NTNU
Norwegian University of
Science and Technology

Aesthetics and Usability in Cross-Cultural Web Sites

Maria Aune Remøy

Master of Science in Computer Science
Submission date: June 2016
Supervisor: Konstantinos Chorianopoulos, IDI

Norwegian University of Science and Technology
Department of Computer and Information Science

Abstract

Although the use of websites keeps increasing all over the world, many websites are not adapted to a global audience. For some time, the issue of how to localise websites have been researched. Moreover, the idea that culture can affect what users see as user-friendly and aesthetically beautiful is now considered to be true. This thesis challenges these ideas by doing a survey inspired by Tractinsky (1997). This study is a replication study of an experiment by Kurosu and Kashimura (1995) but in a different cultural setting. The objective was to investigate the relationships between users' perceptions of interface aesthetics and usability. Moreover, Tractinsky (1997) also examined the differences between his results and the results from Kurosu and Kashimura (1995) seen from a cultural perspective.

The research questions for this thesis are about whether users' perception of beauty affects their perception of usability, and whether they perceive websites designed to fit their culture as more user-friendly or beautiful than other designs. A questionnaire consisting of six different website designs that were evaluated based on their beauty and perceived usability by people from several different countries was created to investigate these questions. For each design, an image was shown to the participant followed by 13 statements. The participant then rated the statements on a 7-point Likert scale from 1 (strongly disagree) to 7 (strongly agree).

For the first research question, the results of the study show that there is a link between the aesthetics of a design and the usability of the design, but this relationship is not strong enough to only rely on the aesthetics of the design when designing a user-friendly system. Moreover, it is not possible to use the results from this study to point out specific components in the design that influence the perceived usability in a specific way.

For the second and third research questions, the results showed that a user's cultural background does not seem to affect which design it prefers, both regarding the aesthetics and the usability. Also, the results indicate that which designs the participants liked varied both just as much inside each country as between countries.

Keywords: localization, website design, aesthetics, Hofstede's dimensions, usability.

Sammendrag

Selv om bruken av nettsider fortsetter å øke over hele verden, er mange nettsider enda ikke tilpasset et globalt publikum. Hvordan man best lokaliserer nettsider er et spørsmål som har blitt forsket på en stund. Det er fastslått at hvilken kultur en bruker tilhører kan påvirke hva denne brukeren ser på som brukervennlig og estetisk vakkert. I denne oppgaven har jeg utfordret disse ideene ved å gjennomføre en studie inspirert av Tractinsky (1997). Studien av Tractinsky (1997) er en replikasjonsstudie av en studie av Kurosu and Kashimura (1995), men i en annen kulturell setting. Målet med studien var å undersøke sammenhengen mellom brukernes oppfatninger av estetikk og brukervennlighet i design. I tillegg undersøkte Tractinsky (1997) forskjellene mellom hans resultater og resultatene fra Kurosu and Kashimura (1995) sett fra et kulturelt perspektiv.

Problemstillingene for denne masteroppgaven handler om hvorvidt brukerens oppfatning av skjønnhet påvirker deres oppfatning av brukervennlighet og om brukere oppfatter en nettside designet for å passe deres kultur som mer brukervennlig eller vakker enn andre nettsider. For å undersøke disse problemstillingene har jeg gjennomført en spørreundersøkelse som består av seks ulike nettsidedesign. Disse designene ble evaluert basert på deres estetikk og brukervennlighet, av folk fra flere forskjellige land. For hver nettside ble et bilde vist til deltakeren etterfulgt av 13 utsagn om nettsiden. Deltakerne skulle deretter vurdere utsagnene på en 7-punkts Likert-skala fra 1 (helt uenig) til 7 (svært enig).

For den første problemstillingen viser resultatene fra studien at det er en sammenheng mellom estetikken og brukervennligheten i et design, men dette forholdet er ikke sterkt nok til at man kun kan basere seg på estetikken i designet for å lage et brukervennlig system. Dessuten er det ikke mulig å bruke resultatene fra denne studien til å påpeke spesifikke komponenter i designet som påvirker brukervennlighet på en bestemt måte da dette varierte fra design til design.

For den andre og tredje problemstillingen viser resultatene at en brukers kulturelle bakgrunn ikke ser ut til å påvirke hvilket design den foretrekker, både når det gjelder estetikk og brukervennlighet. Imidlertid indikerte resultatene at hvilke design deltakerne likte varierte like mye inne i hver kulturelle gruppe som mellom gruppene.

