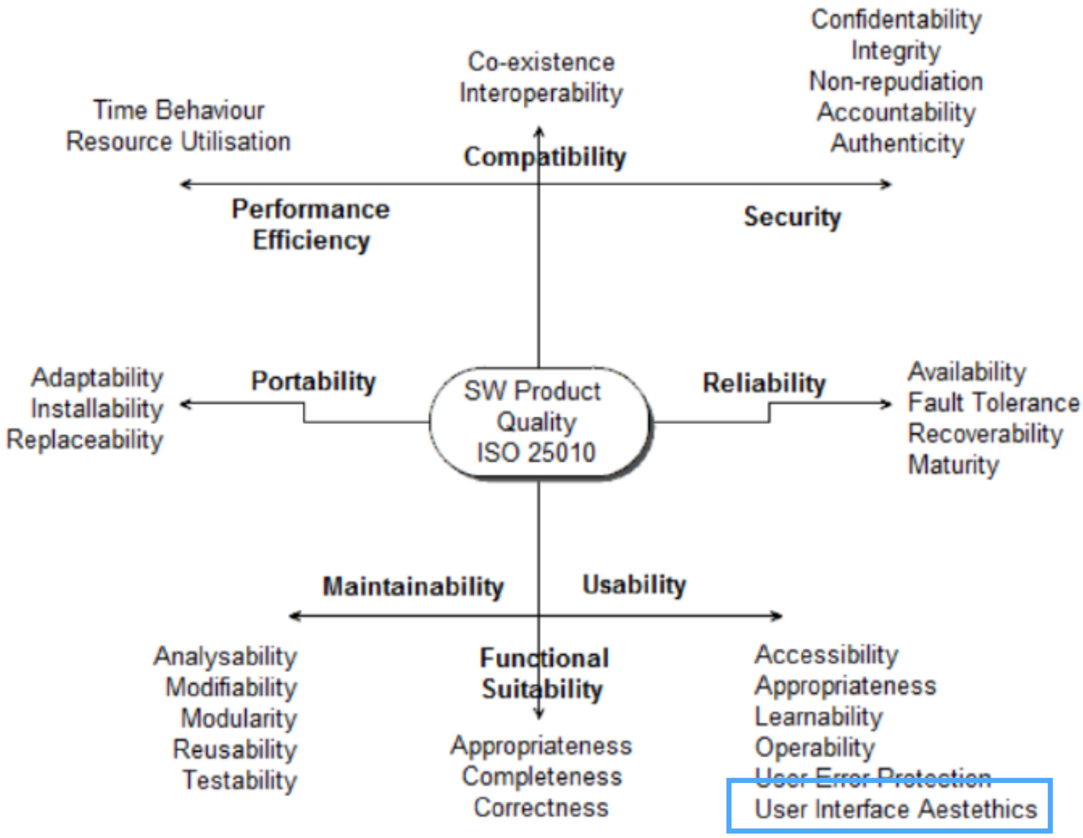


Figure 2 - ISO/IEC 25010[1]

[1] Remodeled version of the ISO/IEC25010 model available on http://iso25000.com/index.php/en/iso-25000-standards/iso-25010.

It should be noted that these factors are equally prevalent, and classify the product/software quality into sub-features. Each factor evaluates specific features that are considered as key indicators for rating software/product quality. For our research, we will focus on the usability. Indeed, this factor includes a major sub-characteristic: user interface aesthetics, which is the inflection point of our analysis.

Nowadays, use interface aesthetics is regarded as an essential matter, because several studies have demonstrated that it has played an important role in influencing users' attitudes and experiences. According to Jordan (1998), aesthetics is a real source of delight for users during their interaction. Besides, one study revealed that the perception and preferences of a website were predicted by its beauty (Schenkman and Johnson, 2000); that being said, Lavie and Tractinsky (2004) claim that there is a true lack of suitable concepts and measures of aesthetics. As a matter of fact, aesthetics remains a complex concept; something hard to master or to assess. Nonetheless, previous research supports the idea that aesthetics can be evaluated by following two different methods: the subjective process and the objective process. Most previous studies were conducted using the subjective method because it is easier to implement: it does not require the involvement of specific metrics as it mainly relies on human reviews. The objective method, on the contrary, relies on complicated formulas and algorithms. There are different types of measures to assess aesthetics objectively. There are geometric measures, which are based on linking shapes to users' emotional responses, as well as statistical measures, which are based on computing statistics.

The main purpose of our analysis though, is to determine a precise measurement method for assessing user interface aesthetics. We aim to suggest a measurement somewhere in-between subjective and objective in order to procure the most accurate and balanced results. Therefore, we will focus, in our study, on geometric measures from which stem visual measures, and compare them with subjective measures. This thus leads us to the use of the dedicated applicative software QUESTIM. Indeed, this software provides visual measures such as the balance, the complexity, the equilibrium and so on.

In order to illustrate and summarize best the context of this analysis, we developed the graph below.

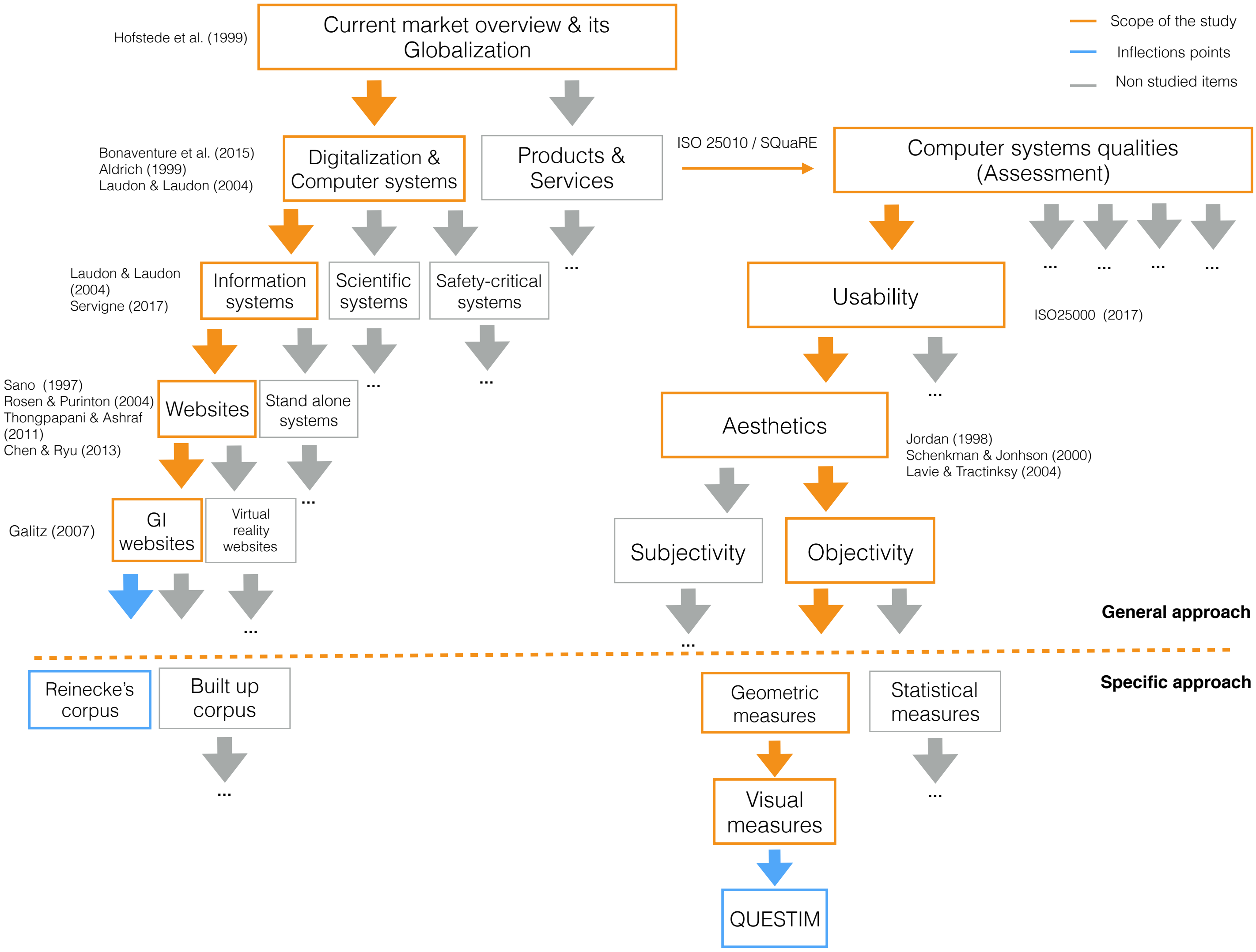


Figure 3 - Framework of the analysis

We can summarize this section, by stating that the heart of our thesis is to address the design procedure of GUIs. We aim to assess GUIs with specific measures (computed with formulas) that consider both the graphical aspects and the disposition of the visual component of the interfaces. It should be noted that we will restrict our research to human computer interaction designs, and thus no other designs will be examined.

In addition, our study primarily targets user interface designers. Indeed, we intend to offer them a new approach in assessing their work. This, of course, also implicitly includes users, as well as companies, as the outcomes of our study could respectively ease their use, but also help their online businesses flourish.

1.4. Thesis structure

This thesis is divided into two distinct parts. The first part focuses on the research question in management, the problematic and proposes a literature review. The second part presents the statistical analysis and its results, which were obtained through a comparative study, using a quantitative approach. In this second part, we present the results and conclusions drawn from our statistical analysis, and the limits of our study. We also suggest new perspectives for future analysis.

The first chapter introduces the problematic and details the different motivations for our research.

The second chapter reviews the existing literature covering the topic under review. In this section, we first define the different concepts of our research question in management, which are the following: aesthetics and graphical user interfaces. In addition, we highlight the impact of the aesthetics on usability and trace the historical evolution of aesthetics up until today. We also focus and develop the theories that were used for our research. Then, we highlight the importance of aesthetics from the point of view of marketing, as well as from business. The last point we address, is the complexity of measuring aesthetics as it remains a highly subjective assessment. In this segment, we interpret the model of aesthetics perception.

The third chapter explains the model and hypothesis that we developed for our statistical analysis.

The fourth chapter describes the methodology we used throughout our statistical analysis.

14.

The fifth chapter focuses on the statistical analysis, and presents the SPSS outcomes and tables.

The sixth chapter presents the results obtained through our analysis, with which validates or refutes our different set of hypotheses.

The seventh chapter describes the discussions and the interpretations of the results, and also suggests new perspectives for future analysis.

The eighth and final chapter concludes our analysis by highlighting different limits of our research.

2. Literature review

Before going through any further analysis, it is fundamental to define precisely the main concepts of our research question in management, which deal with the measurement of the aesthetics of GUIs. This way, it allows us to illustrate and provide a better insight into our problematic. Consequently, we reviewed the literature covering the following concepts:

– aesthetics;
– Graphical User Interfaces.

2.1. Aesthetics

2.1.1. Historical evolution of the term 'aesthetics'

The aesthetics find its origins in the late 18th century, from the Greek "aisthētikos", in the sense 'relating to perception by the senses'. Later, the sense evolved towards the German meaning 'concerned of beauty'. This very same sense was endorsed in English in the early 19th century; however, its use remained moot for many years (Online Oxford Dictionaries, 2017). Aesthetics derives from various meanings as it has been considered from different perspectives throughout the years. Consequently, the term has also had different degrees of importance.

Historically and for a long period of time, the values of aesthetics were considered as the conception of beauty. According to Feagin and Maynard (1997), beauty and its perception were of prime concern. Moreover, Lavie and Tractinsky (2004) assert that the word 'beauty' is usually used to describe things that are pleasing, regardless of their nature. The concept kept evolving and its meaning was, at first, very often combined with order, but in contemporary literature, pleasure and perception are the main qualifiers of the aesthetic interpretation of beauty (Feagin and Maynard, 1997). For a long time, the state of beauty was associated with the state of nature; thereby, appealing conceptions had to pertain to the principles of natural beauty (Kruft, 1994).

The concept of aesthetics, in the sense of beauty has recently been introduced in Western thought. Indeed, the term 'aesthetics' was slowly instituted, and Baumgarten was the first philosopher to publish a philosophical study of art and beauty in the 18th century (Saw & Osborn, 1968). In his research, he claims that the perfection of sensory awareness refers to the perception of beauty. From that moment, the term ‘aesthetics’ was introduced into general language, and no longer considered as a philosophic technical term. The roots of the standard meaning of beauty - and thereby the standard meaning of the aesthetics - were examined using two distinct approaches: humanistic (Porteous, 1997) and empirical (Swede, 1994). They both attempted to determine a generic idea of aesthetics, and were based on previous research on works of art; however, many other frameworks have also been considered for the study of aesthetics.

Although there are various interpretations of aesthetics, in this thesis we will focus on the meaning that refers to “concerned with beauty or the appreciation of beauty” (Online Oxford Dictionaries, 2017, para. 1), or to being “pleasing in appearance” (Merriam-Webster Dictionary, 2017, para. 1).

2.1.2. Definition of the aesthetics

Now that we have better knowledge of the historical background of the aesthetics, let us move on to its definition. As mentioned earlier, the term ‘aesthetics’ has much evolved throughout the years, and in this section, we will provide, in chronological order, some authors’ interpretations thereof.

In 1750, Baumgarten suggested that “The aim of aesthetics is the perfection of sensible cognition as such, that is, beauty, while its imperfection as such, that is, ugliness, is to be avoided” (cited by Gregor, 1983, p.357). Baumgarten was a pioneer, and created a new branch of philosophy with its various conceptions of aesthetics, which rely on the essence of art, taste and beauty, along with its valuation.

Later on, Clay (1908) suggested that aesthetics was used as an evaluation criterion for assessing appealing things, and therefore, should be considered as a value judgement for delightful things only. This allows us to state that aesthetics depends on the resemblance of things to a preset ideal standard.

Then, in 1933, Birkhoff came up with a mathematical theory of aesthetics. According to him, there are two categories for aesthetics: the natural beauty and the artificial beauty, which is a creation of the artist. He believed that aesthetics relies on artificial creations.

Sixty years later, van Damme (1996) introduced the concept of universal aesthetics, which implies the existence of a shared standard of beauty among everyone. He conducted his study on characteristics of forms, and concluded that some of these aspects could be considered as aesthetics principles across all cultures around the world.

Finally, Lavie and Tractinsky (2004) proposed a different approach to define aesthetics. Indeed, they intended to evaluate the perceived visual aesthetics of web pages through various dimensions. This study enabled them to claim that the aesthetics perception could be considered through two dimensions: “classical aesthetics”, which hints at beauty standards, and “expressive aesthetics”, which englobes more than the classical principles.

These different qualifications and approaches provide us with a more global definition of the aesthetics, but it remains a little vague still. Moshagen and Thielsch (2010), suggest the following broad interpretation: “the immediate pleasurable experience that is desired toward an object and not mediated by intervening reasoning” (Moshagen and Tielsch, 2010, p.3).

2.2. Graphical User Interface (GUI)

2.2.1. Historical evolution of Graphical User Interface (GUI)

The very first GUI was born in the late seventies at the Xerox Palo Alto Research Center. Indeed, some researchers conceived the first application endowed with a GUI, namely, the Xerox Star, and proceeded into two designing phases. They designed the computer human interface beforehand, and then worked on building the internal workings of the application, which made the Xerox Star incomparable. Nevertheless, it turned out to be too slow and not commercially successful (Jansen, 1998).

Then, after visiting the Xerox Palo Alto Research Center, Steve Jobs intended to develop his own GUI, with the help of some of the original designers of the Xerox Star. His GUI was known as the Apple Lisa. However, this first attempt resulted in failure as, similarly to the Xerox Star, the Apple Lisa was not commercially successful either. Thus, in 1984, his team produced the Apple Macintosh, which, this time, led to success, and is nowadays used as a reference in terms of the look and feel for all GUIs (Jansen, 1998).

The Apple Macintosh, which was the first to introduce a menu, icons etc., the IBM System Application Architecture (SAA), which is an entire system of interfaces including database tools, networking tools etc., and the X-Windowing System, which has application tools and utilities, are considered the models to follow for all modern GUIs (Jansen, 1998).

2.2.2. Definition of Graphical User Interface

With a better insight into the historical evolution of GUIs, let us now focus on its definition. To do so, we will provide various definitions of the term, in order then find the most accurate definition possible.

According to Van Hoff (1998), "GUIs are well known mechanisms by which users can interact with computer programs. A typical GUI provides windows and/or dialog boxes that enable a user to initiate an operation by the computer program on the user's computer" (Van Hoff, 1998, para.3). This allows us to state that the GUI must provide users with an operating structure thus indicating to them the path to follow.

Norman (2013) continues by explaining that "the design implications are clear: provide meaningful structures. […] This is the power of the traditional graphical user interface with its old-fashioned menu structure" (Norman, 2013, p.100). Meanwhile, Jansen (1998) defines the GUI as follows: "A GUI is a type of computer human interface on a computer. It solves the blank screen problem that confronted early computer users" (Jansen, 1998, para. 6).

Bonsiepe (1990) suggests the following interpretation from a more conceptual level, as the "means by which people and computers communicate with each other" (Bonsiepe, 1990, p.264). However, from a computer science point of view, the GUI should be specified as a visual performing platform that is displayed to computer users (Harding, 1989). Bonsiepe (1990) adds, that “a GUI is a specification for the look and feel of the computer system” (Bonsiepe, 1990, p.265).

Finally, Zen (2017) describes the GUI as “a specialization of a logical user interface that takes the form of a set of graphical components (icons, visual indicators, navigation bars, etc.). Its main goal is to provide the system user with an interaction modality that puts forward a visual logic. It is often opposed to text-based interfaces that are, by contrast, designed with syntactical rules” (Zen, 2017, p.23).

This being said, despite the profusion of GUIs on the current market, its exact definition still remains fuzzy. This can be explained by its novelty, since it has been introduced on the market only quite recently, in the early eighties more precisely. Nonetheless, it is important to note that regardless of the given definition, GUIs always present common attributes, such as WIMP (windows, icons, menus and push-buttons), and that they are usually composed of three major items (a windowing system, an imaging model, and an application program) (Hayes, 1989).

2.3. Importance of the aesthetics

For a long time, efficiency was the sacred word in terms of human-computer interaction. Consequently, human computer scientists’ research was essentially efficiency-driven, and all other criteria were left out. Neale and McCombe (1997) precisely emphasize the functional and usable aspect of the design, when explaining how to create and build a website. Indeed, the idea of a smooth user experience was clearly not the primary issue at that time. According to Tractinsky et al. (2000), the reason for this is two-fold: on the one hand, there has been a blockade reaction to the exaggeration of fashions in new tendencies, while on the other, there was a strict focus on effectualness. So basically, the real matter was the effectiveness of a program regardless of its ease of use or appearance. The major assets of this kind of program were its rapidity of execution and its impressive capacity of combination of outputs.

Nevertheless, this monocentric strategy led to the omission of other significant features, such as the aesthetics (Lavie & Tractinsky, 2004). But in reality, the aesthetics, which was, in the past, referred to as 'beauty' has had its relevance for thousands of years. Indeed, Vitruvius was the very first theoretician of architecture to have considered 'beauty' as a prerequisite of this art (Kruft, 1994). On top of that, Coleman (2007) re-uses Leon Battista Alberti's words to define beauty: "Beauty is that reasoned harmony of all the parts within a body, so that nothing may be added, taken away, or altered, but for the worse. It is a great and holy matter" (cited by Coleman, 2007, p.30).

In many disciplines, the aesthetics has become a fundamental matter, and has shown its importance among people's daily interactions and attitudes towards each other. Indeed, Porteous (1997) claims that the aesthetics influences peoples' feelings. Besides, various authors assert that it also impacts new product development, marketing strategies, and the retail environment (Russell and Pratt, 1980; Russell, 1988; Kotler and Rath 1984; Whitney, 1988). Meanwhile, Bloch (1995) adds the following: ''physical form or design of a product is an unquestioned determinant of its marketplace success'' (Bloch, 1995, p.16). In sum, beauty represents a major advantage for product attractiveness, and many studies have found a relationship between an interface's aesthetics and its perceived usability (Tractinsky, 2000).

Indeed, as mentioned previously in the introduction of this thesis, the visual appearance of an object plays a fundamental role in customers' attitudes. Various studies have shown design to be of prime importance in terms of differentiation. Many customers are more willing to buy a product that is visually pleasing at a higher price, rather than a product offering commonplace packaging at a lower price (Reimann et al., 2010). This design theory is thus relevant and valid for GUIs.

Different studies in human-computer interaction have demonstrated that the aesthetics of computing items have played a significant role in forming people's behavior (Kurosu and Kashimura, 1995; Tractinsky, 1997; Tractinsky et al., 2000). According to Jordan (1998), aesthetics is a real source of delight for users during the interaction, which also applies to other web-related contexts (Schenkman and Jonsson, 2000; van der Heijden, 2003). One study revealed that the perception and preferences of a website were predicted by its beauty (Schenkman and Johnson, 2000). Besides, van der Heijden (2003) supports the notion that the visual aspect of a website clearly influences three main elements: the user's experience, the usability and the usefulness.