Nøkkelord: lokalisering, nettsidedesign, estetikk, Hofstedes dimensjoner, brukervennlighet.

Preface

You are now about to read the thesis "Aesthetics and Usability in Cross-Cultural Web Sites". It is about whether aesthetics and the perceived usability in designs are linked, and whether culture affects what people think of the aesthetics or the usability in designs. This thesis is my master's thesis in Computer Science at NTNU. I was engaged in researching and writing this thesis from January to June 2016.

The study in this thesis is an extension of my specialisation project at NTNU, which I worked with during the fall of 2015. The research questions were formulated together with my supervisor, Konstantinos Chorianopoulos. The research has been difficult as I have challenged my self in many different ways, both as a designer, with making a questionnaire, and with the statistical analysis of the collected data.

Throughout this research, I have always been able to get help from my supervisor, and also from Patrick Mikalef, who has helped me a lot with the questionnaire and statistical analysis. This support has made the project fun to work with, and I have had many positive experiences throughout this period. Also, I managed to collect enough information to be able to discuss and give conclusions to my research questions.

I would like to thank my supervisor for his guidance and support during my master's thesis and the specialisation project. You have challenged my thinking a lot, and I am grateful for that. Also, I would like to thank Patrick for his help during this master' thesis. Finally, I would like to thank the respondents of my design evaluations and my questionnaire, without you I would not have been able to finish this thesis.

I hope you enjoy your reading.

Maria Aune Remøy.

Trondheim, June 8, 2016.

Contents

List of Figures

List of Tables

Acronyms

IC Individualism vs Collectivism.

LTO Long-Term vs Short-Term Orientation.

MF Masculinity vs Femininity.

PD Power Distance.

UA Uncertainty Avoidance.

Introduction and Related Work

1 Introduction

1.1 Motivation

As of July 1, 2015, there were 3.2 billion internet users in the world. Of this 48.4% are Asian, 21.8% are Americans, 19% are Europeans, and 9.8% are African (Internet Live Stats 2015), see figure 1.1. Because of this growth of internet users, the number of websites is also increasing. Reinecke and Bernstein (2011) points out that this increase makes searching for information more like a search for the best information presentation and that you are likely to use the website that you think has the best design. In other words, out of several websites that have approximately the same information; the users are more likely to choose to use the ones that they think are the most beautiful or interesting.

Because users tend to choose websites based on the aesthetics, it is easy to think that the aesthetics of a website also influence the perceived usability of that site. This idea is interesting because it means that the aesthetics do not only affect whether users like looking at the website, but it can influence the entire experience of using the site. There have been several studies investigating the link between aesthetics and usability in designs. Several of them have shown that aesthetics influence usability (Hartmann et al. 2008, Lindgaard et al. 2006, Tractinsky et al. 2006, Lavie and Tractinsky 2004, Tractinsky et al. 2000, Tractinsky 1997, Kurosu and Kashimura 1995). However, some studies do not find any evidence that aesthetics affect the perceived usability (Tuch et al. 2012, Hassenzahl 2004, van Schaik and Ling 2009). These differences in results make further investigation interesting because the inability to find links between aesthetics and usability does not necessarily mean that the links do not exist. Moreover, even if they do not exist, more experiments are needed to come to that conclusion.

Experiments have been done where the results show that whether users like a website depends on their cultural background (Badre 2001, Reinecke and Bernstein 2011, Simon 2001, Siu-Tsen Shen 2006, Tractinsky 1997, Tsikriktsis 2002). These experiments showed that users perform better at systems designed for them and that they enjoy using these designs more. Because of this, it is interesting to investigate, not only a possible link between aesthetics and usability in design but also whether culture affects a user's view of what is beautiful or easy to use.

The inspiration for this thesis is the experiment in "Aesthetics and Apparent Usability: Empirically Assessing Cultural and Methodological Issues" by Tractinsky (1997). This study is a replication study of "Apparent usability vs. inherent usability: experimental analysis on the determinants of the apparent usability" by Kurosu and Kashimura (1995). These two experiments investigated the link between aesthetics and usability by having participants rate images of different ATMs. Because the two experiments were using participants from various countries Tractinsky (1997) also discussed how the participants cultural background could affect their opinion about the usability and the beauty.