Thus, as explained in the previous section, GUIs have evolved considerably throughout the years, making them a lot more complex than ever before. But since GUIs' main purpose is to deliver a message, the content of a page is obviously fundamental; however, the layout and the form account for the same level of importance (Zen, 2017). In the eighties, when microcomputers began to commercialize, scientists realized the necessity for exploring the aesthetics of both, hardware and software. Forgaty et al. (2001) explain that the use of computer technology has risen, and therefore, specifications have changed accordingly. Besides, studies have demonstrated that a user's perception of a system's usefulness and usability depends on its perceived quality (Zhang and Li, 2004). Indeed, visual designs of hardware and user interfaces decidedly impact usability performance and user experience and fulfillment. Thus, the aesthetics and visual appeal are now considered as intrinsic components of an interactive systems design (Forgaty et al., 2000). Investigating the aesthetics of GUIs could potentially lead to an improvement in users' perception and thereby ease their everyday journey online; however, the lack of human-computer interaction references on aesthetics illustrates the fact that this feature has been ignored in the discipline in comparison to others.

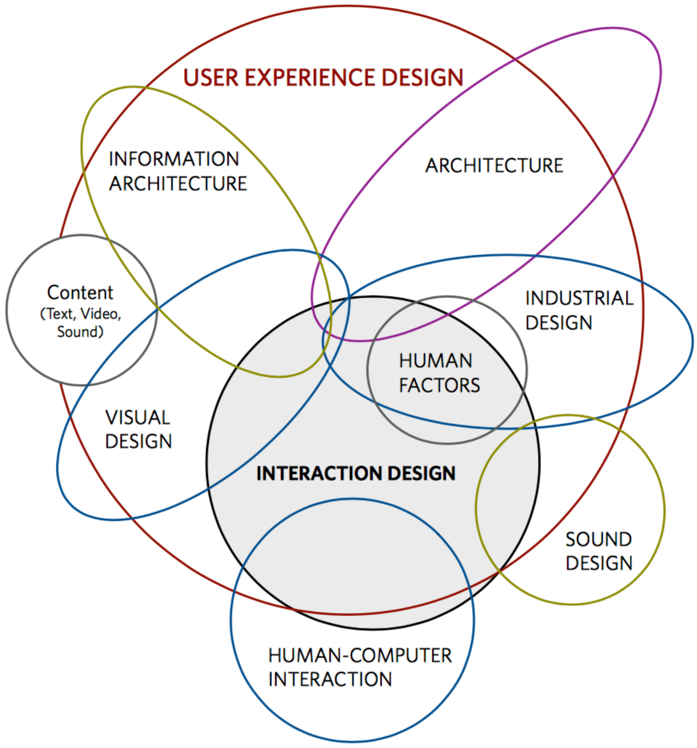


Figure 4 - Illustration of the relationships between the disciplines

Source: Saffer, D. (2010). *Designing for interaction: creating innovative applications and devices*. New Riders, p.21.

Figure 3 above, intends to highlight the relationship among recent disciplines. Most of them are related to each other and appear to be part of the user experience design. For instance, the main goal of visual design is to build and develop a visual language to transmit content, such as colors, font, and layout of user interfaces. Interaction design and human-computer interaction are similar at all points, except for the methods they use. Indeed, human-computer interaction refers to more quantitative data; however, the main aspect to understand here, is that all the disciplines are interconnected with interaction design, and that each time they overlap with one other, there is further ground for study and practice. For example, interface design could be the point where the visual and interaction design meets. So, there should be emphasis on the visual aspect of user interfaces, since this aspect has been largely neglected for too long.

Tractinsky (2000) suggests that the blatant disregard for aesthetical aspects in human computer interaction is regrettable on different levels. Indeed, it first highlights an inconsistency between the practice and research elements of human computer interaction regarding conception guidelines. Furthermore, studies reveal that the needs of interface users are clearly overlooked. Customers of computing items have similar needs to common customers; hence, they present the same valorization for aesthetics as well as usability of products (Darden and Babin, 1994; Jordan, 1998). As researchers have demonstrated that there is a relationship between the perception of aesthetics and usability, users usually consider both dimensions simultaneously.

2.4. Complexity of measuring the aesthetics

Now that we have a better knowledge of the importance of the aesthetics in human computer interaction, let us examine what procedures enable aesthetics to be measured precisely.

As aesthetics has become an integral part of contemporary design processes, it is fundamental to measure and assess this dimension as accurately as possible. This is to make sure new releases fit customers' expectations, but also that they are in accordance with new aesthetics dogmas. Nonetheless, according to Zen (2017), the aesthetics is mostly deemed a subjective issue; hence its recurring association with the following concepts: "visual design", "beauty", "pleasing", which are predominately abstract concepts. Gardner (2002) also argues that aesthetics is characterized by being immediately distinctive, emotional and contemplative. Broadly speaking, according to Moshagen and Thielsch (2010), the aesthetics refers to "the immediate pleasurable experience that is desired toward an object and not mediated by intervening reasoning" (Moshagen and Tielsch, 2010, p.3); while in human computer interaction, it invokes the attractiveness and loveliness of an interactive user interface (Robins and Holmes, 2008) (Tractinsky et al, 2000).

2.4.1. Aesthetics perception

The process of aesthetic perception is still complex to examine. In fact, many approaches have been developed throughout the years to investigate the perception of this dimension. According to some authors (Clay, 1908; Leder et al., 2004; Norman, 2005; Tuch et al., 2010), when facing a new design, people's attitudes and emotions differ. Norman (2005) suggests that there are three different stages in forming an opinion on designs, which are the following: visceral, behavioral and reflective. First, the visceral stage takes place once the user is exposed to the design; or in other words, when the user makes his judgement instantly. Then comes the behavioral stage, which is more cognitive and happens when the user interacts with the design. The last stage, which is the reflective stage, is characterized by the final judgment, which is when the user decides how he feels about the design, and whether he will continue to use it or not, depending on what he experienced. There are thus elements of emotion and satisfaction that play a role in this process.

On top of that, Tuch et al. (2012) studied the aesthetic perception of user interface by conceiving an assessment procedure. Figure 4 illustrates the model of aesthetics perception proposed by Tuch et al. (2012), which is an adapted version of the one thought by Leder et al. (2004). Although this model was conceived for analyzing attitudes toward art, it may be valid for human computer interaction studies.

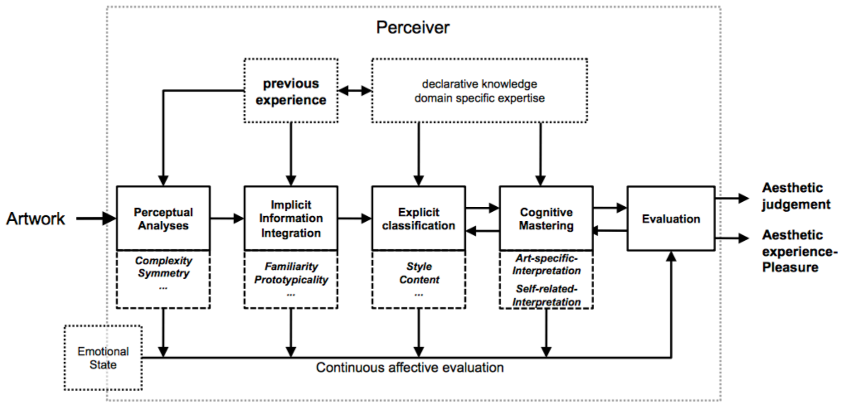


Figure 5 - Model of aesthetics perception

Source: Tuch, A. N., Presslaber, E. E., StöCklin, M., Opwis, K., & Bargas-Avila, J. A. (2012). The role of visual complexity and prototypicality regarding first impression of websites: Working towards understanding aesthetic judgments. *International Journal of Human-Computer Studies*, *70*(11), p.5.

Thus, as shown in figure 4 this approach consists of a five-step process. The first two phases, perceptual analysis and implicit information integration, appear to be more intuitive, as they only reach a basic cognitive level. This implies that through those two steps, the user adopts an unconscious procedure to develop an idea on what they are confronted with. The perceptual analysis is relatively prompt as it consists of superficially scanning the design, such as the color, complexity, balance and symmetry; while the implicit information integration serves in integrating the information of the object/design, and relating its characteristics to previous experiences. In sum, the user seeks for familiarity and prototypicality. Corneille et al. (2005) have demonstrated in a study that attractive faces were more likely to be classified as familiar. We assume that this model is valid for all type of objects and therefore could also apply in this case. Indeed, as user seeks familiarity, if one is confronted to a pleasing object, he would detect familiarity and thus, would be more prone to perceive the object as aesthetic.

The three remaining steps, on the contrary, are qualified as more cognitive, as they require the perceiver to refer to their expertise and knowledge. The explicit classification consists of examining in depth the features of the design, such as its content or style. The precision of the perceiver's description will depend on their knowledge. Cognitive mastering refers to interpreting cognitive properties. The last step consists of assessing the gathered information and determining if the object/design is likable or not (Zen, 2017; Leder et al., 2004; Tuch et al., 2012).

This evaluating procedure leads to two different possible outcomes: aesthetic judgment or aesthetic experience pleasure. The aesthetic experience pleasure results from an instant subjective and positive experience; while the aesthetics judgment requires more analysis, using normative criteria, and deciding whether the object/design should be considered as aesthetic in the sense of art (Leder et al., 2004: Tuch et al., 2012). These two final stages are not inevitably connected. In other words, one can identify an object/design as being aesthetically pleasing in the sense of art, but not being positively influenced by it. Thus, the aesthetics evaluation procedure remains highly personal and subjective.

This model clearly highlights the real problem of aesthetics perception and evaluation that we intend to address in the thesis. As such, the whole process is based on one's own cognitive analysis, which is related to personal knowledge and impressions. This thus involves the emotional state of the perceiver. Depending on one's own mental health and disposition, the perception can vary. Therefore, the evaluation of a design will differ among people, and will always remain subjective and individual. Nevertheless, the aesthetics of an interface should not be regarded as a set of preferences that only designers can master. This vision is too restrictive. This is the reason, why we want to be able to provide a different assessment procedure, which would consider human perception and metric computations. By comparing the outcomes resulting from both approaches, it will enable us to see which features are the most relevant, and which should be considered when designing interfaces. This is illustrated in the graph below.

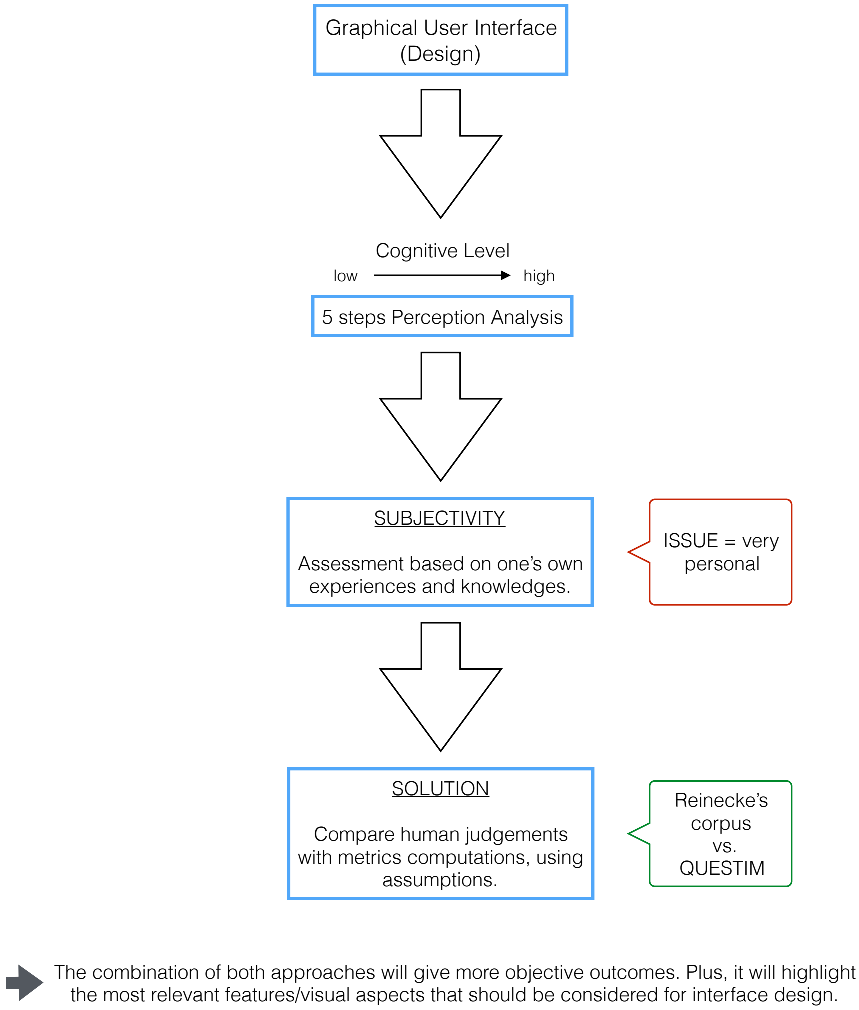


Figure 6 - Representation of the issue

2.4.2. Subjectivity versus Objectivity

That being said, the question of whether the aesthetic can be measured or not remains very controversial. From a rational point of view, this interrogation is doubtful, as the aesthetic appeal continues to be considered as a very personal and subjective matter, and which also depends on one's own cultural background, education, personality, gender, etc.

According to Rantalla (1994), this idea, in fact, refers to two well-known concepts in the field of philosophy: aesthetic subjectivism and relativism. These notions state that an object does not possess any aesthetics characteristics on its own; but rather that it is the perceiver who creates an object's features using their own personal perception. Aesthetics subjectivism thus argues that aesthetic aspects depend on one's own preferences; while relativism suggests that aesthetic aspects depend on one's own culture and not on personal perception. But most importantly, Rantalla (1994) says that one's own subjective perception is clearly influenced by their cultural background.

However, the reverse idea, known as objectivism, claims that aesthetic features emanate from objects, rather than from the individual's perception and culture. And as aesthetic perception relies on aesthetic features, this idea suggests that our evaluation procedure is based on both the object's characteristics and our knowledge/judgment. This implies that there is a universal beauty, or aesthetic features that are significant all around the world regardless of culture; and that we thus examine objects using universal models that are considered pretty.

We aim to assess the aesthetics of interactive user interfaces, and thus must find the best way to do so. It is important to note that for a good evaluation, the main inquiry resides in determining "how appealing" the design is. This involves two antagonist methods: the objectivist and subjectivist view. We must remember that an evaluation is, by nature, an objective method providing quantitative data and results (Zen, 2017); while the subjectivist view refers to human evaluation, which is considered as subjective as it provides qualitative data and results. So, the main concern here is to find a way to compound these two opponent processes in order to get the best assessment possible.

Kant (1987) suggests one solution to bridge the gap between the two methods: inter-subjectivity. In his theory, he states that all human being can think and envisage peers' judgments together, which, when integrated forms the final opinion. This means that any judgment is independent of time and place, because it actually depends on one's own judgment combined with those of their peers. This process thus aggregates a set of subjective opinions to produce a more objective idea, illustrated in figure 5 (below).

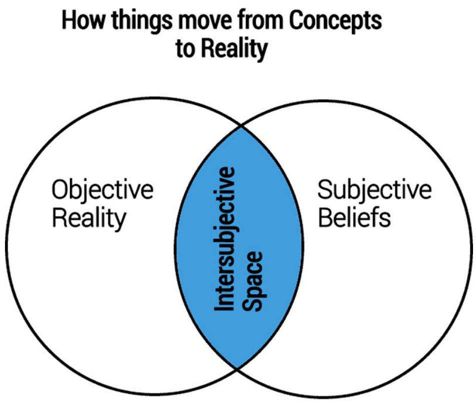


Figure 7 - Illustration of the inter-subjectivity

Source: Amerland, D. (2015). *Why Branding is the Space Where Your Business Values Meet Your Customers' Minds?* Available on: https://davidamerland.com/seo-blog/1052-why-branding-is-the-space-where-your-business-values-meet-your-customers-minds.html

Thus, this can be valid for the assessment of user interfaces. The inter-subjective user experience seeks to collect subjective judgments about the aesthetics of user interfaces, in order to afterwards generate an objective overall evaluation, approved by all users. This will enable quantitative data to be yielded. In other words, to survey the aesthetics of user interfaces on a large population will, in fact, result in assembling subjective opinions that will together form an objective judgment. Thus, this allows us to assert that some aesthetic features may be measurable, if not as a whole (Zen, 2017).

2.4.3. Outcomes of previous studies

Throughout the years, many studies have been conducted to determine what the relevant aesthetic features were to take into consideration with regard user interface design. These studies revealed that the layout of a screen was of prime importance; however, this aspect has been neglected in human computer studies. To create and develop an ideal user interface, there is a need for a more balanced approach.

Ngo et al. (2000) claim that "the aesthetics may not be the only solution, but it definitely plays a significant role in the overall value of a screen design" (Ngo et al., 2000, p.98). Although some assert that interface aesthetic aspects may deteriorate usability, they are actually fundamental and more than relevant in a system's usability and acceptability. This thus highlights the need for creating an evaluation system of interface aesthetics, and not only for measuring usability. Previous researchers have developed metrics, such as the density and the balance for instance (Tullis, 1988; Streveler and Wasserman, 1984). Thus, Ngo et al. (2000) decided to lead a study using five different aesthetics measures (balance, equilibrium, symmetry, sequence, order and complexity), based on previous surveys. It is important to note that the order and complexity refer to the layout of elements of the screen. Therefore, the measure of order and complexity involves the four others, which gives us the following formula (given by Birkhoff (1933)):

$$M = \frac{O_{BM} + O_{EM} + O_{SM} + O_{SQM}}{C} (C > 0)$$

Ngo et al. (2000) applied this formula to a set of two-object layouts that are illustrated in the figure below.

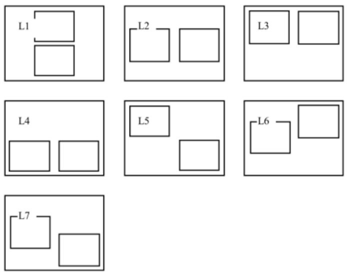


Figure 8 - Illustration of the seven layouts

Source: Ngo, D., Samsudin, A., Abdullah, R. (2000). Aesthetics measures for assessing graphic screens. *J. Inf. Sci. Eng 16*, p.108.

The computation gives rather satisfying results. We can clearly observe in figure 7 below, that the layouts 1, 2, and 5, were measured high, while the layouts, 3, 4, 6 and 7 were measured low. This coincides with what we observe in figure 6, in that the layouts 1, 2 and 5 are in balance and equilibrium, while the four others are only partially in balance and equilibrium. In sum, achieving balance, equilibrium, symmetry and sequence maximizes screen order, and fewer screen objects lowers the complexity.

Layout	O_{BM}	O_{EM}	O_{SM}	O_{SQM}	O
L1	0.5	0.5	1	0.5	2.5
L2	0.5	0.5	1	0.5	2.5
L3	0.25	0.25	0.5	0.5	1.5
L4	0.25	0.25	0.5	0	1
L5	0.5	0.5	0.5	0.5	2
L6	0.25	0.25	0	0.5	1
L7	0.25	0.25	0	0.5	1

Figure 9 - Results of the seven layouts

Source: Ngo, D., Samsudin, A., Abdullah, R. (2000). Aesthetics measures for assessing graphic screens. *J. Inf. Sci. Eng 16*, p.108.

To ensure the validity of these measures, Ngo et al (2000) replicated the study, but this time, using human beings to rate the layouts on a scale from 0 (worst) to 3 (best). They first confronted six GUI designers with the layouts, then a group of 22 undergraduates studying sciences. The scores given by both, the students and the designers were strongly similar to those computed with the formula. This thus translates a high correlation of judgments with measured aesthetics values, meaning that interfaces can be measured objectively.

This study highlights interesting aesthetic measures and results, which we will consider for our statistical analysis. Moreover, given this literature review, we now have a better knowledge and control of the concepts of our research topic. This will guide us in our statistical analysis and allow us to support our comments, conclusions and recommendations on the issue. Besides, it is interesting to bring together the outcomes of several studies in order to give more weight to our own results and above all, it helps in forming a valid model. It should be noted that many studies have been conducted but many of them only focused on the subjective part of our analysis. That is the reason why we want to go beyond and provide user interfaces' designers with fresher insights.

On the whole, this literature review allowed us to understand the general framework and fields of study related to the aesthetics and GUIs. However, we believed it was appropriate to summarize and illustrate in more detail the main references involved in our analysis. Therefore, we have created a diagram gathering the relevant features on the matter (see figure below).

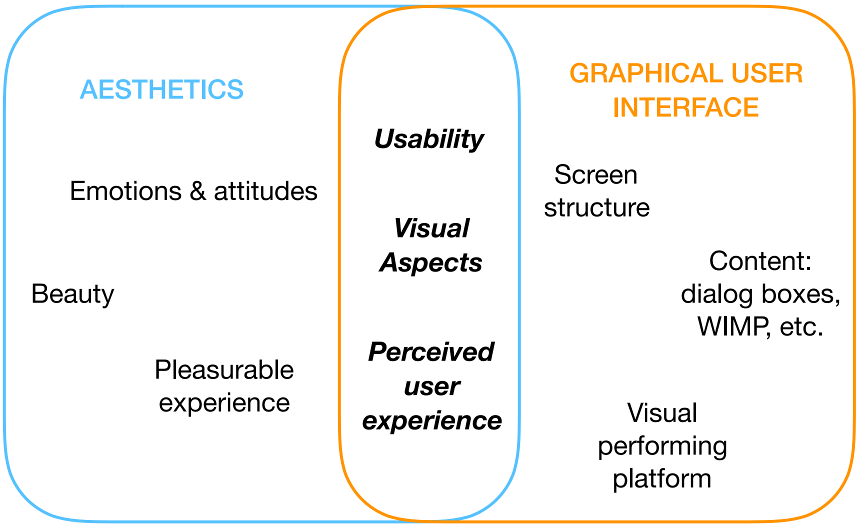


Figure 10 - Close up overview of the framework

This figure gives a closer overview and a refined framework that enables us to distinguish aesthetics feature from GUIs features, and the ones that apply to both fields. We can thus observe that the usability, the visual aspects and the perceived user experience hold for both domains. These three features thus stand in the hearth of our study and are the key factors. It should be noted that the best products often rely on all the disciplines all together (Saffer, 2010). This highlights the importance of mastering many disciplines along with their related features.

32.

3. Model and Hypothesis

Our analysis stems from a similar analysis previously conducted by Katharina Reinecke (Reinecke et al., 2013), a German professor of Computer Science and Engineering at the University of Washington. She designed two different studies. The first one aimed to report how people evaluate the visual complexity and colorfulness of websites; while the second aimed to gather aesthetic scores given by the participants to the same set of websites. We could find the explanations, outcomes and the surveys for both experiments on an online platform developed for this very specific purpose, namely Lab in the Wild (available on http://www.labinthewild.org).

The first study was implemented as follows. During 10-minute online tests, the participants were asked to rate a set of 30 screenshots, which were shown to them for only 500ms. Participants had to rate each website on a 9-point Likert scale, going from "not at all complex" to "very complex" and from "not at all colorful", to "very colorful". Then participants were asked to repeat the task a second time, after a short break, to validate the model. This time though the screenshots were ordered differently.

The second experiment followed the same pattern. As mentioned previously, it aimed to gather aesthetics scores. Therefore, participants were asked to rate a set of 30 websites on a 9-point Likert scale going from "not at all appealing" to "very appealing".

Thus, as mentioned earlier, the literature suggests that the complexity, the colorfulness and the aesthetics score of websites might be correlated. Indeed, in Reinecke's (Reinecke et al., 2013) experiments, we observe that people often gave low scores to some websites according to their colorfulness and complexity. Therefore, this allow us to assume that these variables must be correlated. Moreover, Reinecke is not the only study on which we based our assumptions. After going through Mathieu Zen's study, we could confirm the relevance of the assumptions. In its thesis, Zen (2017) compares a small part of Reinecke's results, which, as a reminder, are human reviews, with the results he collected with the dedicated software he designed, known as QUESTIM. This applicative software intents to measure the aesthetics of graphical user interfaces using a semi-automated process. QUESTIM computes various variables, which were considered as relevant and validated to measure the aesthetics of graphical user interface (Zen, 2017) Nonetheless, to validate the model properly, we believed it was important to broaden the analysis to verify our assumptions. We hereby propose an illustration of QUESTIM.

Figure 11 - Illustration of the applicative software QUESTIM

Therefore, to run a relevant study, we came up with some research questions and hypotheses, which guided our statistical analysis. We based our research questions and hypotheses on the two main concepts of our research question in management, namely: ***aesthetics*** and ***graphical interfaces***. Therefore, we came up with the following assumptions.

Research Question 1: Is the aesthetics judged the same way by a human being and by a semi-automated software (QUESTIM)?

- *Hypothesis 1: There is a link between the same type of variables computed in QUESTIM and those computed in Reinecke's study.*
- *Hypothesis 2: The variables computed in each study impact in the same way the visual appeal (aesthetics score).*

Research Question 2: Do both studies measure variables that enable aesthetics to be judged accurately?

- <u>*Hypothesis 3*</u>*: All the studied variables (as measured by the specific formulas) influence the visual appeal (aesthetics score).*
- <u>*Hypothesis 4*</u>*: The variables influencing the visual appeal (aesthetics score) can be ranked.*

Research Question 3: Would it be possible to anticipate the aesthetics success of a Graphical Interface?

- <u>*Hypothesis 5*</u>*: The ten templates show similar results as the 417 websites previously analyzed after being processed in QUESTIM.*

To better illustrates the framework of our research, we conceived a visual representation of the assumed relations of our model (see graph below). It demonstrates on which variables and previous studies we based our investigation. We intended to identify the independent as well as the dependent variables, and clarified which tests we conducted to get the best analysis.

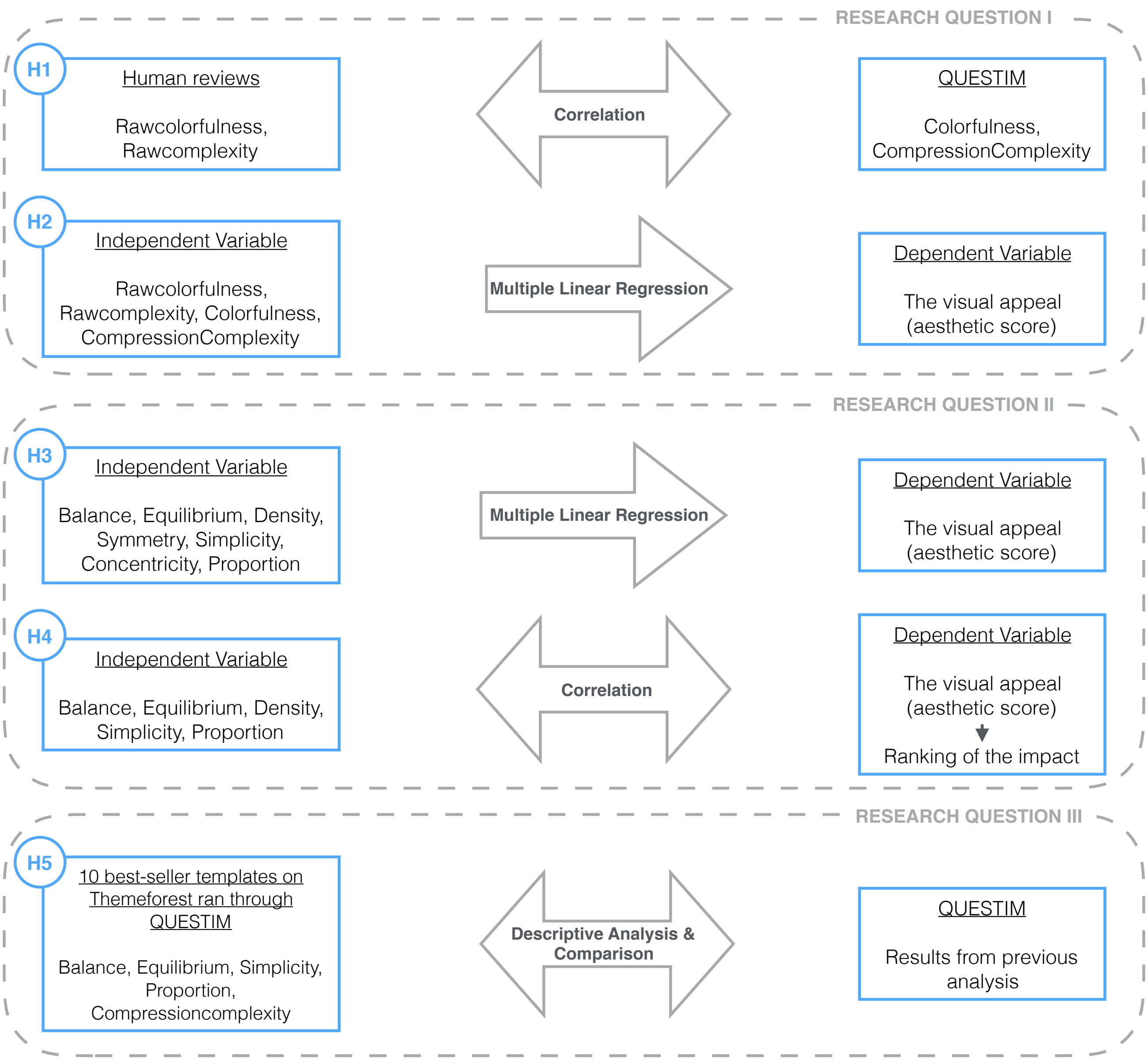


Figure 12 - Visual representation of the assumptions

As the main analysis of this thesis aims to find a way to accurately measure the aesthetics of graphical user interfaces, we believed it was appropriate to conduct an overall comparative study, based on the assumptions and research questions. This study involves different tests, such as correlation, multiple linear regression and a descriptive analysis combined with a specific comparison. As explained previously, it would allow us to come up with some conclusions as well as recommendations, but most importantly validate or refute our assumptions.

4. Methodology

Before going through any statistical analysis, it is fundamental to clarify and describe the methodology we followed to articulate this thesis. We will first explain how we defined the main topic and its goal. Then we will describe how data were collected and analyzed. It is important to mention that we based our methodology on the concept and theory explained by Creswell (Creswell, 2013).

4.1. Study Design

4.1.1. Research topic and its purpose

Defining a research topic requires following a few mandatory steps. First, the topic was picked as we have shown a strong interest in this domain. Then, we made sure that there was enough related literature available, and that our analysis could rely on previous experiments showing convincing results. It was also of prime importance to define a precise scope of study: not too wide but not too small. This was to ensure a leading structure and to avoid getting off track our main matter; or in other words, a framework that would not be too restrictive, and that would provide new and fresh outcomes.

This thesis has a main purpose of confirming and validating our assumptions, and at the same time defining precise aesthetics measurement methods. As a matter of fact, we want to provide our readers, and most importantly user interface designers with new insights. Therefore, we designed a comparative study to correlate the results of different experiments, using theory to support and enhance our reasoning. As mentioned previously, in the section above, we defined various hypotheses aiming to prove that there are some relations between the overall aesthetics score and the components of a website, such as the colorfulness, the complexity and the balance. By confronting the results of different studies, we would potentially be able to detect a general pattern, or at least point out some key features that should always be considered in an interface design.

After choosing a clear topic and designing the structure of our study, we focused on the problematic to develop our assumptions and research questions, which has already been described in the previous section.

4.1.2. Research method

According to Creswell (2013), as our analysis aims in a first step to test, and in a second step to verify various hypotheses, we can assert that we use a quantitative approach for our study. This type of approach examines a theory by defining a set of assumptions, and collecting a large amount of data so as to have an accurate representation of the studied population. The data are gathered and then analyzed employing hypothesis testing and statistical procedures. This approach enables attitudes, and opinions toward a specific topic, to be measured. This process also allows variables of interest to be identified and then relates them in questions or hypotheses.

4.1.3. Data collection, cleaning and preparation

To begin our comparative study, we first had to collect data by computing a total of 430 websites in the applicative software QUESTIM. It should be noted that we used the sample suggested in Reinecke's research involving a set of 430 different websites. Indeed, this sample was empirically validated and covers a representative population gathering diverse cultures. Therefore, we believed it would be suitable for our research. For each website, we collected 4 different files, namely, a .csv (comma-separated value), containing the experiment results, a .json (JavaScript Object Notation), gathering the coordinates of defined zones, and two .png (Portable Network Graphics), one original screenshot, and a second one with the defined zones that are required to run in QUESTIM (see appendices). After that, we had to gather all the measurement results in a single file. At the same time, as our analysis aims to compare the results obtained by Reinecke and those obtained using QUESTIM, we had to aggregate Reinecke's results in that same file. In figure 9 (below) we decided to present a non-exhaustive illustration of the zone definition on QUESTIM.

Figure 13 - Illustration of the zone definition in QUESTIM

We thought it would be appropriate to clean and prepare our database to exclude any absurd or missing data, and to prevent them from interfering in our analysis. Moreover, as our analysis is based on a comparison between two different studies, it was important to match the number of results. Therefore, after proceeding to this refining phase, our sample size amounts to 417 websites. Indeed, we excluded the following data: english_6, 31, 53, 84, 110, 126, 127, 227, 285, 296, 301 and foreign_45, 56 of our sample because there were some missing data such as some scores. We assumed that a loss of only 13 results would not affect our analysis considering that the rest of our sample remained sufficient.

Afterwards, we had to define the type of each variable we were confronted with, which is a crucial step as the hypothesis testing procedure depends on the type of variable we assign in our SPSS database. Indeed, depending on the type of variable, the adequate test will vary.

4.1.4. Data analysis and comprehension

Before starting any statistical analysis, it is fundamental to understand the role of each variable and the way in which they are measured and computed. Both studies provided a certain number of variables: 44 in Reinecke's experiment to be precise and 17 when using

QUESTIM. We focused our analysis on only three key variables computed by Reinecke (rawcomplexity, rawcolorfulness and score (visual appeal)), because these three variables were evaluated by human beings. Regarding the data provided by QUESTIM, we decided to focus on nine variables out of the 17 (balance, equilibrium, density, concentricity, simplicity, proportion, symmetry, colorfulness and compression complexity), as they seemed the most relevant to test our hypotheses. As explained previously, the data from Reinecke's experiment were collected among humans, while those from QUESTIM were computed using different algorithms. Regarding what they measure, most of them are explicit by their name, but to ensure a good comprehension of our study, we believed it was important to detail them and give a brief illustration of each in order to help visualize what they refer to.

- **Visual appeal**

The visual appeal refers to the overall aesthetics of an interface (rated on a 9-point Likert scale in Reinecke's experiment).

- **Balance**

According to Ngo (2001), the balance refers to the distribution of the optical weight in a picture. The balance can be understood as being the weight perception given to specific objects in a picture.

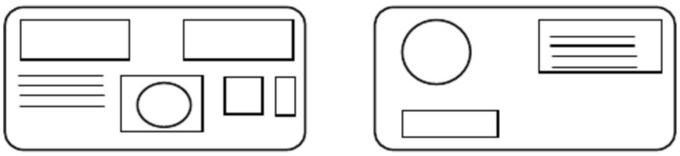


Figure 14 - Illustration of the balance

Source: Vanderdonckt, J., & Gillo, X. (1994, June). Visual techniques for traditional and multimedia layouts. In *Proceedings of the workshop on Advanced visual interfaces* (pp. 95-104). ACM, p.2.

Equilibrium

The equilibrium refers to the disposition of the contents of a page, which is considered as the layout. According to Ngo (2000), there is equilibrium when the layout is centered, and thus can be superposed with the frame center.

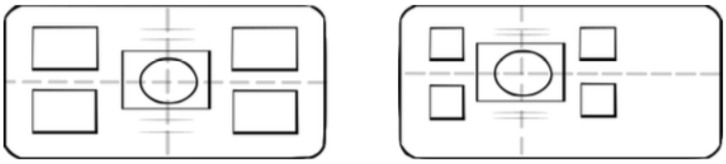


Figure 15 - Illustration of the equilibrium

Source: Zen, M. (2017). *A methodology for assessing aesthetics of a Graphical User Interface of an Information System by (Semi-) automatic Evaluation of Visual Metrics.* (PhD thesis). Université catholique de Louvain, Louvain-la-Neuve, p.135.

Density

The density of an interface refers to the number of objects covering the screen (Ngo, 2000).

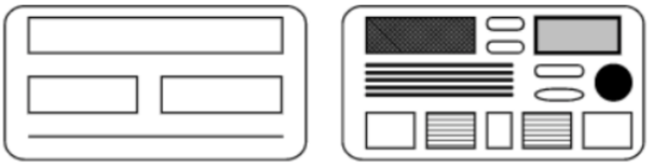


Figure 16 - Illustration of the density

Source: Zen, M. (2017). *A methodology for assessing aesthetics of a Graphical User Interface of an Information System by (Semi-) automatic Evaluation of Visual Metrics.* (PhD thesis). Université catholique de Louvain, Louvain-la-Neuve, p.132.

Concentricity

As the name of the variable indicates, the concentricity evaluates whether the objects on the screen are concentrated in the center of the page or not (Zen, 2016).

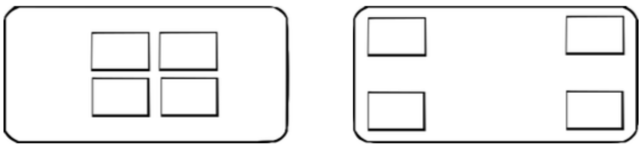


Figure 17 - Illustration of the concentricity

Source: Zen, M. (2017). *A methodology for assessing aesthetics of a Graphical User Interface of an Information System by (Semi-) automatic Evaluation of Visual Metrics.* (PhD thesis). Université catholique de Louvain, Louvain-la-Neuve, p.135.

Simplicity

According to Vanderdonckt (1994), the simplicity is the directness and the singleness of the layout. It must be free from any secondary complications or sophistications.

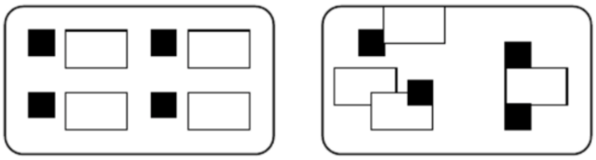


Figure 18 - Illustration of the simplicity

Source: Vanderdonckt, J., & Gillo, X. (1994, June). Visual techniques for traditional and multimedia layouts. In *Proceedings of the workshop on Advanced visual interfaces* (pp. 95-104). ACM, p.3.

Proportion

The proportion refers to the design of an interface. The objects of the screen must follow some ratio considered as aesthetically pleasant (Zen, 2016).

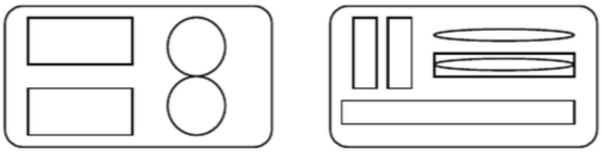


Figure 19 - Illustration of the proportion

Source: Vanderdonckt, J., & Gillo, X. (1994, June). Visual techniques for traditional and multimedia layouts. In *Proceedings of the workshop on Advanced visual interfaces* (pp. 95-104). ACM, p.3.

Symmetry

According to Arnheim (1954), the symmetry is the exact correspondence of form, or object on the opposite sides of a center or axis.

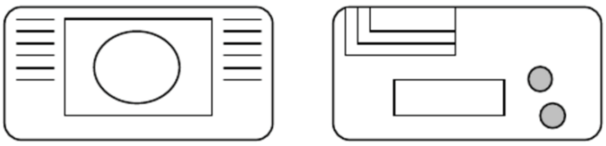


Figure 20 - Illustration of the symmetry

Source: Vanderdonckt, J., & Gillo, X. (1994, June). Visual techniques for traditional and multimedia layouts. In *Proceedings of the workshop on Advanced visual interfaces* (pp. 95-104). ACM, p.2.

Colorfulness

The colorfulness of an interface refers to its intensity of colors (Zen, 2016).

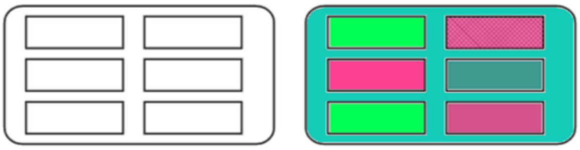


Figure 21 - Illustration of colorfulness

Source: Zen, M. (2017). *A methodology for assessing aesthetics of a Graphical User Interface of an Information System by (Semi-) automatic Evaluation of Visual Metrics.* (PhD thesis). Université catholique de Louvain, Louvain-la-Neuve, p.136.

Compression Complexity

The compression complexity is a variable that intents to compute the complexity, using the size of a compressed image, as this measure has been previously validated and suggested by Riglis (1998).

Now that we have a better insight of the variables used in our study, we can move on to the statistical part of our analysis, and proceed to test our different hypotheses. It is important to mention that these variables have been tested and validated to be used in previous studies (Ngo et al., 2000; Vanderdonckt and Gillo, 2013).

44.

5. Statistical Analysis

In this section, we will develop our statistical procedure and its results. As a reminder, we based our investigation on the different assumptions and research questions that we developed earlier. Thus, this section is articulated around one global question: *"Is it possible to measure the aesthetics of Graphical Interfaces*?", from which stem all our hypotheses and research questions. Conducting a very specific study on our database will enable us to validate or refute our assumptions, and therefore come up with some relevant conclusions.