This study follows the same principles as the studies of Kurosu and Kashimura (1995) and Tractinsky (1997), by having different designs that are created to be more or less beautiful and user-friendly and asking the participants to rate the beauty and user-friendliness based on an image of the design. This study takes it a step further, however, by involving

Hofstede's dimensions from the beginning, and creating each design to fit one particular culture. Because of this, it is not a replicate study, but it is very similar.

It is more likely for a research claim to be false than true (Ioannidis 2005). Therefore, a lot of the research performed should be getting negative results. However, a lot of the published research is showing positive results. One reason for this might be that everyone involved in research, from the researchers to the scientific journal, are being evaluated by their number of citations and how many papers they publish (Fanelli 2011), and negative results are not as interesting as positive results. Therefore, the focus is sometimes more on getting a new important finding than to do replicate studies, and increasingly more research questions are seen as true even though it only has a few experiments to back it up (Ioannidis 2005). Because of this, it is important to do several identical or similar studies to try to figure out if the claim is, in fact, correct, or if it was simply a fortunate result. The study in this thesis combines positive findings done by other researchers, and the study design of Kurosu and Kashimura (1995) and Tractinsky (1997) to investigate the proposed questions even further.

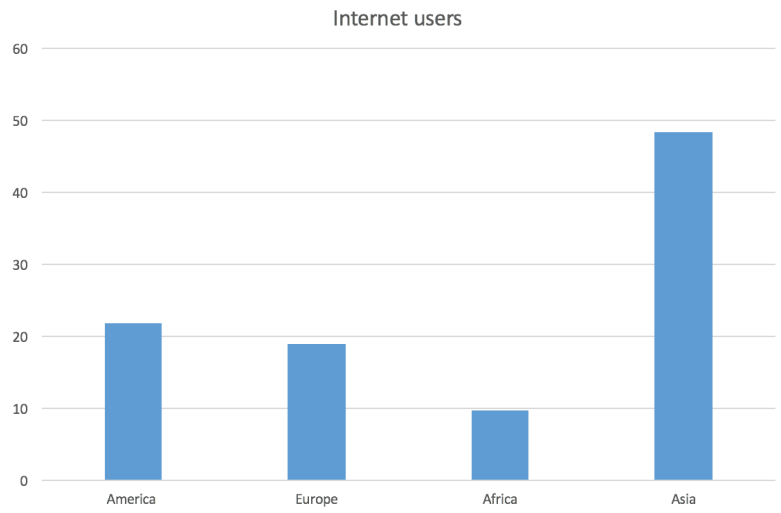


Figure 1.1: Internet users in the world (Internet Live Stats 2015)

1.2 Research Questions

The theme of this thesis is to investigate if there is any link between perceived usability and aesthetics on web shops based on first impressions, and whether users from different cultures have different opinions about which designs that are most beautiful or easy to use. Three research questions have been made to investigate these issues, and the research model in Figure 1.2 shows the links between the different research questions.

1.2.1 Aesthetics and Usability

The first research question is about whether a user's perception of beauty and aesthetics is related to their perception of usability regardless of their culture. It is agreed upon that any system, including web shops, should be easy to use. Some researchers (Hartmann et al. 2008, Lindgaard et al. 2006, Tractinsky et al. 2000, Tractinsky 1997) claims that users are more positively inclined towards systems that they perceive as beautiful and that this positive attitude can affect their opinion of the entire system, including the usability. If this is true, it means that aesthetics is an important part of the usability. Not only to please the eye of the user but also to decide how easy the users think the functionality of the system is to use. Therefore, the first research question is:

> **RQ 1:** Does a user's perception of aesthetics affect their perception of usability?

1.2.2 Cultural Differences

The second and third research questions are about whether members of different cultures have different views on what is beautiful and easy to use. These issues will be analysed by looking at how people perceive aesthetics and usability in various designs, and whether there are any differences between the opinions of people from different cultures and similarities between the opinions of people from similar cultures.

Because usability of systems is important for the user experience, it is natural to think about this when designing user interfaces. However, is usability the same everywhere or are there cultural differences in what designs users perceive as easy to use? This issue is what the second research question aim to answer.

> **RQ 2:** Does a user's cultural background affect which design they perceive as user friendly?

The third and last research question in this thesis also investigate cultural differences in how the users perceive different designs, but the focus is on how beautiful the designs are rather than how usable they are.