5.1. Is it possible to measure the aesthetics of Graphical Interfaces?

In this phase of our analysis, we intend to answer the main question by testing our hypotheses and relating our variables of interest. To do so, we must choose the appropriate tests accordingly to the different type of variables we are confronted with. To examine best our database, we decided to use a predictive analytics software proposed by IBM, known as SPSS Statistics. Indeed, this program offers a large spectrum of tests, giving precise statistical analysis, reporting, predictive modeling, data mining, decision management/deployment and big data analytics (IBM, 2017). It should be noted that before running any test, we made sure that the application conditions were met accordingly to the nature of the test (see appendices). Moreover, it is also important to note that we decided to resort to the Pearson correlation instead of the Kendall correlation for different reasons. Firstly, according to Field (2009) the Pearson correlation is "an accurate measure of the linear relationship between two variables" when the data are considered as interval data, which is our case since our variables are computed on a scale going from 1 to 9 (Field, 2009, p. 177). Secondly, the Pearson correlation is the most widely used in comparison to the two others correlation methods (Statistics Solutions, 2017). Besides, the studied variables should be normally distributed, and therefore the related application conditions (aka assumptions) should be validated. This is the case in our study. Lastly, the Kendall correlation is a non-parametric correlation and should be used when examining a small data set, which is not our case. Indeed, we have a large data set.

Before going through the testing section, it is important to remind the nature of each variables to understand their role in the analysis. Therefore, we have gathered all the variables in a table in order to have a clear idea of their nature.

VARIABLES	NATURE
Rawcolorfulness	Independent
Colorfulness	Independent
Rawcomplexity	Independent
Compressioncomplexity	Independent
Equilibrium	Independent
Simplicity	Independent
Balance	Independent
Proportion	Independent
Density	Independent
Concentricity	Independent
Symmetry	Independent
Score	Dependent

Table 1 - Recapitulative table of the nature of each variable

5.1.1. *Research Question 1: Is the aesthetics judged the same way by a human being and by a dedicated software (QUESTIM)?*

For this very first research question, we came up with two different assumptions, both of which intend to compare and evaluate the relation between the variables computed by QUESTIM and those resulting from Reinecke's experiments, which are qualified as human reviews, implying that they are supposedly more subjective.

- *Hypothesis 1: There is a link between the same type of variables computed in QUESTIM and those computed in Reinecke's study.*

In this hypothesis, we decided to correlate the variables that potentially measured the same feature among each experiment. More precisely, we decide to analyze the correlation between the two variables of colorfulness, and between those of complexity. All four variables were defined as "scale" variables, which is the reason why we decided to use the Pearson correlation test. This analysis produced two correlation tables: the first set of outputs refers to the variables of colorfulness, while the second set of outputs refers to the variables of complexity.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
rawcolorfulness	417	1,00000000	9,00000000	4,77053647	1,51458070	,085	,120	-,540	,238
Colorfulness	417	,039177424	1,00000000	,522543034	,238808652	,255	,120	-,609	,238
Valid N (listwise)	417								

Table 2- Descriptive Statistics

Table 2 allows us to claim that the two variables have a positive skewness (0,085 and 0,255), which reflects a long distribution tail on the right. In addition, both variables have a negative kurtosis (-0,540 and -0,609), which reflects a platykurtic[2] distribution.

Correlations

		rawcolorfulness	Colorfulness
rawcolorfulness	Pearson Correlation	1	,344**
	Sig. (2-tailed)		,000
	N	417	417
Colorfulness	Pearson Correlation	,344**	1
	Sig. (2-tailed)	,000	
	N	417	417

**. Correlation is significant at the 0.01 level (2-tailed).

Table 3 - Correlations

Hypothesis test

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

Table 3 allows us to state that the two variables of colorfulness are correlated, as the level of significance is lower than 0,05 (0,000). Therefore, we can reject the null hypothesis (H_0). Moreover, we observe that the coefficient of correlation is 0,344, which indicates the direction or the degree of the relation between those two variables. In this case, the variables are positively correlated, which implies that if the colorfulness is rated as high by humans in Reinecke's study, it will be classified as very colorful by QUESTIM. Let us now have a look at the second set of outputs referring to the variables of complexity.

[2] A platykurtic distribution refers to a distribution less clustered around the mean.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
rawcomplexity	417	1,00000000	9,00000000	4,78149139	1,50782994	,077	,120	-,578	,238
compressionComplexity	417	,187950792	,802450951	,441883347	,110004428	,380	,120	,326	,238
Valid N (listwise)	417								

Table 4 - Descriptive Statistics

Table 4 allows us to claim that the two variables have a positive skewness (0,077 and 0,380), which reflects a long distribution tail on the right. In addition, rawcomplexity has a negative kurtosis (-,578), which reflects a platykurtic distribution, while the compressioncomplexity has a positive kurtosis (0,326), which reflects a leptokurtic[3] distribution.

Correlations

		rawcomplexity	compressionComplexity
rawcomplexity	Pearson Correlation	1	,757**
	Sig. (2-tailed)		,000
	N	417	417
compressionComplexity	Pearson Correlation	,757**	1
	Sig. (2-tailed)	,000	
	N	417	417

**. Correlation is significant at the 0.01 level (2-tailed).

Table 5 - Correlations

<u>Hypothesis test</u>

$$H_0: \rho = 0$$
$$H_1: \rho \neq 0$$

Table 5 enables us to claim that the two variables of complexity are correlated as well, as the level of significance is lower than 0,05 (0,000). Therefore, we can reject the null hypothesis (H_0). Moreover, we observe that the coefficient of correlation is 0,757, which implies a strong positive correlation between those variables. Again, it allows us to say that if the complexity is rated as high by humans in Reinecke's study, it will be defined as very complex by QUESTIM as well.

[3] A leptokurtic distribution refers to a distribution more clustered around the mean.

Therefore, these two first tests suggest that human beings and semi-automated applicative software such as QUESTIM potentially evaluate and judge the aesthetics in a similar manner, as the variable of interest appears to be positively correlated. Let us now focus on the second hypothesis, which is the following.

- *Hypothesis 2: The variables computed in each study impact in the same way the visual appeal (aesthetics score).*

In this hypothesis, we intend to evaluate the impact of the four same variables (rawcomplexity, compressioncomplexity, colorfulness, and rawcolorfulness) on the aesthetics score. Therefore, as we were confronted with "scale" variables, we assumed that the appropriated test would be a multiple linear regression; however, we decided to run the Pearson Correlation test first, to get the individual correlations. In this case, the aesthetics score is classified as the dependent variable, while the four others are considered as independent variables.

Correlations

		score	rawcomplexity	rawcolorfulness	compressionComplexity	Colorfulness
score	Pearson Correlation	1	-,363**	-,369**	-,415**	-,074
	Sig. (2-tailed)		,000	,000	,000	,129
	N	417	417	417	417	417
rawcomplexity	Pearson Correlation	-,363**	1	,944**	,757**	,339**
	Sig. (2-tailed)	,000		,000	,000	,000
	N	417	417	417	417	417
rawcolorfulness	Pearson Correlation	-,369**	,944**	1	,775**	,344**
	Sig. (2-tailed)	,000	,000		,000	,000
	N	417	417	417	417	417
compressionComplexity	Pearson Correlation	-,415**	,757**	,775**	1	,360**
	Sig. (2-tailed)	,000	,000	,000		,000
	N	417	417	417	417	417
Colorfulness	Pearson Correlation	-,074	,339**	,344**	,360**	1
	Sig. (2-tailed)	,129	,000	,000	,000	
	N	417	417	417	417	417

**. Correlation is significant at the 0.01 level (2-tailed).

Table 6 - Correlations

On Table 6 above though, we see that all the variables are correlated to the aesthetics score, except for the colorfulness. Regarding the coefficient of correlation, we observe that all the variables are negatively correlated. This thus implies that the higher the complexity and the colorfulness are rated, the lower the aesthetics score will be. Let us move on to the outputs of the multiple linear regression.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	,432[a]	,187	,179	,942219814	,187	23,621	4	412	,000

a. Predictors: (Constant), compressionComplexity, Colorfulness, rawcomplexity, rawcolorfulness

b. Dependent Variable: score

Table 7 - Model Summary

In Table 7, we observe that the adjusted R Square amounts to 0,179, which implies that 17,9% of the total variation of the score is explained by the variation of the four independent variables. This presents the importance of the association of these variables. The significance of the association is presented in the appendices (see appendix 3). Let us examine the test results.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	6,041	,195		30,964	,000
	rawcomplexity	-,047	,093	-,068	-,500	,618
	rawcolorfulness	-,050	,096	-,073	-,521	,603
	Colorfulness	,422	,209	,097	2,020	,044
	compressionComplexity	-3,240	,677	-,343	-4,784	,000

a. Dependent Variable: score

Table 8 - Coefficients

<u>Hypothesis test</u>

(1) $H_0: \beta_1 = 0$
$H_1: \beta_1 \neq 0$

(2) $H_0: \beta_2 = 0$
$H_1: \beta_2 \neq 0$

(3) $H_0: \beta_3 = 0$
$H_1: \beta_3 \neq 0$

(4) $H_0: \beta_4 = 0$
$H_1: \beta_4 \neq 0$

In this test, we intend to see if the four independent variables play a role in influencing the aesthetics score. On the coefficients table above, which results from the multiple linear regression, we observe that only the colorfulness and the compressioncomplexity are significant. We can thus reject the null hypothesis for the following tests: (3) and (4). Indeed, their level of significance is lower than 0,05. These results could be explained as followed: the four studied variables are supposed to predict the aesthetic score but, when computed together in a multiple linear regression, some appear to have a stronger correlation in predicting the aesthetic score. In our case, the rawcomplexity and rawcolorfulness appear not significant in contrast with the two others. This could imply that the variables computed by QUESTIM are better predictors of the aesthetics than human reviews.

The Pearson correlation table gives us the individual correlations, while the multiple linear regression model gives us the correlation where all the variables are combined with each other. It is important to mention that the multiple linear regression model is considered a more accurate prediction. Therefore, this table enables us to claim that the compressioncomplexity, the colorfulness both contribute to predicting the aesthetics score.

These first outcomes enable us to answer our first research question, and assert that the aesthetics of graphical interfaces is only partially judged similarly by a human being and by dedicated applicative software. Indeed, half the variables predict the aesthetics score. In the multiple linear regression, the rawcomplexity and the rawcolorfulness, resulting from Reinecke's study, are not significant, which implies that these variables should not be considered in the model. On top of that, when we run a multiple linear regression using the "stepwise method", which intends to define the regression model gathering the most significant variables only, we observe that only the compressioncomplexity is significant (see table below). The program progressively adds more variables in the model (one at the time), until all the significant predictors appear in the model.

Model Summary[b]

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	,415[a]	,173	,171	,946817430	,173	86,577	1	415	,000

a. Predictors: (Constant), compressionComplexity

b. Dependent Variable: score

Table 9 - Model Summary

In Table 9, we observe that the adjusted R Square amounts to 0,171, which implies that 17,1% of the total variation of the score is explained by the variation of the compressioncomplexity.

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	6,104	,192		31,768	,000
	compressionComplexity	-3,927	,422	-,415	-9,305	,000

a. Dependent Variable: score

Table 10 - Coefficients

It gives us the following equation:

$$\mathbf{Score} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1 compressioncomplexity}$$

So,

$$\mathbf{Score} = \mathbf{6,104} - \mathbf{3,927}\ (\boldsymbol{compressioncomplexity})$$

The stepwise method gives the strongest model, which implies that the compressioncomplexity is, statistically, the only strong predictor for the aesthetic score. However, the three other variables might not appear as significant, statistically, but they may have an impact anyway, although less obvious.

5.1.2. Research Question 2: Do both studies measure variables that enable aesthetics to be judged?

This second research question focuses on the variables that have not been considered yet. Therefore, we intend to see if all the other variables of interest computed by QUESTIM, play a role in predicting the aesthetics score. We also attempt to highlight which of them influence the score the most, and in which manner (positively, or negatively).

- *Hypothesis 3: All the studied variables (as measured by the specific formulas) influence the visual appeal (aesthetics score).*

As explained just above, for this third hypothesis, we focus on the rest of the independent variables, i.e. the simplicity, the balance, the equilibrium, the concentricity, the proportion, the symmetry, and the density, in order to evaluate their influence on the aesthetics score. Therefore, we decided to first study each of them separately and independently from the others, and analyze their correlation with the aesthetic score using the Pearson correlation. After that, we decided to run a multiple linear regression, to consider their impact when studied all together. This gives us the following tables.

Correlations

		score	Balance	Equilibrium	Density	Concentricity	Simplicity	Proportion	Symmetry
score	Pearson Correlation	1	,122*	-,001	,176**	-,095	-,235**	,209**	,097*
	Sig. (2-tailed)		,013	,988	,000	,052	,000	,000	,049
	N	417	417	417	417	417	417	417	417
Balance	Pearson Correlation	,122*	1	,462**	,618**	,119*	-,578**	,010	-,137**
	Sig. (2-tailed)	,013		,000	,000	,015	,000	,831	,005
	N	417	417	417	417	417	417	417	417
Equilibrium	Pearson Correlation	-,001	,462**	1	,319**	,399**	-,522**	-,066	-,157**
	Sig. (2-tailed)	,988	,000		,000	,000	,000	,178	,001
	N	417	417	417	417	417	417	417	417
Density	Pearson Correlation	,176**	,618**	,319**	1	-,015	-,896**	,232**	-,088
	Sig. (2-tailed)	,000	,000	,000		,761	,000	,000	,071
	N	417	417	417	417	417	417	417	417
Concentricity	Pearson Correlation	-,095	,119*	,399**	-,015	1	-,039	-,186**	-,045
	Sig. (2-tailed)	,052	,015	,000	,761		,425	,000	,355
	N	417	417	417	417	417	417	417	417
Simplicity	Pearson Correlation	-,235**	-,578**	-,522**	-,896**	-,039	1	-,209**	,088
	Sig. (2-tailed)	,000	,000	,000	,000	,425		,000	,074
	N	417	417	417	417	417	417	417	417
Proportion	Pearson Correlation	,209**	,010	-,066	,232**	-,186**	-,209**	1	,109*
	Sig. (2-tailed)	,000	,831	,178	,000	,000	,000		,027
	N	417	417	417	417	417	417	417	417
Symmetry	Pearson Correlation	,097*	-,137**	-,157**	-,088	-,045	,088	,109*	1
	Sig. (2-tailed)	,049	,005	,001	,071	,355	,074	,027	
	N	417	417	417	417	417	417	417	417

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Table 11 - Correlations

Hypothesis test

(1) $H_0: \rho_1 = 0$

$H_1: \rho_1 \neq 0$

(2) $H_0: \rho_2 = 0$

$H_1: \rho_2 \neq 0$

(3) $H_0: \rho_3 = 0$

$H_1: \rho_3 \neq 0$

(4) $H_0: \rho_4 = 0$

$H_1: \rho_4 \neq 0$

(5) $H_0: \rho_5 = 0$

$H_1: \rho_5 \neq 0$

(6) $H_0: \rho_6 = 0$

$H_1: \rho_6 \neq 0$

(7) $H_0: \rho_7 = 0$

$H_1: \rho_7 \neq 0$

On table 11, on the one hand, we observe that the balance, the density, the simplicity, the proportion and the symmetry are correlated to the aesthetics score. Indeed, their level of significance appears to be lower than 0,05. We can thus reject the null hypothesis for the following tests: (1), (3), (5), (6) and (7), and claim that these five variables play a role in predicting the aesthetics score independently from each other; while on the other hand, the equilibrium and the concentricity are not correlated, as their significance level is higher than 0,05, which implies that their impact on the aesthetics score is not statistically observed. Therefore, we cannot reject the null hypothesis for the tests (2) and (4).

This leads us to the analysis of the multiple linear regression table. It should be noted that again, we ran the test using the "stepwise method", which enables only the most significant variables to be gathered. The program thus ran the test several times to get the best model. Therefore, as we can observe on the output below, there remain only five relevant variables, namely, the simplicity, proportion, equilibrium, density and balance.

Model Summary[f]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,235[a]	,055	,053	1,01183196
2	,286[b]	,082	,077	,998678865
3	,307[c]	,094	,088	,992927698
4	,341[d]	,116	,108	,981966285
5	,356[e]	,126	,116	,977610444

a. Predictors: (Constant), Simplicity

b. Predictors: (Constant), Simplicity, Proportion

c. Predictors: (Constant), Simplicity, Proportion, Equilibrium

d. Predictors: (Constant), Simplicity, Proportion, Equilibrium, Density

e. Predictors: (Constant), Simplicity, Proportion, Equilibrium, Density, Balance

f. Dependent Variable: score

Table 12 - Model Summary

On table 12, we observe that adjusted R Square amounts to 0,116, which implies that 11,6% of the total variation of the score is explained by the variation of the five independent variables. The significance of the association is presented in the appendices (see appendix 4). Let us examine the coefficients of this multiple linear regression.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95,0% Confidence Interval for B Lower Bound	95,0% Confidence Interval for B Upper Bound	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	4,903	,119		41,074	,000	4,669	5,138		
	Simplicity	-1,752	,356	-,235	-4,918	,000	-2,452	-1,052	1,000	1,000
2	(Constant)	4,430	,180		24,546	,000	4,075	4,785		
	Simplicity	-1,492	,359	-,200	-4,151	,000	-2,199	-,786	,956	1,045
	Proportion	,489	,141	,167	3,465	,001	,211	,766	,956	1,045
3	(Constant)	4,736	,220		21,549	,000	4,304	5,168		
	Simplicity	-2,059	,428	-,276	-4,813	,000	-2,899	-1,218	,668	1,497
	Proportion	,416	,143	,142	2,901	,004	,134	,698	,914	1,094
	Equilibrium	-,094	,039	-,135	-2,410	,016	-,171	-,017	,695	1,438
4	(Constant)	6,623	,628		10,549	,000	5,389	7,858		
	Simplicity	-4,764	,944	-,638	-5,046	,000	-6,620	-2,908	,134	7,459
	Proportion	,427	,142	,146	3,008	,003	,148	,706	,914	1,094
	Equilibrium	-,146	,042	-,209	-3,475	,001	-,228	-,063	,594	1,683
	Density	-1,595	,498	-,363	-3,205	,001	-2,573	-,617	,167	5,973
5	(Constant)	6,464	,629		10,270	,000	5,227	7,701		
	Simplicity	-5,082	,951	-,681	-5,341	,000	-6,952	-3,211	,131	7,641
	Proportion	,462	,142	,158	3,249	,001	,182	,742	,902	1,109
	Equilibrium	-,181	,045	-,259	-4,036	,000	-,269	-,093	,517	1,936
	Density	-2,081	,544	-,473	-3,825	,000	-3,151	-1,012	,139	7,201
	Balance	,751	,347	,139	2,163	,031	,069	1,434	,518	1,930

a. Dependent Variable: score

Table 13 - Coefficients

<u>Hypothesis test</u>

(1) $H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

(2) $H_0: \beta_2 = 0$

$H_1: \beta_2 \neq 0$

(3) $H_0: \beta_3 = 0$

$H_1: \beta_3 \neq 0$

(4) $H_0: \beta_4 = 0$

$H_1: \beta_4 \neq 0$

(5) $H_0: \beta_5 = 0$

$H_1: \beta_5 \neq 0$

It gives us the following equation:

$$\mathbf{Score} = \boldsymbol{\beta_0 + \beta_1 simplicity + \beta_2 proportion + \beta_3 equilibrium + \beta_4 density + \beta_5 balance}$$

So,

$$\mathbf{Score} = \mathbf{6,464 - 5,082}\ (\boldsymbol{simplicity}) + \mathbf{0,462}\ (\boldsymbol{proportion}) - \mathbf{0,181}\ (\boldsymbol{equilibrium}) - \mathbf{2,081}\ (\boldsymbol{density}) + \mathbf{0,751}\ (\boldsymbol{balance})$$

In this test, we intend to see which independent variables influence the aesthetics score. In the coefficient table above, we observe that the fifth model is the version that gathers the most significant variables, as their level of significance is lower than 0,05. The fact that they are significant indicates that they influence the aesthetics score and are considered as predictors. We can thus reject the null hypothesis for the following tests: (1), (2), (3), (4) and (5). These results could be explained as followed: since the concentricity and the symmetry have been removed from the model, it implies that they do not predict the aesthetic score. Truth is that a study conducted by Swaddle and Cuthill (1995) have revealed that the face's symmetry did not predict its attractiveness. Indeed, more symmetric faces were rated as less attractive. This, thus confirms our results and the removal of the symmetry variable. Regarding the concentricity, as it refers to the concentration of the objects of the screen, we believe that this variable is less relevant in assessing the aesthetics, because when modeling an interface, designers usually tend to use the whole screen space instead of concentrating everything in one area. This may be the reason why the concentricity was removed from the model as it did not predict the aesthetics score as much as the other variables.

Correlations

		score	Density	Simplicity	Proportion	Balance	Equilibrium
score	Pearson Correlation	1	,176**	-,235**	,209**	,122*	-,001
	Sig. (2-tailed)		,000	,000	,000	,013	,988
	N	417	417	417	417	417	417
Density	Pearson Correlation	,176**	1	-,896**	,232**	,618**	,319**
	Sig. (2-tailed)	,000		,000	,000	,000	,000
	N	417	417	417	417	417	417
Simplicity	Pearson Correlation	-,235**	-,896**	1	-,209**	-,578**	-,522**
	Sig. (2-tailed)	,000	,000		,000	,000	,000
	N	417	417	417	417	417	417
Proportion	Pearson Correlation	,209**	,232**	-,209**	1	,010	-,066
	Sig. (2-tailed)	,000	,000	,000		,831	,178
	N	417	417	417	417	417	417
Balance	Pearson Correlation	,122*	,618**	-,578**	,010	1	,462**
	Sig. (2-tailed)	,013	,000	,000	,831		,000
	N	417	417	417	417	417	417
Equilibrium	Pearson Correlation	-,001	,319**	-,522**	-,066	,462**	1
	Sig. (2-tailed)	,988	,000	,000	,178	,000	
	N	417	417	417	417	417	417

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 14 - Correlations

On table 14 above (gathering the five significant variables only), we observe similar levels of significance, except for the equilibrium, which is not significant anymore. As mentioned earlier, the variables that appear significant play a role in predicting the aesthetics score. However, it should be noted that the Pearson correlation table give us the individual correlations, while the multiple linear regression model gives us the correlation where all the variables are combined with each other. It is important to mention that the multiple linear regression model is considered as a better prediction. Therefore, this output enables us to claim that the density, the simplicity, the proportion, the balance, and the equilibrium all contribute together in predicting the aesthetics score. We can thus support that fact that some of the variables of interest influence the visual appeal. However, the concentricity and the symmetry do not influence the aesthetic score as they were excluded from the model (see appendix 4 "excluded variables"). As explained earlier, the "stepwise method" sorts the variables according to their significance and provides a model that gathers the most significant variables. Thereby, these two-last variables were not considered as they appeared not to be significant.

- *Hypothesis 4: The studied variables influencing the visual appeal (aesthetics score) can be ranked.*

In this hypothesis, we intent to determine if the five significant variables can be ranked, depending on their level of impact on the aesthetics score. Therefore, we analyzed the correlation table that results from a multiple linear regression involving the five relevant variables only (defined above). As explained earlier, the multiple linear regression is a better predictor than the Pearson correlation because it provides the correlations of the variables of interest in association; hence the use of this correlation output.

Correlations

		score	Density	Simplicity	Proportion	Balance	Equilibrium
Pearson Correlation	score	1,000	,176	-,235	,209	,122	-,001
	Density	,176	1,000	-,896	,232	,618	,319
	Simplicity	-,235	-,896	1,000	-,209	-,578	-,522
	Proportion	,209	,232	-,209	1,000	,010	-,066
	Balance	,122	,618	-,578	,010	1,000	,462
	Equilibrium	-,001	,319	-,522	-,066	,462	1,000
Sig. (1-tailed)	score	.	,000	,000	,000	,006	,494
	Density	,000	.	,000	,000	,000	,000
	Simplicity	,000	,000	.	,000	,000	,000
	Proportion	,000	,000	,000	.	,415	,089
	Balance	,006	,000	,000	,415	.	,000
	Equilibrium	,494	,000	,000	,089	,000	.
N	score	417	417	417	417	417	417
	Density	417	417	417	417	417	417
	Simplicity	417	417	417	417	417	417
	Proportion	417	417	417	417	417	417
	Balance	417	417	417	417	417	417
	Equilibrium	417	417	417	417	417	417

Table 15 - Correlations

On table 15, we observe that the density, the proportion, and the balance positively influence the aesthetics, while the simplicity negatively influences the aesthetics. It should be noted that the equilibrium is not considered on its own, as this variable does not appear to be correlated to the aesthetics score. Consequently, on the one hand, the higher the density, the proportion and the balance are, the higher the aesthetics score will be; while on the other hand, the higher the simplicity is, the lower the aesthetics score will be. In order to rank these variables, we looked at their correlation. Regarding the positively correlated variables, first comes the proportion (influencing the score by 20,9%), then the density (influencing the score by 17,6%), and lastly the balance (influencing the score by 12,2%). Regarding the negatively correlated variable, the simplicity is the only variable remaining, influencing the score by -23,5%. However, we observe that for all the variable examined in this test, the correlation is lower than the correlation of the variables studied earlier (compressioncomplexity and colorfulness). We can thus argue that the different variable of interest can be ranked accordingly to their correlation.

These tests enable us to answer our second research question. Reinecke’s study clearly aims to measure the aesthetics of graphical user interface; however, it is human reviews-based. And as explained in the literature review, the real struggle is to gather accurate statistical results; hence, the importance of the results given through QUESTIM. Indeed, it provides interesting measures that could be used in interface design. Moreover, thanks to these tests, we can determine the importance of each variable of the visual appeal of graphical interfaces.

To get a better overview of the variables influencing the aesthetics score, we decided to run another stepwise multiple linear regression gathering all the significant variables together (i.e. the equilibrium, the simplicity, the balance, the proportion, the density, the compressioncomplexity and the colorfulness), in order to observe their impact on the score, when in combination. This gave us the following output.

Model Summary[h]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	,415[a]	,173	,171	,946817430	,173	86,577	1	415	,000
2	,547[b]	,299	,296	,872539049	,126	74,665	1	414	,000
3	,563[c]	,317	,313	,862021483	,018	11,164	1	413	,001
4	,574[d]	,329	,323	,855493337	,012	7,327	1	412	,007
5	,592[e]	,351	,343	,842895160	,021	13,408	1	411	,000
6	,604[f]	,365	,356	,834433272	,015	9,378	1	410	,002
7	,604[g]	,365	,357	,833420478	,000	,003	1	410	,957

a. Predictors: (Constant), compressionComplexity
b. Predictors: (Constant), compressionComplexity, Density
c. Predictors: (Constant), compressionComplexity, Density, Proportion
d. Predictors: (Constant), compressionComplexity, Density, Proportion, Equilibrium
e. Predictors: (Constant), compressionComplexity, Density, Proportion, Equilibrium, Balance
f. Predictors: (Constant), compressionComplexity, Density, Proportion, Equilibrium, Balance, Simplicity
g. Predictors: (Constant), compressionComplexity, Proportion, Equilibrium, Balance, Simplicity
h. Dependent Variable: score

Table 16 - Model Summary

On table 16, we observe that the adjusted R Square amounts to 0,357, which implies that 35,7% of the total variation of the score is explained by the variation of the five independent variables.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	6,104	,192		31,768	,000	5,727	6,482		
	compressionComplexity	-3,927	,422	-,415	-9,305	,000	-4,756	-3,097	1,000	1,000
2	(Constant)	5,609	,186		30,138	,000	5,243	5,975		
	compressionComplexity	-5,269	,419	-,558	-12,582	,000	-6,093	-4,446	,862	1,160
	Density	1,684	,195	,383	8,641	,000	1,301	2,067	,862	1,160
3	(Constant)	5,338	,201		26,555	,000	4,943	5,733		
	compressionComplexity	-5,171	,415	-,547	-12,465	,000	-5,986	-4,355	,858	1,166
	Density	1,524	,198	,347	7,683	,000	1,134	1,914	,812	1,231
	Proportion	,410	,123	,140	3,341	,001	,169	,651	,941	1,062
4	(Constant)	5,382	,200		26,889	,000	4,988	5,775		
	compressionComplexity	-5,335	,416	-,564	-12,820	,000	-6,153	-4,517	,840	1,191
	Density	1,737	,212	,395	8,193	,000	1,320	2,153	,701	1,427
	Proportion	,356	,123	,121	2,880	,004	,113	,598	,916	1,091
	Equilibrium	-,082	,030	-,118	-2,707	,007	-,142	-,022	,859	1,164
5	(Constant)	4,932	,232		21,231	,000	4,476	5,389		
	compressionComplexity	-5,573	,415	-,590	-13,426	,000	-6,390	-4,757	,819	1,221
	Density	1,294	,241	,294	5,364	,000	,820	1,768	,525	1,905
	Proportion	,408	,122	,139	3,328	,001	,167	,648	,904	1,106
	Equilibrium	-,124	,032	-,178	-3,875	,000	-,187	-,061	,749	1,335
	Balance	1,097	,300	,202	3,662	,000	,508	1,686	,518	1,932
6	(Constant)	6,419	,537		11,948	,000	5,363	7,475		
	compressionComplexity	-5,259	,424	-,556	-12,415	,000	-6,092	-4,426	,771	1,297
	Density	-,027	,493	-,006	-,054	,957	-,996	,943	,123	8,115
	Proportion	,381	,122	,130	3,134	,002	,142	,620	,899	1,112
	Equilibrium	-,189	,038	-,271	-4,957	,000	-,264	-,114	,516	1,937
	Balance	1,206	,299	,222	4,037	,000	,619	1,793	,510	1,960
	Simplicity	2,563	,837	,343	3,062	,002	4,209	,918	,123	8,118
7	(Constant)	6,397	,351		18,210	,000	5,707	7,088		
	compressionComplexity	-5,267	,399	-,557	-13,216	,000	-6,050	-4,483	,869	1,151
	Proportion	,380	,121	,130	3,144	,002	,143	,618	,906	1,104
	Equilibrium	-,188	,034	-,270	-5,498	,000	-,256	-,121	,640	1,563
	Balance	1,200	,280	,221	4,286	,000	,650	1,751	,579	1,726
	Simplicity	-2,524	,405	-,338	-6,231	,000	-3,320	-1,728	,525	1,906

a. Dependent Variable: score

Table 17 - Coefficients

In this output (table 17), we notice that the seventh model involves only five significant variables out of the seven, namely, the compressioncomplexity, the proportion, the equilibrium, the balance and the simplicity. Indeed, the significance level of these five variables is lower than 0,05, implying the existence of an impact on the score when in combination. In order to rank them just like in the previous analysis, we ran a Pearson correlation (see table below).

Correlations

		score	compression Complexity	Simplicity	Balance	Proportion	Equilibrium
score	Pearson Correlation	1	-,415**	-,235**	,122*	,209**	-,001
	Sig. (2-tailed)		,000	,000	,013	,000	,988
	N	417	417	417	417	417	417
compressionComplexity	Pearson Correlation	-,415**	1	-,211**	,308**	,022	,001
	Sig. (2-tailed)	,000		,000	,000	,661	,977
	N	417	417	417	417	417	417
Simplicity	Pearson Correlation	-,235**	-,211**	1	-,578**	-,209**	-,522**
	Sig. (2-tailed)	,000	,000		,000	,000	,000
	N	417	417	417	417	417	417
Balance	Pearson Correlation	,122*	,308**	-,578**	1	,010	,462**
	Sig. (2-tailed)	,013	,000	,000		,831	,000
	N	417	417	417	417	417	417
Proportion	Pearson Correlation	,209**	,022	-,209**	,010	1	-,066
	Sig. (2-tailed)	,000	,661	,000	,831		,178
	N	417	417	417	417	417	417
Equilibrium	Pearson Correlation	-,001	,001	-,522**	,462**	-,066	1
	Sig. (2-tailed)	,988	,977	,000	,000	,178	
	N	417	417	417	417	417	417

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 18 - Correlations

On table 18, we observe that the proportion and the balance both positively influence the aesthetics score, by 20,9% and by 12,2% respectively; while the compressioncomplexity and the simplicity negatively influence the score, by -41,5%, and by -23,5%, respectively. The equilibrium appears as not significant on its own, and when in combination with the others, it negatively influences the score (see coefficient table above). These results thus allow us to rank the variables according to their level of impact on the aesthetics score.

5.1.3. Research Question 3: Would it be possible to anticipate the aesthetics success of a Graphical Interface?

For this last research question, we want to see if it is possible to anticipate the aesthetics success of graphical interfaces. Therefore, we pushed our analysis a little further. Indeed, we decided to analyze the ten best-seller website templates from Themeforest (available on https://themeforest.net), which is a famous website offering various types of website templates. We thought it would be interesting to run these ten websites through QUESTIM, and to then examine their results.

- *<u>Hypothesis 5</u>: The ten templates show similar results as the 417 websites previously analyzed after being computed in QUESTIM.*

To orient our analysis, we formulated this last hypothesis. To determine if some similarities do exist, we believed it was appropriate to examine the descriptive statistics, and then to compare them to our previous results.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Balance	10	,712409049	,987018754	,877176733	,095720663
Equilibrium	10	,955120070	,997027797	,982614239	,013563113
Density	10	,742053986	1,28347647	1,02889036	,177080795
Concentricity	10	,415350892	,603222640	,516751538	,062368346
Simplicity	10	-,07923824	,298061889	,107221487	,101882998
Proportion	10	1	1	1,00	,000
Symmetry	10	,418864232	,501970516	,458404619	,025757033
Colorfulness	10	,049992043	,650966749	,362735052	,200783175
compressionComplexity	10	,212634783	,314269893	,264440317	,042000631
Valid N (listwise)	10				

Table 19 - Descriptive Statistics

In this table, we observe that the mean for the balance, the equilibrium and the proportion are quite high; while the mean for the simplicity and the compressioncomplexity appear to be low. In the results obtained previously, we noted that the equilibrium, the simplicity as well as the compressioncomplexity were negatively correlated to the aesthetics score which means that the lower they are, the higher the aesthetics score will be. Besides, we observed the exact opposite regarding the balance and proportion. Indeed, these variables, on the contrary, are positively correlated with the aesthetics score. This thus allow us to state that the balance and proportion should be the highest possible to get a high aesthetics score, while the equilibrium, the compressioncomplexity and the simplicity should be the lowest possible, which is the case in our analysis of the ten templates, except for the equilibrium which is surprisingly highly ranked. That being said, in the previous analysis, the equilibrium has always shown no impact when individually correlated to the score. It should also be noted that we do not consider the concentricity, the symmetry, the colorfulness, or the density, as they did not predict the aesthetics score in the previous analysis.

According to these results, we believe it would possible to anticipate the aesthetics success of a graphical interface. Therefore, in the design process of a graphical user interface (website), we are convinced that if we respect the “guidelines”, and intend to provide a website with high balance and proportion, but at the same time with low compressioncomplexity, colorfulness and simplicity, there would be a high probability that the website will be considered as aesthetically pleasing, as we would have followed the same designing pattern as the then best-seller websites templates.

It should be noted that this last research question represents a major challenge as we wonder if it would possible to predict the aesthetic perception of a screen, before even testing it with participants. Uribe et al. (2017), suggest an evaluation method based on low-level parameters. Indeed, computing a set of twenty, or more, formulas to predict the aesthetics seems a lot, and thus complex and time-taking. The approach and results presented in Uribe et al. (2017) analysis seem promising and maybe more appropriate for such a pre-evaluation phase. Therefore, we believe this method should also be considered for anticipating the aesthetic success of interfaces, using a simple and fast approach.

6. Results

In this section, we will present the results obtained through our statistical analysis. These outcomes will enable us to validate or refute our different assumptions for the entire sample. It will also allow us to test the relevance of our conceptual framework. It should be noted that our sample was empirically validated in Reinecke's study. Indeed, it gathers a total of 430 websites from all types, which was identified as a representative sample of the population. However, we must keep in mind that our study was conducted at a given time and that our society evolves constantly, therefore, our results and findings may not be valid at some point in the future.

When we tested the two first hypotheses, which attempt to establish a link between Reinecke's study and QUESTIM, we first observed that the similar variables (rawcomplexity, rawcolorfulness, compressioncomplexity, and colorfulness) given in both studies, were correlated two by two. In addition to that, when running the multiple linear regression, we observed that only two variables (the compressioncomplexity and the colorfulness) had an impact on predicting the aesthetics score. As a matter of fact, when computing them all together, the rawcomplexity and the rawcolorfulness appeared not to be significant anymore, which implies that their impact on the aesthetic score is not statistically observed when in combination with the other variables. Another interesting fact, is that when looking at the individual correlation (Pearson correlation) of these four variables, this time the colorfulness was not significant. Then we ran a second multiple linear regression, using the stepwise method, which highlighted that only the compressioncomplexity was significant and probably is the strongest predictor. However, we must keep in mind that the other variables may have an impact as well, although their significance is less obvious. So, on the whole, two variables influence the aesthetics score when in combination: the compressioncomplexity and the colorfulness, while the rawcomplexity and rawcolorfulness are left out due to their statistical irrelevance. This thus allowed us to state that human reviews and computed measures somewhat judge the aesthetics in a similar manner due to their correlation. However, the variables issued from QUESTIM were stronger predictors for the aesthetic score, statistically speaking. Therefore, we can validate the two first hypotheses.

Regarding the second set of hypotheses, which aim to determine if both studies fundamentally judge the aesthetics of graphical user interfaces, we focused on testing the variables that had not been considered yet. This thus involved variables given by QUESTIM only, as those resulting from Reinecke's study had all already been examined earlier. Thanks to a multiple linear regression, using the stepwise method, we found out that only five variables (density, simplicity, proportion, balance and equilibrium) out of the seven remaining had an impact on predicting the aesthetics score, as their significance level were lower than 0,05. After defining which variables were significant, we intended to determine if they could be ranked in order to appraise their importance individually. Thus, the correlation table (Pearson correlation) allowed us to claim that the proportion positively influenced the score by 20,9%, closely followed by the density and the balance, which positively influenced the scored by 17,6% and by 12,2% respectively. The simplicity though negatively influenced the score by 23,5%. It should be noted that the equilibrium is not presented because its significance level is higher than 0,05, and thus does not influence the score on its own. Thanks to these results, we could argue that both studies did indeed computed measures judging the aesthetics. We can thus validate the second set of hypotheses.

In order to have a general overview of all the significant variables, we decided to analyze the regression model gathering all the variables that had been previously identified as significant in predicting the score (i.e. the equilibrium, the simplicity, the balance, the proportion, the density, the compressioncomplexity and the colorfulness), using the stepwise method. It followed that the colorfulness and the density were not significant anymore, and therefore were removed from the model.

In the last research question, we wanted to go beyond the comparison of Reinecke's and QUESTIM results. Therefore, we intended to determine to what extent it was possible to anticipate the aesthetics success of graphical user interfaces. To do so, we conducted an analysis of the scores given by QUESTIM to the ten best-seller templates from Themeforest, using the descriptive tables. This allowed us to uncover that the mean of the five relevant variables matched the previous diagnoses. In fact, as classified as best-seller templates, we expected that their aesthetics would be scored according to our previous findings through QUESTIM. Therefore, we can validate this last hypothesis as its outcomes appear convincing. However, as explained earlier, computing a large amount of formulas is quite inconvenient. Therefore, we believe that resorting to another method such as the one proposed by Uribe et al. (2017), using few low-level parameters for evaluation the aesthetics in a predicting phase, would be interesting and maybe easier.

To visualize better these results, we decided to design the recapitulative table below.

VARIABLES	CORRELATIONS
Rawcolorfulness	**Not significant**
Colorfulness	Negative correlation with the score
Rawcomplexity	**Not significant**
Compressioncomplexity	Negative correlation with the score
Equilibrium	Negative correlation with the score
Simplicity	Negative correlation with the score
Balance	Positive correlation with the score
Proportion	Positive correlation with the score
Density	Positive correlation with the score
Concentricity	**Not significant**
Symmetry	**Not significant**

Table 20 - Recapitulative table of the results

Table legend:

- *Significant variables (green)*
- *Not significant variables (grey)*
- *Significant variables at first, then turned out to be not significant when computed all together (orange)*

7. Discussions on the outcomes and suggestions for future perspectives

In this chapter, we will first discuss and interpret the results obtained; and in a second step, we will suggest and develop new perspectives for future analysis.

7.1. Discussions and interpretations of the outcomes

As a quick reminder, our main goal in this analysis was to compare the outcomes of Reinecke's study with the results given through QUESTIM, in terms of aesthetics measurement. Indeed, in the literature review, we uncovered that assessing the aesthetics remained a complex concept as the clear majority of aesthetics tests previously conducted resulted in highly subjective outcomes. This was largely due to a lack of focus on human reviews, but also because of the lack of methods available to numerically measure the aesthetics. It clearly illustrated the need for new in depth research on the topic. Comparing such studies helped us to determine whether our applicative software (QUESTIM) computed relevant aesthetics measures or not; and if so, to demonstrate which of the variables fundamentally predicted the aesthetics score. Let us thus focus on the interpretations of the outcomes.