> **RQ 3:** Does a user's cultural background affect which design they consider to be aesthetically beautiful?

1.3 Research Strategy

For this master's thesis, the goal is to test whether the aesthetics of design has an impact on a user's perceived usability of that design and whether users' cultural backgrounds affects their perception on different designs. Figure 1.3 shows the research process I used to perform this study. These steps will be further explained later in this section, but first summarised here to give an impression of the process.

A literature review was performed to be able to develop the research questions and the conceptual framework for this thesis. As can be seen in the figure, personal experiences and

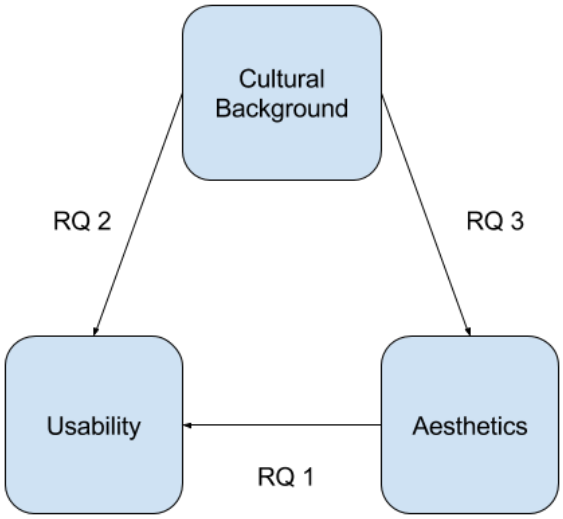


Figure 1.2: Research model

motivation for doing this thesis were also a part of the creation of the research questions. Then a survey was performed the research strategy, with a questionnaire as the data generation method. Finally, a quantitative data analysis was conducted to discuss and conclude on the research questions.

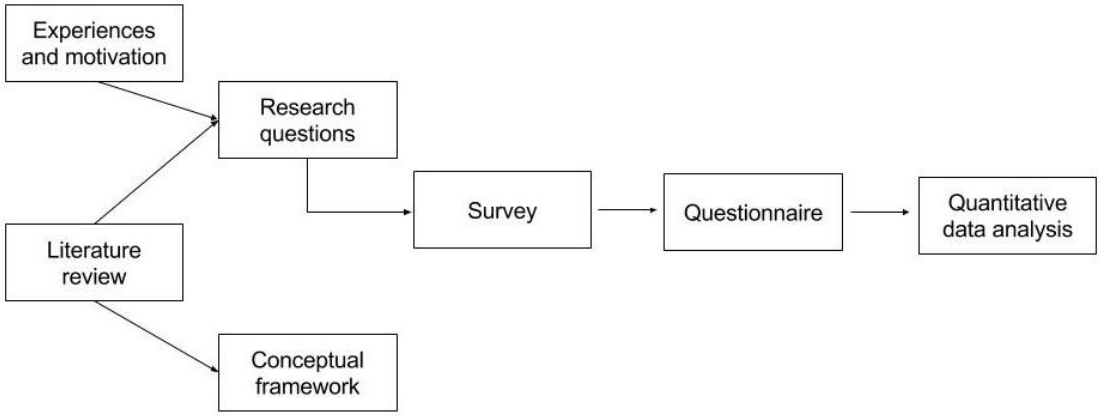


Figure 1.3: Model of the research process

1.3.1 Literature Review

During the specialisation project on NTNU fall 2015, a literature review on how culture can affect the usability and trust in websites was carried out. The purpose of the literature review was first to learn more about, and get a feel for the research and literature on the topic and choose research questions. Then the literature was analysed to be used as a basis for discussing the research questions. Selected parts of this review have been altered and reused in this master's thesis, mostly in Chapter 2 Related Work. The chapter *Reviewing*

the Literature in Oates (2006) was the inspiration for the method used for this literature review.

In the first part of the literature review, the goal was to identify and read as much literature as needed to get a feel of the research field and what other researchers has found out before. To be able to find good articles for the review a systematic literature search was conducted, and all relevant literature was evaluated on a set of criteria.

In the systematic literature search, the research problem was divided into concepts, see Table 1.1. Then concepts from different categories were paired and used as keywords in the searches. This technique made finding relevant and contrasting literature easier.