As we can observe in the previous section, we could basically validate all our assumptions as well as all our research questions for the studied sample given the results obtained. It is important to keep in mind though, that the results are balanced and sometimes only partially validate the hypotheses, as demonstrated in the first set of hypotheses. The comparison between Reinecke's study and QUESTIM led to the following conclusion: human reviews and the applicative software judge the aesthetics in a partially similar manner. Indeed, the four studied variables have first shown positive correlations two by two, and then the results have shown that two (the colorfulness and the compressioncomplexity) out of the four variables of interest, had an impact on predicting the aesthetics score when in combination. Therefore, we can assert that the measures computed by the applicative software QUESTIM are relevant in assessing the aesthetics as they have demonstrated strong correlation with human reviews. Although this statement only accounts for four variables, it illustrates the effectiveness of the software. We thus expect other variables to play a role in predicting the aesthetics. This leads us to the next analysis, which consists of determining which variables given through QUESTIM fundamentally impact the aesthetics and how.

So, when deepening our analysis, and examining the further outcomes closely, we could assert that the applicative software QUESTIM computed few relevant aesthetics measures. Indeed, it should be noted that not all the variables are significant. Out of the nine variables given through QUESTIM, only five of them (the equilibrium, the proportion, the balance, the compressioncomplexity and the simplicity) were observed as significant, those of which therefore predicted the aesthetics score. In addition, it is important to mention that their significance level varied depending on their associations. We observed that some variables played a role in predicting the aesthetics score on their own, while when in combination with others, they did not anymore. That is the reason why we resorted to the stepwise method, which enabled us to gather significant variables only. Therefore, we can claim that out of the nine studied variables given through QUESTIM, only five of them predict the aesthetics score. This suggests that these significant variables should be closely looked at in terms of interface design and at a primary stage. Nevertheless, the rest of the variables should not be ignored. As some appeared as significant when correlated on their own, we believe they accounts as much as the five others, but maybe should be considered in a second or later stage of the designing process. Moreover, the last hypothesis, referring to the anticipation of the aesthetics success of an interface, allowed us to reinforce this statement, as its results matched the previous findings.

On the whole, all these tests have demonstrated that QUESTIM provides some interesting insights and could definitely be used in interface design. We believe that referring to these measures and variables would decidedly improve the design process, and benefit human computer interaction designers. On top of that, improving the GUI conception process would lead to better results, which would then positively impact companies in many ways. Indeed, as explained previously, an online presence is increasingly important and therefore, we are convinced that great interface designs play a role in marketing as well as on users' satisfaction. As a matter of fact, we have learned in the literature that interface design was a slow and thorough process, and hard to master, which impedes fast and effective user interface conception (Baumer et al., 1996). This thus clearly highlights the need for innovation and advances. Using the applicative software QUESTIM could represent just part of the process of course, such as a preliminary phase for instance, but we are convinced that it would speed up the decision-making process in the very early steps of the design process; therefore, it would lead towards a faster and more powerful approach, thus proving fruitful outcomes.

7.2. Suggestions for future perspectives

As clarified in the introduction of this chapter, we will now suggest and develop new perspectives that we believed that could be relevant for future analysis.

In this section, we will thus focus on other measurement tools and methods that, although they have not been addressed yet, represent a great opportunity to investigate deeper our main issue: UI aesthetics and interface design. Thereby, we will suggest three other measurements, and optimization methods on aesthetics and performance, namely, A/B testing, multivariate testing, and Bayesian test. We will also present their utility and functioning.

7.2.1. A/B testing

Let us now focus on the first measurement method, which is called A/B testing, but is also known as split testing or bucket testing, and how it functions. Optimizely, a web experimentation platform dedicated to testing across channel, devices, and touchpoints defines this method as followed: "A/B testing is a method of comparing two versions of a webpage or app against each other to determine which one performs better. A/B testing is essentially an experiment where two or more variants of a page are shown to users at random, and statistical analysis is used to determine which variation performs better for a given conversion goal" (Optimizely, 2017, para. 1). This technique is also known as the "frequentist method", which is based on the law of observations. It is therefore characterized as experimental or inductive. This approach enables organizations to directly compare two distinct situations, in terms of interface design, but also to gather results and data on both situations. Therefore, this process allows the leverage of a change from one situation to another to be examined and evaluated. The testing method provides organizations decision-makers with web optimization advice, that are data-driven and thereby guarantee positive outcomes if applied correctly.

The functioning of this approach is rather easy. As mentioned earlier, it consists of testing two situations (webpage or app screen) against each other. To do so, we select one page, then reproduce a second version of that same page, which would have undergone modifications. The shifts can vary in their nature, from very simple to very complex, depending on the needs of the organization. Afterwards, we confront both versions with an audience, and measure and gather their arousal towards each situation. This allows statistical data to be collected, which are then examined to determine how the modifications influenced users' attitudes (Optimizely, 2017).

A/B testing should be considered as a performance measurement tool; indeed, it empowers companies as well as individuals in statistically assessing their websites and/or apps, using a two-phase process, which allows the precise change to be determined in order to proceed. This can result in giving valuable and proven outcomes on users' experience that guide interface design. This technique could be used endlessly in order to improve a definite goal over time; however, A/B testing only allows one feature to be tested at the time, which, in fact, results in specifying a local maximum. In other words, this procedure cannot analyze a set of features; on the contrary, it focuses only on one, then another.

7.2.2. Multivariate testing

Regarding the second assessment method, namely the multivariate testing, which is often used as a synonym for A/B testing mistakenly, this involves testing several versions of various features of a web page or app together. On the website of Optimizely, it is specified that "multivariate testing is a technique for testing a hypothesis in which multiple variables are modified. The goal of multivariate testing is to determine which combination of variations performs the best out of all of the possible combinations" (Optimizely, 2017, para. 1). This approach is mostly used by big companies, such as Google or Facebook. The needs of such organizations, in terms of website structure and design, mean that they must reach very high standards, and produce a high traffic level to establish relevant results. Therefore, allowing them to assess a set of items together, represents a huge advantage.

The functioning of multivariate testing is comparable to the A/B testing, except that the A/B testing only focuses on one item at the time, while the multivariate testing includes various items and examines them all together. The reason for this is to determine the key association that will enhance the main target effectively. This technique is similar to A/B testing, which uses a matrix algebra and statistical analysis (Optimizely, 2017; Pemberton Levy, 2015).

This process has advantages and drawbacks. On the one hand, it offers extended testing which enables the testing procedure to be shortened, to that in A/B testing, if used properly. On the other hand, multivariate testing requires significant traffic in order to produce significant outcomes. On top of that, it requires a pre-test phase, which involves evaluating the traffic level needed to reach statistically valuable results.

Multivariate testing is thus also classified as an effective optimization technique. It allows priceless user data to be collected, offers a deep perspective on their attitudes, and most importantly establishes a thorough action plan regarding which strategic variations to bring. This approach undeniably addresses the shakiness of implementing blind changes.

7.2.3. Bayesian test

The third and final assessment method developed in this section is known as the Bayesian test. Testing in graphical user interfaces has become increasingly important; however, it remains a challenging field of study, due to the large amount of event interactions. As a matter of fact, the infinite amount of event interaction possibilities prevents all of them from being tested. Therefore, there is a need for new testing methods to be used to improve the process of detecting defects. Fortunately, Bayesian testing enables us to do so. Gelman et al. (2014) define the Bayesian inference as follows: “the process of fitting a probability model to a set of data and summarizing the result by a probability distribution on the parameters of the model and on unobserved quantities such as predictions for new observations” (Gelman et al., 2014, p.1).

This approach refers to a very specific model optimization scheme that enables testing to be speeded up and results to be stronger. The Bayesian test is more flexible compared to other techniques as it adapts itself overtime and has a large scope of study, which is important in graphical user interface testing. Unlike A/B testing, this technique is characterized as theoretical or deductive, which allows data information to be combined with a priori knowledge, coming from previous studies or from expert advice. This to obtain a set of a posteriori information. The Bayesian approach thus focuses on a priori law. The statistics jointly measure the uncertainty of what you are going to give up with the A variation and the uncertainty of what you get with the B variation. By having an accurate view of earnings, in the form of confidence intervals, informed decisions can be formed accordingly (Brebion, 2016). Figure 10 below illustrates the Bayesian statistical engine.

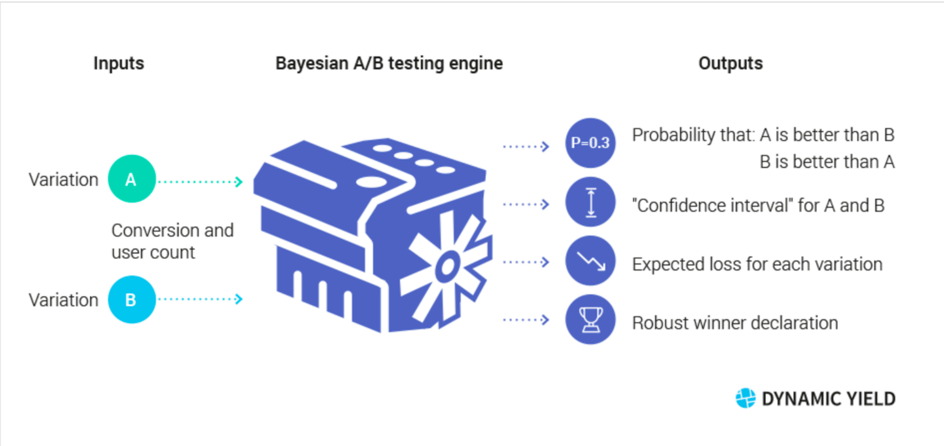


Figure 22 - Illustration of the Bayesian statistical engine

Source: Michaeli, I. (2016). *Going Bayesian: a fresh approach to A/B testing.* Available on the website Dynamic Yield: https://www.dynamicyield.com/2016/09/bayesian-testing/

This method is robust when it is possible to refer to similar previous experiences. The main advantage of this approach is therefore that it gets rid of fixed horizons and that results can be produced quickly. It is unnecessary to fix the size of a necessary sample beforehand and a level of traffic to carry out a test; rather, the results can be consulted throughout the experiment and are thus faster to obtain (Michaeli, 2016). However, the Bayesian statistic presents some disadvantages, such as being much more difficult to apprehend. Bayesian statistics try to calculate a probability distribution, which is a more complex concept than a simple indicator of confidence (Michaeli, 2016).

In brief, these three techniques provide useful data and insights for reaching definite goals. Indeed, these methods should be used for decision making in terms of performance as well as aesthetics, and thus interface design too. Each of them has their own advantages and drawbacks, and should all be considered when deciding which methods to implement. They are undeniably essential to succeed online, as they enable optimizations in interface design leading to better performances. Besides, we are convinced that A/B testing and Bayesian testing both present interesting advantages that could allow "machine learning" to be investigated further.

Thanks to our statistical analysis, we could demonstrate the relevance of QUESTIM for partially measuring the aesthetics of GUI. However, as explained previously, we are convinced that it should be paired with other measurement methods, in order to provide deeper insights not only on the aesthetics, but also on the performance and users experience, which, as we read in the literature, are highly correlated (Forgaty et al., 2000). These features are key when it comes to designing great interfaces; therefore, improving how they are assessed seems fundamental. To do so, we thought about the concept of "machine learning", which could be an adapted version of QUESTIM. Using probability distribution would be more in keeping with the subjective character of aesthetics evaluation, which would ultimately lead to a more balanced response from the computer in charge of comparing the aesthetics of several interfaces. It would go beyond predicting the aesthetics, and would provide the probability of aesthetics appreciation. This procedure would allow an aesthetic judgment to be obtained combined with a level of confidence of the relevance of this judgment. Thanks to its probability distribution, machine learning would ease the designing process, because it could generate GUIs and test each parameter very quickly, until it comes across the "perfect match". A main advantage of such a technique, is its speed and effectiveness in outcomes supply. On top of that, this new method would provide designers with more information, such as the performance level, which therefore allow them to focus on user experience optimization.

8. Conclusion

8.1. Summary of the study

In order to fill the gaps that still exist in the literature regarding the aesthetics of graphical user interface, we conducted a quantitative study based on empirical data, from previous analysis, coupled with new data collected through a dedicated applicative software, QUESTIM. The purpose of our study was to determine whether QUESTIM provided appropriate and adequate aesthetics measures for graphical user interfaces or not, by comparing its results to previous studies' outcomes (Reinecke's study). Our theoretical researches enabled us to get a better insight on the topic and its key concepts, but also to come up with several research questions and assumptions regarding our applicative software, which are the following:

Research Question 1: Is the aesthetics judged the same way by a human being and by a semi-automated software (QUESTIM)?

- *Hypothesis 1: There is a link between the same type of variables computed in QUESTIM and those computed in Reinecke's study.*
- *Hypothesis 2: The variables computed in each study impact in the same way the visual appeal (aesthetics score).*

Research Question 2: Do both studies measure variables that enable the aesthetics to be judged accurately?

- *Hypothesis 3: All the studied variables (as measured by the specific formulas) influence the visual appeal (aesthetics score).*
- *Hypothesis 4: The variables influencing the visual appeal (aesthetics score) can be ranked.*

Research Question 3: Would it be possible to anticipate the aesthetics success of a Graphical Interface?

- *Hypothesis 5: The ten templates show similar results as the 417 websites previously analyzed after being processed in QUESTIM.*

Through our statistical analysis, we examined specific sets of variables, by gathering only those judged as the most relevant in our case. The first series of tests performed on our sample of 430 interfaces allowed us to validate, even if sometimes partially, all our hypotheses, which enabled us to draw some interesting conclusions. Therefore, we could assert that the dedicated applicative software QUESTIM did provide appropriate aesthetics measures, although some appeared more relevant than others. Testing our assumptions induced the isolation of some measures as they did not produce significant results. By the end of the testing phase, we could count five significant variables.

Then in the last part of our statistical analysis, we wanted to push our analysis further; therefore, we applied QUESTIM to a new sample of websites (best-seller templates from Themeforest), aiming to evaluate if their aesthetics success could be anticipated. The strong outcomes allowed us to reinforce our previous conclusions, and validate once again QUESTIM's ability to measure the aesthetics. On top of that, as mention earlier, Uribe et al. (2017) suggested another method using fewer formulas to evaluate the aesthetic, which should be considered for such anticipating tests.

However, this software is not the best in measuring the aesthetics and presents some weaknesses, which will be developed below. Therefore, although it already offers great knowledge on GUI aesthetics, we are convinced that QUESTIM should be paired with other evaluation methods to get the best GUI assessment possible; hence the suggestion of other measurement methods for future analysis. We believe that adapting and reviewing QUESTIM features could lead to the development of highly productive software, which could benefit and maybe revolutionize the domain of GUI conception in human computer interaction.

In short, this study has allowed us to test the relevance of our theoretical framework, as well as realize that factors other than the equilibrium, the simplicity, the balance, the proportion or the compressioncomplexity, play a role in predicting the aesthetics. This has allowed us to suggest new perspective for futures analysis on this matter.

Finally, there are some limitations associated with our research method, some of which have already been expressed earlier. These limits thus give room for fresh research opportunities. Therefore, we have suggested that it would be interesting to continue the research using a mixed approach, combining QUESTIM with other evaluation techniques, judged as relevant.

8.2. Study limits and threats of validity

In conclusion, it is important to recall the limitations of our study.

Firstly, we decided to study a specific sample, gathering a set of 430 websites, which had already been empirically validated, in terms of distribution and population representation in a previous study. However, we are aware that this database included some outdated websites. Therefore, one should be cautious when resorting to our framework model for future analysis. The results might vary along with the sample chosen for the study. However, that sample still allowed us to validate our research question and assumptions. Indeed, we believed it was the most appropriate for our study because it fitted the best our needs.

Secondly, as uncovered through our literature review there are many ways to examine and evaluate GUI aesthetics. In our research, we decided to focus on measuring the aesthetics using dedicated applicative software, QUESTIM, and to analyze closely its results. However, the software has shown some weaknesses. It is not the most user-friendly system, and requires some time to get used to the complex procedure. Moreover, in the second step of our statistical analysis, we intended to see if it was possible to anticipate the aesthetics success of ten best-seller templates issued from Themeforest. Hence, we applied QUESTIM to those ten templates. Of course, it would have been interesting and enriching to our findings, to reproduce the procedure on the ten worst templates, but we were not able to do so as it was difficult to determine which of the templates available on the list were actually bad. Moreover, we observed that QUESTIM presented some problems, hence, we would recommend an update. That being said, our analysis enabled us to claim that the software was a strong and powerful tool for measuring GUI aesthetics.

Thirdly, our analysis focused only on the assessment of static websites screens, which limits the results. As reviewed in the literature, the usability is a key factor for determining the aesthetics, therefore, we are convinced that conducting a similar study, but this time, assessing more dynamic and more complex websites designs would definitely be interesting. The new outcomes and model would bring fresher insights. In addition, as user interfaces keep evolving and expand to other devices, such as smartphones, tablets, etc., we believe it would also be interesting to extend the analysis toward new devices to develop a new and more complete framework.

Fourthly, the aesthetic level of an interface might vary not only along with the objects of the screen, but also along with the textual content of each interface. Therefore, the textual aspect might as well impact user's experience and their evaluation. In our study, we resorted to the applicative software QUESTIM, which does not consider any textual aspect. The software only computes specific variables and none of them refer to textual content. Nonetheless, we are persuaded that the wording used on the interfaces might influence user's emotions and therefore their appreciation. It would thus be interesting to further study this aspect, and compare the results.

Despite these limits, the deep research, from both a theoretical and a statistical point of view, led us to new and fresh perspectives that should be considered for futures analysis. This in order to keep up with the continuously evolving trends of the digital era we currently live in. Besides, we are convinced they would produce promising results, and that they would always allow advances in the domain of human computer interaction.

9. Bibliography

Aldrich, D. F. (1999). *Mastering the digital marketplace: practical strategies for competitiveness in the new economy*. New York, NY: John Wiley & Sons.

Amerland, D. (2015). *Why branding is the space where your business values meet your customers' minds.* Available on: https://davidamerland.com/seo-blog/1052-why-branding-is-the-space-where-your-business-values-meet-your-customers-minds.html

Arnheim, R. (1954). *Art and Visual Perception: a Psychology of the Creative Eye*. (3rd ed.). Los Angeles, CA: University of California Press.

Bäumer, D., Bischofberger, W. R., Lichter, H., & Züllighoven, H. (1996, May). User interface prototyping—concepts, tools, and experience. In *Proceedings of the 18th international conference on Software engineering* (pp. 532-541). IEEE Computer Society.

Birkhoff, G. D. (1933). *Aesthetic measure* (Vol. 9). Cambridge, MA: Harvard University Press.

Bloch, P. H. (1995). Seeking the ideal form: Product design and consumer response. *The Journal of Marketing*, 59, 16-29.

Bonaventure, O., Detal, G., & Paasch, C. (2015). *Systèmes informatiques. SINF1252*. Syllabus, Université catholique de Louvain, Diffusion Universitaire CIACO. Available on : https://sites.uclouvain.be/SystInfo/distrib/SINF1252-Theorie.pdf

Bonsiepe, G. (1990). INTERFACE DESIGN-LANGUAGE-GRAPHICS: Interpretations of Human User Interface. *Visible Language*, *24*(3), 262-285.

Brebion, A. (2016). *Comment améliorer ses taux de conversion avec les données et les nouveaux algorithmes ?* Available on Frenchweb.fr, le magazine de l'innovation : http://www.frenchweb.fr/comment-ameliorer-ses-taux-de-conversion-avec-les-donnees-et-les-nouveaux-algorithmes/243331

80.

Chen, M., & Ryu, Y. U. (2013). Facilitating effective user navigation through website structure improvement. *IEEE Transactions on Knowledge and Data Engineering*, *25*(3), 571-588.

Clay, F. (1908). The origin of the aesthetic emotion. *Sammelbande der Int. Musik.* 282–290.

Coleman, N. (2007). *Utopias and architecture* (2nd ed.). Padstow, Cornwall: Routledge.

Corneille, O., Monin, B., & Pleyers, G. (2005). Is positivity a cue or a response option? Warm glow vs evaluative matching in the familiarity for attractive and not-so-attractive faces. *Journal of Experimental Social Psychology*, *41*(4), 431-437.

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*. Los Angeles, CA: Sage publications.

Darden, W. R., & Babin, B. J. (1994). Exploring the concept of affective quality: expanding the concept of retail personality. *Journal of Business Research*, *29*(2), 101-109.

Feagin, S. L., & Maynard, P. (Eds.). (1997). *Aesthetics* (p. 84). Oxford: Oxford University Press.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: Sage publications.

Fogarty, J., Forlizzi, J., & Hudson, S. E. (2001, November). Aesthetic information collages: generating decorative displays that contain information. In *Proceedings of the 14th annual ACM symposium on User interface software and technology* (pp. 141-150). ACM.

Galitz, W. O. (2007). *The essential guide to user interface design: an introduction to GUI design principles and techniques* (3rd ed.). Indianapolis, IN: John Wiley & Sons.

Gardner, S. (2002). Aesthetics, in N. Bunnin and E. P. Tsui-James (eds.), T*he Blackwell Companion to Phylosophy* (pp. 232-256). Oxford: Blackwell Publishing Ltd.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (Vol. 2). Boca Raton, FL: CRC press.

Gregor, M. J. (1983). Baumgarten's" Aesthetica". *The Review of Metaphysics*, 357-385.

Harding, B. A. (1989, October). Windows & Icons & Mice, Oh My! The Changing Face of Computing. In *Frontiers in Education Conference* (pp. 337-342). IEEE.

Hayes, F., & Baran, N. (1989). A guide to GUIs. *Byte*, *14*(7), 250-257.

Hoegg, J., Alba, J. W., & Dahl, D. W. (2010). The good, the bad, and the ugly: Influence of aesthetics on product feature judgments. *Journal of Consumer Psychology*, *20*(4), 419-430.

Hofstede, F. T., Steenkamp, J. B. E., & Wedel, M. (1999). International market segmentation based on consumer-product relations. *Journal of Marketing Research*, 1-17. doi: 10.2307/3151911

IBM. (2017). *IBM SPSS Software*. Available on: https://www.ibm.com/analytics/us/en/technology/spss/

Cisco, V.N.I. (2017). Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021, white paper.

ISO25000. (2017). *ISO/IEC25010.* Available on http://iso25000.com/index.php/en/iso-25000-standards/iso-25010.

Jansen, B. J. (1998). The graphical user interface. *ACM SIGCHI Bulletin*, *30*(2), 22-26.

Jordan, P. W. (1998). Human factors for pleasure in product use. *Applied ergonomics*, *29*(1), 25-33.

Kant, I. (1987). *Critique of judgment* (W. S. Pluhar, Trad.). Indianapolis, IN: Hackett Publishing. (Original version published in 1790).

Kotler, P., & Alexander Rath, G. (1984). Design: A powerful but neglected strategic tool. *Journal of business strategy*, *5*(2), 16-21.

82.

Kruft, H.-W. (1994). *A History of Architectural Theory: From Vitruvius to the Present* (1st ed.). New York, NY: Princeton Architectural Press.

Kurosu, M., & Kashimura, K. (1995, May). Apparent usability vs. inherent usability: experimental analysis on the determinants of the apparent usability. In *Conference companion on Human factors in computing systems* (pp. 292-293). ACM.

Laudon, K. C., & Laudon, J. P. (2004). *Management information systems: Managing the digital firm* (8th ed.). Upper Saddle River, NJ: Pearson.

Lavie, T., & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites. *International journal of human-computer studies*, *60*(3), 269-298.

Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British journal of psychology 95*(4), 489–508.

Merriam-Webster Dictionary. (2017). *Definition of aesthetic*. Available on: https://www.merriam-webster.com/dictionary/aesthetic

Michaeli, I. (2016). *Going Bayesian: a fresh approach to A/B testing.* Available on the website Dynamic Yield: https://www.dynamicyield.com/2016/09/bayesian-testing/

Moshagen, M., & Thielsch, M. T. (2010). Facets of visual aesthetics. *International Journal of Human-Computer Studies*, *68*(10), 689-709.

Myers, B. A., & Rosson, M. B. (1992, June). Survey on user interface programming. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 195-202). ACM.

Neale, W., & McCombe, C. (1997). Designing usable and visually appealing web sites. In *CHI'97 Extended Abstracts on Human Factors in Computing Systems* (pp. 170-171). ACM.

Ngo, D. C. L., & Byrne, J. G. (2001). Another look at a model for evaluating interface aesthetics. *International Journal of Applied Mathematics and Computer Science*, *11*, 515-535.

Ngo, D. C. L., Samsudin, A., & Abdullah, R. (2000). Aesthetic measures for assessing graphic screens. *J. Inf. Sci. Eng*, *16*(1), 97-116.

Norman, D. (2013). *The design of everyday things: Revised and expanded edition*. New York, NY: Basic Books.

Norman, D. A. (2005). *Emotional design: Why we love (or hate) everyday things*. New York, NY: Basic books.

Online Oxford Dictionaries. (2017). *Definition of aesthetic*. Available on: https://en.oxforddictionaries.com/definition/aesthetic

Optimizely. (2017). *A/B testing.* Available on: https://www.optimizely.com/ab-testing/

Optimizely. (2017). *What is multivariate testing?* Available on: https://www.optimizely.com/optimization-glossary/multivariate-testing/

Pemberton Levy, H. (2015). *The Secret to A/B testing for Customer Experience*. Available on the website of Gartner Inc.: http://www.gartner.com/smarterwithgartner/the-secret-to-ab-testing-for-customer-experience/

Porteous, J.D. (1997). Environmental Aesthetics: Ideas, Politics and Planning. *Environment International*, *1*(23), 142.

Queensland Government. (2016). *Creating a website*. Available on: https://www.business.qld.gov.au/starting-business/internet-start-ups/online-basics/creating-websites

Rantalla, V. (1994). *Environmental experience: Beyond aesthetic subjectivism and objectivism*. The Thingmount Working Paper Series on the Philosophy of Conservation", TWP 99-104, Lancaster: Lancaster University.

Reimann, M., Zaichkowsky, J., Neuhaus, C., Bender, T., & Weber, B. (2010). Aesthetic package design: A behavioral, neural, and psychological investigation. *Journal of Consumer Psychology*, *20*(4), 431-441.

Reinecke, K., Yeh, T., Miratrix, L., Mardiko, R., Zhao, Y., Liu, J., & Gajos, K. Z. (2013, April). Predicting users' first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2049-2058). ACM.

Riglis, E. (1998). *Modeling complexity in visual images*. (PhD thesis). Heriot-Watt University, Edinburgh.

Robins, D., & Holmes, J. (2008). Aesthetics and credibility in web site design. *Information Processing & Management*, *44*(1), 386-399.

Rosen, D. E., & Purinton, E. (2004). Website design: Viewing the web as a cognitive landscape. *Journal of Business Research*, *57*(7), 787-794.

Russell, J. A. (1988). Affective appraisals of environments. In: Nasar, J.L. (Ed.), *Environmental Aesthetics, Theory, Research, and Applications* (pp.120-132). Cambridge: Cambridge University Press.

Russell, J. A., & Pratt, G. (1980). A description of the affective quality attributed to environments. *Journal of personality and social psychology*, *38*(2), 311-322.

Saffer, D. (2010). *Designing for interaction: creating innovative applications and devices* (2nd ed.). Berkeley, CA: New Riders.

Sano, D. (1997). Designing large-scale web sites. *SIGCHI Bulletin*, *29*(3).

Saw, R., & Osborne, H. (1968). Aesthetics as a branch of philosophy. In: Osborn, H. (Ed.), *Aesthetics in the Modern World*. New York, NY: Weybright and Talley.

Schenkman, B. N., & Jönsson, F. U. (2000). Aesthetics and preferences of web pages. *Behaviour & Information Technology*, *19*(5), 367-377.

Servigne, S. (2017). Systèmes Informatiques – Conception, architecture et urbanisation des systèmes d'information. *Encyclopædia universalis* online. Available on : http://www.universalis.fr/encyclopedie/systemes-informatiques-conception-architecture-et-urbanisation-des-systemes-d-information/

Streveler, D. J., & Wasserman, A. I. (1984, September). Quantitative measures of the spatial properties of screen designs. In *Proceedings of the IFIP TC13 First International Conference on Human-Computer Interaction* (pp. 81-89).

Swaddle, J. P., & Cuthill, I. C. (1995). Asymmetry and human facial attractiveness: symmetry may not always be beautiful. *Proceedings of the Royal Society of London B: Biological Sciences*, *261*(1360), 111-116.

Swede, G. (1994). Basic theoretical issues. *The Psychology of Art: An Experimental Approach*. Toronto: Canadian Scholars' Press Inc.

Thongpapanl, N., & Ashraf, A. R. (2011). Enhancing online performance through website content and personalization. *Journal of computer information systems*, *52*(1), 3-13.

Tractinsky, N. (1997, March). Aesthetics and apparent usability: empirically assessing cultural and methodological issues. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems* (pp. 115-122). ACM.

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

Tuch, A. N., Bargas-Avila, J. A., & Opwis, K. (2010). Symmetry and aesthetics in website design: It'sa man's business. *Computers in Human Behavior*, *26*(6), 1831-1837.

Tuch, A. N., Presslaber, E. E., StöCklin, M., Opwis, K., & Bargas-Avila, J. A. (2012). The role of visual complexity and prototypicality regarding first impression of websites: Working towards understanding aesthetic judgments. *International Journal of Human-Computer Studies*, *70*(11), 794-811.

Tullis, T. S. (1988). Screen design, In M. Helander (ed.), *Handbook of Human-Computer Interaction* (pp. 377- 411). Amsterdam: Elsevier Science Publishers.

Uribe, S., Álvarez, F., & Menéndez, J. M. (2017). User's Web Page Aesthetics Opinion: A Matter of Low-Level Image Descriptors Based on MPEG-7. *ACM Transactions on the Web (TWEB)*, *11*(1), 5.

Van Bremen, E. J. J., Knoop, W. G., Horvath, I., Vergeest, J. S., & Pham, B. (1998, September). Developing a methodology for design for aesthetics based on analogy of communication. In *Proc. of the 1998 ASME Design Engineering Technical Conferences*.

Van Damme, W. (1996). *Beauty in context: Towards an anthropological approach to aesthetics* (Vol. 17). Brill, Leiden.

Van der Heijden, H. (2003). Factors influencing the usage of websites: the case of a generic portal in The Netherlands. *Information & management*, *40*(6), 541-549.

Van Hoff, A. A. (1998). *U.S. Patent No. 5,802,530*. Washington, DC: U.S. Patent and Trademark Office.

Vanderdonckt, J., & Gillo, X. (1994, June). Visual techniques for traditional and multimedia layouts. In *Proceedings of the workshop on Advanced visual interfaces* (pp. 95-104). ACM.

Whitney, D.E. (1988). Manufacturing by design. *Harvard Business Review*, 66(4), 83–90.

Wolfram Mathworld. (2017). *Golden ratio*. Available on: http://mathworld.wolfram.com/GoldenRatio.html

Zen, M. (2017). *A methodology for assessing aesthetics of a Graphical User Interface of an Information System by (Semi-) automatic Evaluation of Visual Metrics.* (PhD thesis). Université catholique de Louvain, Louvain-la-Neuve.

Zen, M., and Vanderdonckt, J. (2016). Assessing user interface aesthetics based on the inter-subjectivity of judgment. In *Proceedings of the 30th International BCS Human Computer Interaction Conference: Fusion!* (p. 25). BCS Learning & Development Ltd.

Zhang, P., & Li, N. (2004). Love at first sight or sustained effect? The role of perceived affective quality on users' cognitive reactions to information technology. *ICIS 2004 Proceedings*, 22.

10. Appendices

Appendix 1 – Correlation between rawcolorfulness and colorfulness (hypothesis 1).

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
rawcolorfulness	417	1,00000000	9,00000000	4,77053647	1,51458070	,085	,120	-,540	,238
Colorfulness	417	,039177424	1,00000000	,522543034	,238808652	,255	,120	-,609	,238
Valid N (listwise)	417								

Correlations

		rawcolorfulness	Colorfulness
rawcolorfulness	Pearson Correlation	1	,344**
	Sig. (2-tailed)		,000
	N	417	417
Colorfulness	Pearson Correlation	,344**	1
	Sig. (2-tailed)	,000	
	N	417	417

**. Correlation is significant at the 0.01 level (2-tailed).

Appendix 2 – Correlation between rawcomplexity and compressioncomplexity (hypothesis 1).

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
rawcomplexity	417	1,00000000	9,00000000	4,78149139	1,50782994	,077	,120	-,578	,238
compressionComplexity	417	,187950792	,802450951	,441883347	,110004428	,380	,120	,326	,238
Valid N (listwise)	417								

Correlations

		rawcomplexity	compression Complexity
rawcomplexity	Pearson Correlation	1	,757**
	Sig. (2-tailed)		,000
	N	417	417
compressionComplexity	Pearson Correlation	,757**	1
	Sig. (2-tailed)	,000	
	N	417	417

**. Correlation is significant at the 0.01 level (2-tailed).

Appendix 3 – SPSS output of the multiple linear regression of the score, compressioncomplexity, colorfulness, rawcolorfulness, rawcomplexity (hypothesis 2).

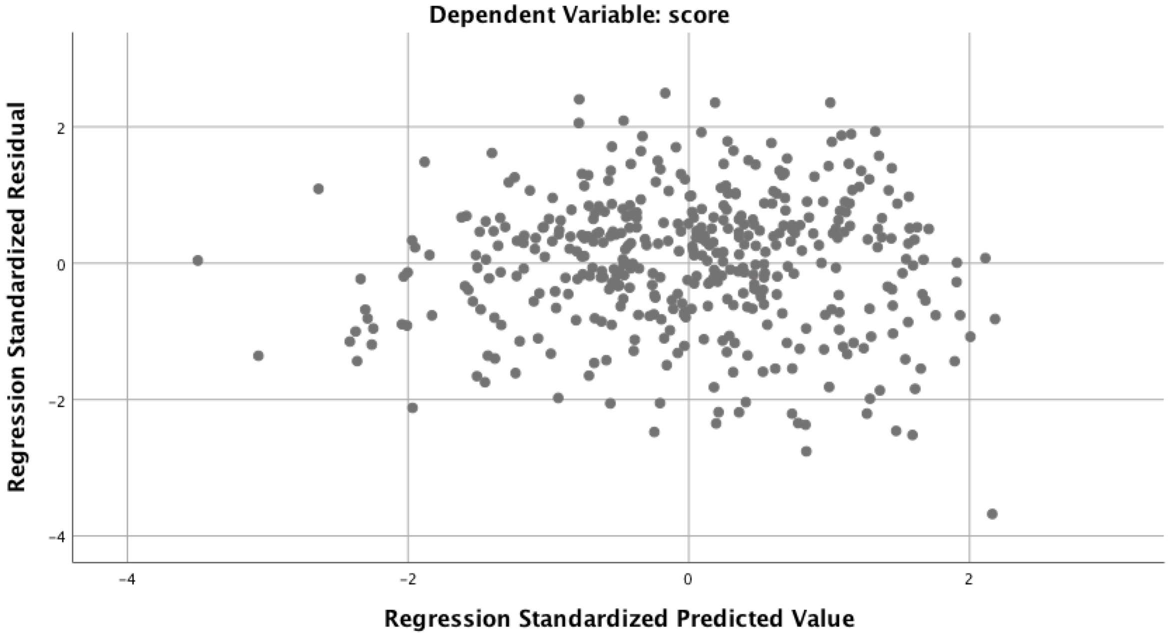
- Descriptive analysis

Descriptive Statistics

	Mean	Std. Deviation	N
score	4,36921610	1,03965309	417
rawcomplexity	4,78149139	1,50782994	417
rawcolorfulness	4,77053647	1,51458070	417
Colorfulness	,522543034	,238808652	417
compressionComplexity	,441883347	,110004428	417

Scatterplot

Dependent Variable: score



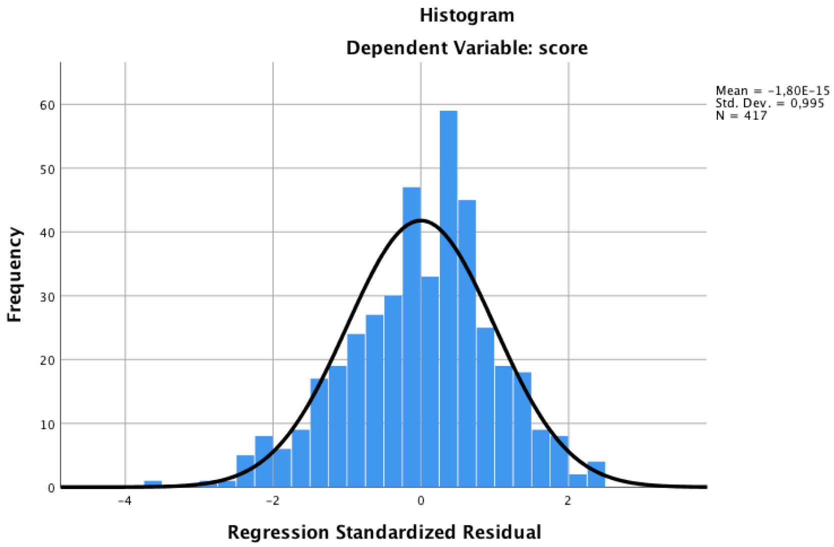
On the scatterplot above, we observe that there is no tendency in the dispersion of the observations, and that most the observations are between -2 and 2. This allow us to say that our sample is normally distributed and that there is no tendency in the distribution.

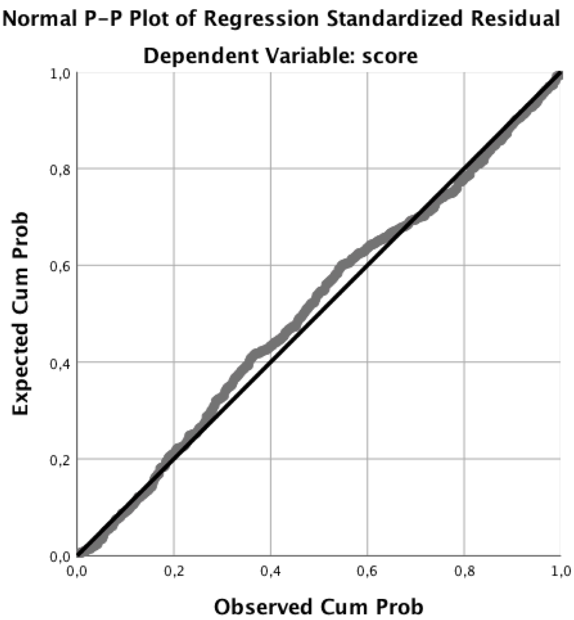
- Application Conditions for the regular multiple linear regression

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	2,79855776	5,34904051	4,36921610	,449039756	417
Residual	-3,4665318	2,35022545	,000000000	,937678969	417
Std. Predicted Value	-3,498	2,182	,000	1,000	417
Std. Residual	-3,679	2,494	,000	,995	417

a. Dependent Variable: score





The histogram and the normal p-p plot allow us to say that our sample is normally distributed as it follows a bell-shaped curve. The normality condition is therefore validated.