Concept 1	**Concept 2**	**Concept 3**	**Concept 4**	**Concept 5**
Design	GUI	Multicultural	Audience	Usability
Plan	Front-end	Universal	Users	Trust
Outline	Web site	Intercontinental	Customers	

Table 1.1: Concepts for literature search

To decide whether an article should be read and analysed it was, as mentioned earlier, evaluated on a few criteria. The criteria were how many citations the article had, how many other articles the author(s) have written before and how successful they were, and whether or not the journal publishing the article is considered to be of high quality. If an article met at least one of these criteria, it was found to be relevant for the specialisation project or master's thesis.

Then, as the last way of finding literature, two other methods were also used. The first method was to look for more relevant literature written by the same researchers as one or several of the articles from the first search. The second method was to look for relevant literature in the references of the articles already included in the review.

After this extensive search and reading part, some research questions had been created. After making these questions, the literature was evaluated and recorded to be able to write the review. The recording was meant to be an aid to the memory during the final stage, which is the writing of the review, by writing down summaries of the articles. Then, in the end, the literature was used as a basis to write a review on how culture can affect the usability and trust in websites.

1.3.2 Survey

The research method for this thesis is a survey. A survey obtains the same kinds of data from a large group of people or events, in a standardised and systematic way (Oates 2006). This strategy makes it possible to look for patterns in the data to generalise the findings to a larger population than the group that participated in the survey.

Figure 1.4 shows the process of creating the survey. Before starting on the questionnaire, a requirements elicitation process was performed together with the literature review. These were done at the same time because the requirements were collected from literature about

previous research. Then there was an iterative process of designing and evaluating the web shops, designing the questions for the questionnaire and doing pilot tests on the current version of the questionnaire. This process was performed twice before the results were satisfiable. Then final changes were made, and the survey was shared. The reason why the pilot test and the evaluation were finished before redesigning the web shops, was to get feedback on the designs both from the people evaluating them and of the participants of the pilot test before changing anything.

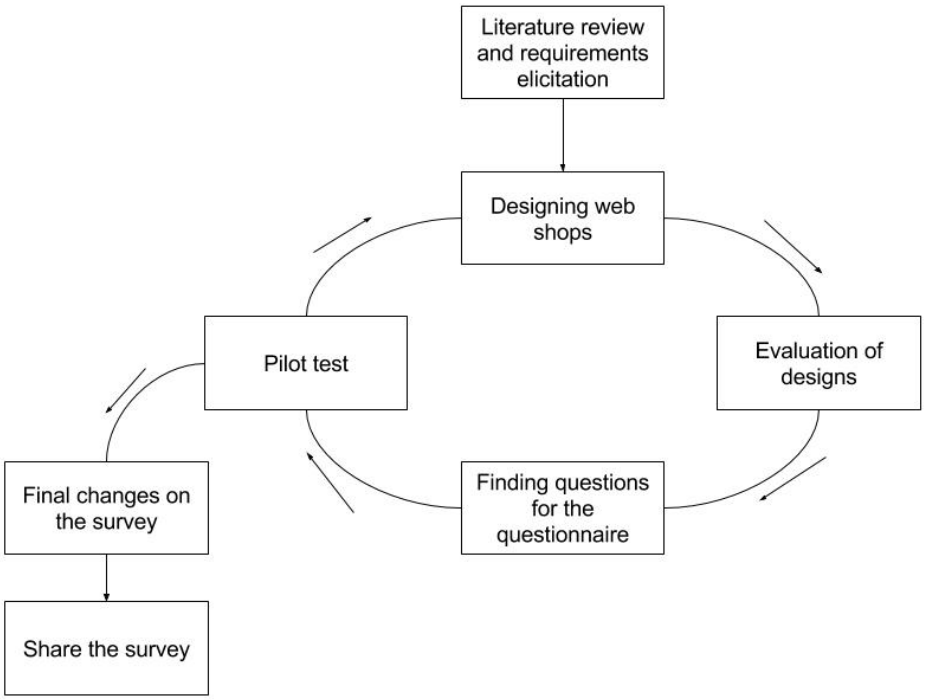


Figure 1.4: Model of the survey creation process

Data Generation Method

The data generation method in this survey is a questionnaire with closed questions, which means that the respondent was forced to pick an answer from a range of available choices on every question. The questions generated both factual data, which is the respondent's gender, age and nationality, and opinions. The opinions were measured using a 7-point Likert scale.

The questionnaire was self-administered, which means that the researcher, was not present when the participants took the questionnaire. This method saves time and makes it easier for the researcher to get more respondents to take the questionnaire. Also, when the researcher is not present, the participants are less likely to try to give the answer they think *correct*, but rather state their opinion (Oates 2006).

Sampling Technique

The sampling technique used in this survey is the self-selection sampling (Oates 2006), which means that the information about the questionnaire was shared to people and that everyone that wanted to participate could do so. The questionnaire was shared through Facebook, more precisely through groups for international students, and the participants were everyone that took the questionnaire.

Research Paradigm

This approach to doing research is underlying the positivism paradigm. This means that this thesis assumes that the world is ordered and regular and that it is possible to investigate it objectively (Oates 2006). This thesis, therefore, expects that it is possible to create hypotheses that is either right or wrong and then test these hypotheses, and also to do this objectively.

The aim of this survey is to investigate whether there is a link between the aesthetics and the perceived usability of a web shop, and whether a user's cultural background affects its views on aesthetics and usability in different web shops. Because the web shops designed are based on guidelines collected from other researchers' experiments and analyses, the researchers personal beliefs on the subject do not influence the thesis, and it is, therefore, objective. Also, as mentioned before, the data generation method is a questionnaire that the participants will answer without interacting with the researcher. There is, therefore, no way that the researchers opinions and thoughts can influence the participants of the questionnaire. Moreover, a person's preferences or cultural background is not something that changes fast, if it changes at all. Therefore, the environment that is tested, and the theory behind it should be regular over time and also reliable.

The survey is strongly influenced by earlier research, by combining the findings of other researchers and testing them in a new environment. This way of doing research is within the positivism paradigm called repeatability (Oates 2006). Repeatability is an important concept because a hypothesis is not true only because one experiment ended on a positive result. Moreover, it is impossible to prove that a hypothesis is true, one can only try to strengthen the hypothesis by performing more experiments and test it as well as possible. Also, by doing this, the hypothesis might end up being disproved. Therefore, it is important to do replicate studies or similar experiments to disprove or strengthen a hypothesis.

Validity of the Survey

Four criteria will be discussed in this section to assess the validity of the survey in this thesis. These criteria are objectivity, reliability, internal validity and external validity (Oates 2006).

The first criterion when evaluating the validity of positivist research is the objectivity. This criterion is about whether the data collected is in fact objective, or if someone, like the researcher, or something has affected the participants when answering the questions (Oates 2006). In this study, the researcher has no idea whom the participants are and have had no contact with any of them during the period of the research. Also, there were no indications on the designs as to which country it was intended. Therefore, the participants could not

know which design that was created to fit their culture. Moreover, the researcher has no reason to be favouring any particular result. This thesis is not written for a company or organisation that would gain from a particular result, but solely because it is an interesting topic for the researcher to investigate. Therefore, it is no reason to believe that the study is not objective.

The second criterion is about the reliability of the survey. This criterion involves whether the instruments used in the research is neutral, accurate and reliable, and if it is possible to replicate the research and get the same results (Oates 2006). In this thesis, the research instrument is the questions in the questionnaire. Both the reliability of the questions and whether the questions are unambiguous and easy to understand will be discussed in Chapter 3 Description of the Survey, and will therefore not be discussed any further here.

The third criterion for assessing the validity of the survey is the internal validity. The internal validity is about whether the correct things were investigated, and that the right data was collected (Oates 2006). Evaluations were performed on the designs before the survey was shared with the public to assess whether the designs meet the requirements. Also, pilot tests were conducted on the questionnaire to get feedback on the survey and to use the data for a trial statistical analysis. Both the evaluation and the pilot tests are discussed in detail in Chapter 3 Description of the Survey, and will therefore not be discussed here.

The fourth criterion is the external validity, which is about whether or not the results of this study is generalizable (Oates 2006). This particular experiment has only been performed once. Therefore, it is difficult to say whether the results are general for the entire population, or if they are more specific to a group of the population or a specific situation. The positive aspects of this research when it comes to generalisation is that there are enough participants in the study to test the research questions, and the participants are from all over the world. Because this thesis is trying to connect design preferences with Hofstede's dimensions, it is good that the participants are from a large sample of countries. However, as can be seen in Section 3.6 Participants, most of the participants in this study are students or young adults. Because younger people is more used to using computers and other electrical devices, they might also be more used to foreign or international user interfaces. Therefore, this group of people might not be representative of the entire population that this research is interested in, which is everyone using websites. Older people might like other types of designs and use the Internet differently than the participants in this survey.