Correlations

		score	rawcomplexity	rawcolorfulness	Colorfulness	compressionComplexity
Pearson Correlation	score	1,000	-,363	-,369	-,074	-,415
	rawcomplexity	-,363	1,000	,944	,339	,757
	rawcolorfulness	-,369	,944	1,000	,344	,775
	Colorfulness	-,074	,339	,344	1,000	,360
	compressionComplexity	-,415	,757	,775	,360	1,000
Sig. (1-tailed)	score	.	,000	,000	,065	,000
	rawcomplexity	,000	.	,000	,000	,000
	rawcolorfulness	,000	,000	.	,000	,000
	Colorfulness	,065	,000	,000	.	,000
	compressionComplexity	,000	,000	,000	,000	.
N	score	417	417	417	417	417
	rawcomplexity	417	417	417	417	417
	rawcolorfulness	417	417	417	417	417
	Colorfulness	417	417	417	417	417
	compressionComplexity	417	417	417	417	417

Model Summary[b]

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	,432[a]	,187	,179	,942219814	,187	23,621	4	412	,000

a. Predictors: (Constant), compressionComplexity, Colorfulness, rawcomplexity, rawcolorfulness

b. Dependent Variable: score

On this table, we observe that adjusted R Square amounts to 0,179, which implies that 17,9% of the total variation of the score is explained by the variation of the independent variables.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	83,881	4	20,970	23,621	,000[b]
	Residual	365,765	412	,888		
	Total	449,645	416			

a. Dependent Variable: score

b. Predictors: (Constant), compressionComplexity, Colorfulness, rawcomplexity, rawcolorfulness

Hypothesis test

$$H_0: R^2 = 0$$

$$H_1: R^2 \neq 0$$

In this table, we observe that the p-value of the third model is smaller than 0,05. Therefore, we can reject the null hypothesis, and state that a model of multiple linear regression exists.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	6,041	,195		30,964	,000		
	rawcolorfulness	-,050	,096	-,073	-,521	,603	,101	9,880
	rawcomplexity	-,047	,093	-,068	-,500	,618	,108	9,270
	Colorfulness	,422	,209	,097	2,020	,044	,859	1,164
	compressionComplexity	-3,240	,677	-,343	-4,784	,000	,385	2,600

a. Dependent Variable: score

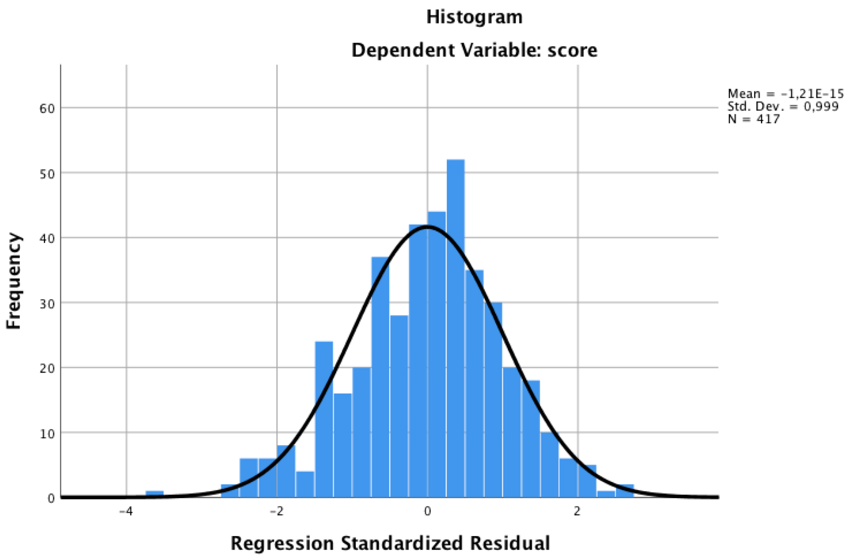
The coefficient table presents the significance of our variables. We observe that only the colorfulness and the compressioncomplexity are significant, which implies that these variables play a role in predicting the aesthetics score. Besides, we observe that the VIF is lower than 10 which implies that there is no multicollinearity between these three variables

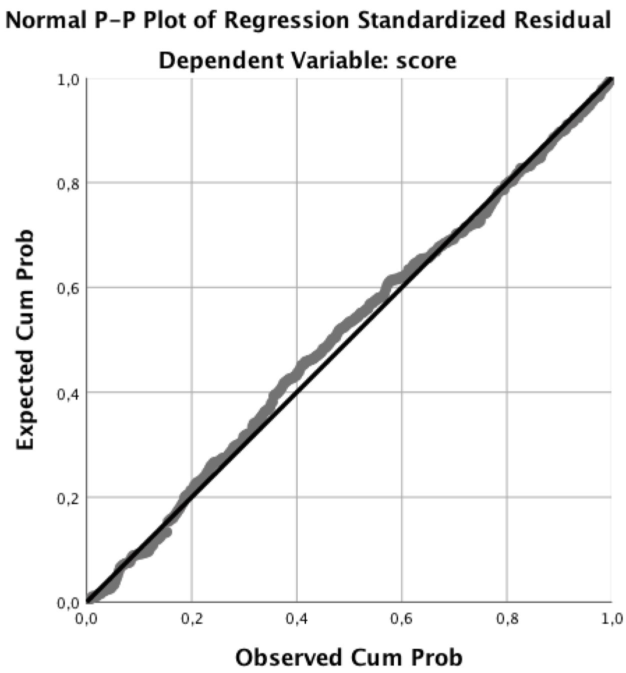
- Application Conditions for the stepwise multiple linear regression

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	2,95342970	5,36629486	4,36921610	,431937803	417
Residual	-3,4926958	2,55298138	,000000000	,945678744	417
Std. Predicted Value	-3,278	2,308	,000	1,000	417
Std. Residual	-3,689	2,696	,000	,999	417

a. Dependent Variable: score





The histogram and the normal p-p plot allow us to say that our sample is normally distributed as it follows a bell-shaped curve. The normality condition is therefore validated.

Correlations

		score	rawcolorfulness	rawcomplexity	Colorfulness	compression Complexity
Pearson Correlation	score	1,000	-,369	-,363	-,074	-,415
	rawcolorfulness	-,369	1,000	,944	,344	,775
	rawcomplexity	-,363	,944	1,000	,339	,757
	Colorfulness	-,074	,344	,339	1,000	,360
	compressionComplexity	-,415	,775	,757	,360	1,000
Sig. (1-tailed)	score	.	,000	,000	,065	,000
	rawcolorfulness	,000	.	,000	,000	,000
	rawcomplexity	,000	,000	.	,000	,000
	Colorfulness	,065	,000	,000	.	,000
	compressionComplexity	,000	,000	,000	,000	.
N	score	417	417	417	417	417
	rawcolorfulness	417	417	417	417	417
	rawcomplexity	417	417	417	417	417
	Colorfulness	417	417	417	417	417
	compressionComplexity	417	417	417	417	417

Model Summary[b]

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	,415[a]	,173	,171	,946817430	,173	86,577	1	415	,000

a. Predictors: (Constant), compressionComplexity

b. Dependent Variable: score

On this table, we observe that adjusted R Square amounts to 0,171, which implies that 17,1 % of the total variation of the score is explained by the variation of the independent variables.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	77,613	1	77,613	86,577	,000[b]
	Residual	372,032	415	,896		
	Total	449,645	416			

a. Dependent Variable: score

b. Predictors: (Constant), compressionComplexity

Hypothesis test

$$H_0: R^2 = 0$$
$$H_1: R^2 \neq 0$$

In this table, we observe that the p-value of the third model is smaller than 0,05. Therefore, we can reject the null hypothesis, and state that a model of multiple linear regression exists.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	6,104	,192		31,768	,000		
	compressionComplexity	-3,927	,422	-,415	-9,305	,000	1,000	1,000

a. Dependent Variable: score

The coefficient table presents the significance of our variables. We observe that only the compressioncomplexity is significant, which implies that it plays a role in predicting the aesthetics score.

Excluded Variables[a]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics		
						Tolerance	VIF	Minimum Tolerance
1	rawcolorfulness	-,117[b]	-1,663	,097	-,081	,400	2,499	,400
	rawcomplexity	-,113[b]	-1,660	,098	-,081	,426	2,345	,426
	Colorfulness	,086[b]	1,806	,072	,088	,871	1,149	,871

a. Dependent Variable: score

b. Predictors in the Model: (Constant), compressionComplexity

This output presents the variables that have been excluded when running each model.

Appendix 4 – SPSS output of the multiple linear regression of the score, balance, equilibrium, density, concentricity, simplicity, proportion, symmetry (hypothesis 3 & 4).

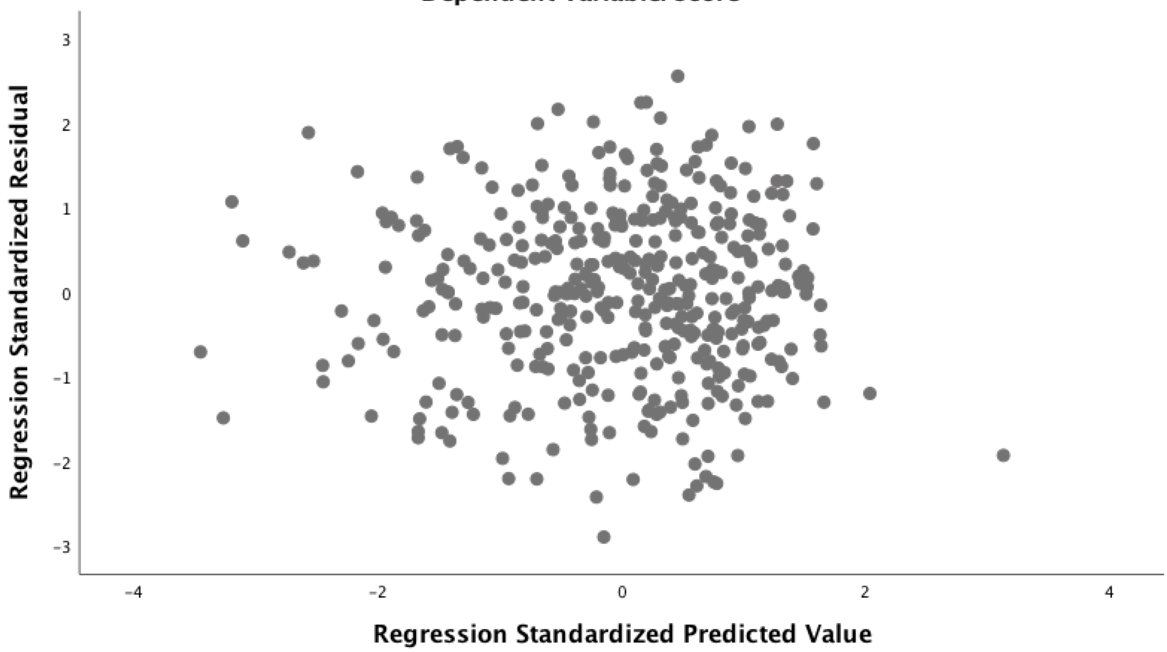
- Descriptive analysis

Descriptive Statistics

	Mean	Std. Deviation	N
score	4,36921610	1,03965309	417
Balance	,758764330	,191720153	417
Equilibrium	,790545343	1,49076003	417
Density	,646472945	,236418016	417
Concentricity	,488724678	,077613173	417
Simplicity	,304894771	,139264869	417
Proportion	,806894187	,354871270	417
Symmetry	,464768045	,026137190	417

Scatterplot

Dependent Variable: score



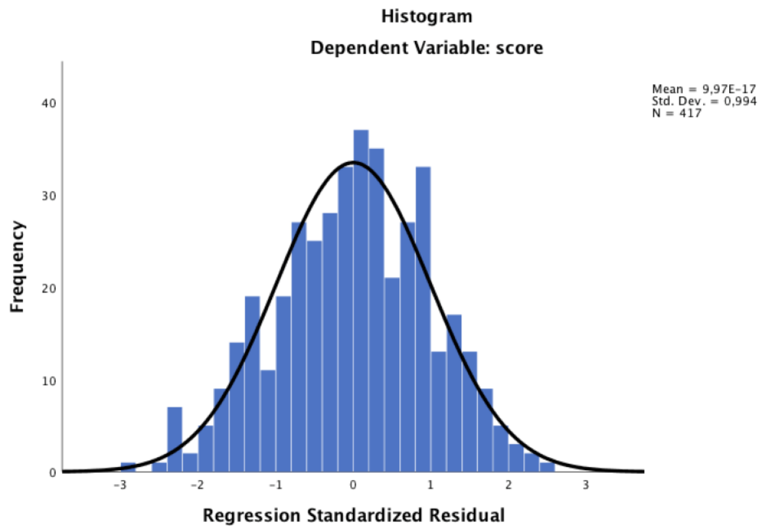
On the graph above, we observe that there is no tendency in the dispersion of the observations, and that most the observations are between -2 and 2. This allow us to say that our sample is normally distributed and that there is no tendency in our distribution.

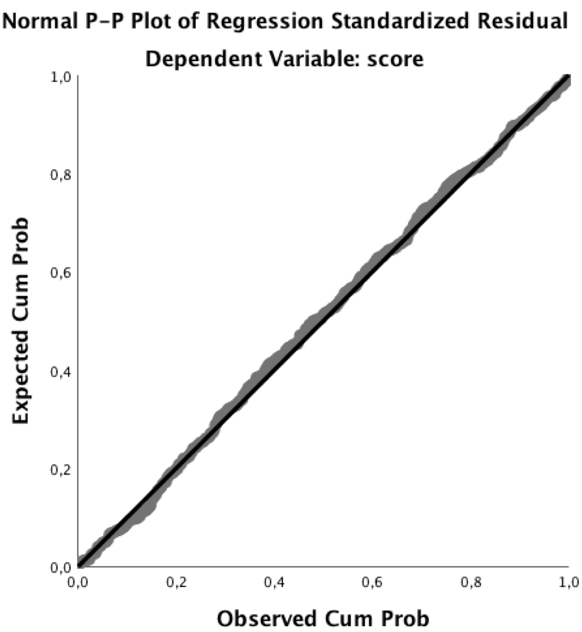
- Application Conditions

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	3,09286141	5,52530813	4,36921610	,369653104	417
Residual	-2,8280358	2,50212717	,000000000	,971717621	417
Std. Predicted Value	-3,453	3,128	,000	1,000	417
Std. Residual	-2,893	2,559	,000	,994	417

a. Dependent Variable: score





The histogram and normal p-p plot allow us to say that our sample is normally distributed as it follows a bell-shaped curve. The normality condition is therefore validated.

Correlations

		score	Balance	Equilibrium	Density	Concentricity	Simplicity	Proportion	Symmetry
Pearson Correlation	score	1,000	,122	-,001	,176	-,095	-,235	,209	,097
	Balance	,122	1,000	,462	,618	,119	-,578	,010	-,137
	Equilibrium	-,001	,462	1,000	,319	,399	-,522	-,066	-,157
	Density	,176	,618	,319	1,000	-,015	-,896	,232	-,088
	Concentricity	-,095	,119	,399	-,015	1,000	-,039	-,186	-,045
	Simplicity	-,235	-,578	-,522	-,896	-,039	1,000	-,209	,088
	Proportion	,209	,010	-,066	,232	-,186	-,209	1,000	,109
	Symmetry	,097	-,137	-,157	-,088	-,045	,088	,109	1,000
Sig. (1-tailed)	score	.	,006	,494	,000	,026	,000	,000	,024
	Balance	,006	.	,000	,000	,008	,000	,415	,003
	Equilibrium	,494	,000	.	,000	,000	,000	,089	,001
	Density	,000	,000	,000	.	,381	,000	,000	,036
	Concentricity	,026	,008	,000	,381	.	,212	,000	,178
	Simplicity	,000	,000	,000	,000	,212	.	,000	,037
	Proportion	,000	,415	,089	,000	,000	,000	.	,013
	Symmetry	,024	,003	,001	,036	,178	,037	,013	.
N	score	417	417	417	417	417	417	417	417
	Balance	417	417	417	417	417	417	417	417
	Equilibrium	417	417	417	417	417	417	417	417
	Density	417	417	417	417	417	417	417	417
	Concentricity	417	417	417	417	417	417	417	417
	Simplicity	417	417	417	417	417	417	417	417
	Proportion	417	417	417	417	417	417	417	417
	Symmetry	417	417	417	417	417	417	417	417

Model Summary[f]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,235[a]	,055	,053	1,01183196
2	,286[b]	,082	,077	,998678865
3	,307[c]	,094	,088	,992927698
4	,341[d]	,116	,108	,981966285
5	,356[e]	,126	,116	,977610444

a. Predictors: (Constant), Simplicity

b. Predictors: (Constant), Simplicity, Proportion

c. Predictors: (Constant), Simplicity, Proportion, Equilibrium

d. Predictors: (Constant), Simplicity, Proportion, Equilibrium, Density

e. Predictors: (Constant), Simplicity, Proportion, Equilibrium, Density, Balance

f. Dependent Variable: score

On this table, we observe that adjusted R Square amounts to 0,116, which implies that 11,6% of the total variation of the score is explained by the variation of the five independent variables (significant variables only).

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	24,767	1	24,767	24,191	,000[b]
	Residual	424,879	415	1,024		
	Total	449,645	416			
2	Regression	36,739	2	18,369	18,418	,000[c]
	Residual	412,907	414	,997		
	Total	449,645	416			
3	Regression	42,467	3	14,156	14,358	,000[d]
	Residual	407,179	413	,986		
	Total	449,645	416			
4	Regression	52,371	4	13,093	13,578	,000[e]
	Residual	397,274	412	,964		
	Total	449,645	416			
5	Regression	56,844	5	11,369	11,895	,000[f]
	Residual	392,802	411	,956		
	Total	449,645	416			

a. Dependent Variable: score

b. Predictors: (Constant), Simplicity

c. Predictors: (Constant), Simplicity, Proportion

d. Predictors: (Constant), Simplicity, Proportion, Equilibrium

e. Predictors: (Constant), Simplicity, Proportion, Equilibrium, Density

f. Predictors: (Constant), Simplicity, Proportion, Equilibrium, Density, Balance

Hypothesis test

$$H_0: R^2 = 0$$

$$H_1: R^2 \neq 0$$

In this table, we observe that the p-value of the fifth model is smaller than 0,05. Therefore, we can reject the null hypothesis, and state that a model of multiple linear regression exists.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	4,903	,119		41,074	,000	4,669	5,138		
	Simplicity	-1,752	,356	-,235	-4,918	,000	-2,452	-1,052	1,000	1,000
2	(Constant)	4,430	,180		24,546	,000	4,075	4,785		
	Simplicity	-1,492	,359	-,200	-4,151	,000	-2,199	-,786	,956	1,045
	Proportion	,489	,141	,167	3,465	,001	,211	,766	,956	1,045
3	(Constant)	4,736	,220		21,549	,000	4,304	5,168		
	Simplicity	-2,059	,428	-,276	-4,813	,000	-2,899	-1,218	,668	1,497
	Proportion	,416	,143	,142	2,901	,004	,134	,698	,914	1,094
	Equilibrium	-,094	,039	-,135	-2,410	,016	-,171	-,017	,695	1,438
4	(Constant)	6,623	,628		10,549	,000	5,389	7,858		
	Simplicity	-4,764	,944	-,638	-5,046	,000	-6,620	-2,908	,134	7,459
	Proportion	,427	,142	,146	3,008	,003	,148	,706	,914	1,094
	Equilibrium	-,146	,042	-,209	-3,475	,001	-,228	-,063	,594	1,683
	Density	-1,595	,498	-,363	-3,205	,001	-2,573	-,617	,167	5,973
5	(Constant)	6,464	,629		10,270	,000	5,227	7,701		
	Simplicity	-5,082	,951	-,681	-5,341	,000	-6,952	-3,211	,131	7,641
	Proportion	,462	,142	,158	3,249	,001	,182	,742	,902	1,109
	Equilibrium	-,181	,045	-,259	-4,036	,000	-,269	-,093	,517	1,936
	Density	-2,081	,544	-,473	-3,825	,000	-3,151	-1,012	,139	7,201
	Balance	,751	,347	,139	2,163	,031	,069	1,434	,518	1,930

a. Dependent Variable: score

The coefficient table present the significance of our variables. In the fifth model, we observe that they are all significant, which implies that these five variables play a role in predicting the aesthetics score. Besides, we observe that the VIF is lower than 10 which implies that there is no multicollinearity between these five variables.

Excluded Variables[a]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics		
						Tolerance	VIF	Minimum Tolerance
1	Balance	-,021[b]	-,358	,721	-,018	,665	1,503	,665
	Equilibrium	-,170[b]	-3,061	,002	-,149	,727	1,375	,727
	Density	-,173[b]	-1,613	,107	-,079	,198	5,051	,198
	Concentricity	-,104[b]	-2,196	,029	-,107	,998	1,002	,998
	Proportion	,167[b]	3,465	,001	,168	,956	1,045	,956
	Symmetry	,118[b]	2,481	,014	,121	,992	1,008	,992
2	Balance	,007[c]	,116	,907	,006	,653	1,532	,624
	Equilibrium	-,135[c]	-2,410	,016	-,118	,695	1,438	,668
	Density	-,213[c]	-2,008	,045	-,098	,196	5,106	,196
	Concentricity	-,075[c]	-1,559	,120	-,077	,959	1,043	,919
	Symmetry	,098[c]	2,073	,039	,101	,976	1,025	,940
3	Balance	,037[d]	,630	,529	,031	,625	1,601	,531
	Density	-,363[d]	-3,205	,001	-,156	,167	5,973	,134
	Concentricity	-,032[d]	-,610	,542	-,030	,790	1,266	,573
	Symmetry	,087[d]	1,834	,067	,090	,964	1,037	,667
4	Balance	,139[e]	2,163	,031	,106	,518	1,930	,131
	Concentricity	-,019[e]	-,367	,713	-,018	,785	1,274	,131
	Symmetry	,075[e]	1,588	,113	,078	,957	1,044	,134
5	Concentricity	-,016[f]	-,308	,758	-,015	,785	1,275	,128
	Symmetry	,079[f]	1,680	,094	,083	,956	1,046	,131

a. Dependent Variable: score

b. Predictors in the Model: (Constant), Simplicity

c. Predictors in the Model: (Constant), Simplicity, Proportion

d. Predictors in the Model: (Constant), Simplicity, Proportion, Equilibrium

e. Predictors in the Model: (Constant), Simplicity, Proportion, Equilibrium, Density

f. Predictors in the Model: (Constant), Simplicity, Proportion, Equilibrium, Density, Balance

This output presents the variables that have been excluded when running each model.

Appendix 5 – Descriptive analysis (hypothesis 5)

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Balance	10	,712409049	,987018754	,877176733	,095720663
Equilibrium	10	,955120070	,997027797	,982614239	,013563113
Density	10	,742053986	1,28347647	1,02889036	,177080795
Concentricity	10	,415350892	,603222640	,516751538	,062368346
Simplicity	10	−,07923824	,298061889	,107221487	,101882998
Proportion	10	1	1	1,00	,000
Symmetry	10	,418864232	,501970516	,458404619	,025757033
Colorfulness	10	,049992043	,650966749	,362735052	,200783175
compressionComplexity	10	,212634783	,314269893	,264440317	,042000631
Valid N (listwise)	10				

Place des Doyens, 1 bte L2.01.01, 1348 Louvain-la-Neuve, Belgique **www.uclouvain.be/lsm**