1.3.3 Quantitative Data Analysis

The types of quantitative data obtained in the questionnaire are nominal data for the gender and nationality questions, and ordinal data for the Likert scale questions and the age question. Because the aesthetics and usability of the system are hard to measure with only one variable each, several variables are used to try to capture the participants' actual opinion on each of these two concepts. In this way, both negative and positive views of the aesthetics and usability are captured, instead of one general impression about the entire usability or aesthetics of the design. Because several variables are used to measure each of these concepts, and all the data is important in the measurement, it makes sense to use all the data to compute a general aesthetics or usability score. Therefore, the mean values are

used to calculate the usability or aesthetics scores, not the median or mode values, even though the median or mode is usually considered a better fit for ordinal variables (Oates 2006). Also, to be sure that the set of variables used to evaluate these components measure the same underlying concept, Cronbach's alpha have been used, as explained next.

Cronbach's Alpha

Cronbach's alpha is used to assess the reliability of the questionnaire. This alpha is a measure of the internal consistency of a set of items, which in this case is a set of questions. It can be used to determine to what degree these questions are measuring the same underlying dimension (Leard Statistics 2015a). In this research, the alpha is used to test that the different questions regarding the aesthetics and the usability in the questionnaire, is, in fact, measuring the aesthetics and the usability of the design. Table 1.2 shows how the internal consistency is evaluated based on the value of the alpha. The alpha can have values between zero and one, and higher values indicate a more reliable questionnaire.

Cronbachs alpha	**Internal Consistency**
$\alpha \geq 0.9$	Excellent
$0.9 > \alpha \geq 0.8$	Good
$0.8 > \alpha \geq 0.7$	Acceptable
$0.7 > \alpha \geq 0.6$	Questionable
$0.6 > \alpha \geq 0.5$	Poor
$0.5 > \alpha$	Unacceptable

Table 1.2: Values for Cronbach's alpha (Wikipedia contributors 2016)

Simple Linear Regression

A simple linear regression can be used to evaluate the linear relationship between two continuous variables. In other words, it can be used to determine whether the linear regression between two variables, one dependent and one independent, is statistically significant, and to predict the value of this dependent variable based on the value of the independent variable (Leard Statistics 2015c). In this research, simple linear regression is used to assess the relationship between the mean usability score and the mean aesthetics score to decide whether the perceived usability is related to the aesthetics of the design. Also, effect size, which is an objective and standardised measure of the magnitude of the effect (Field 2013), will be used to help to describe the effect aesthetics has on usability. This effect will be measured using the multiple correlation coefficient R, with values of $\pm.1$ representing a small effect size, $\pm.3$ representing a medium effect size and $\pm.5$ representing a large effect size (Field 2013).

To run simple linear regression on the data, it needs to meet the following assumptions (Leard Statistics 2015c):

1. One dependent variable measured at the continuous level.
2. One independent variable measured at the continuous level.
3. There must be a linear relationship between the dependent and independent variables.
4. The observations should be independent.
5. There should be no significant outliers.
6. The data needs to show homoscedasticity.
7. The residuals (errors) of the regression line need to be approximately normally distributed.

The first two assumptions are possible to check without doing any analysis. The dependent variable in this test will be the mean usability score from all the designs and all the participants, and the independent variable will be the mean aesthetics score. Which design the participants evaluated does not matter because it is the relationship between the aesthetics and the usability that is tested, not what the participants thought about the different designs.

The third assumption was checked by using a scatterplot of the usability score against the aesthetics score. Visual inspection of the scatterplot indicated a linear relationship between the variables. The fourth assumption could be easily checked using the Durbin-Watson statistic, and this will be explained further in Chapter 4 Results. The fifth assumption is that there should be no outliers. This issue was checked during the linear regression procedure and will also be discussed further in Chapter 4 Results. The sixth assumption, that the data show homoscedasticity was checked by using visual inspection of plots of standardised residuals versus standardised predicted values, which showed homoscedasticity. The seventh assumption was checked using visual inspection of both a histogram of the regression standardised residual against the frequency, and a normal probability plot. Both showed that the residuals were approximately normally distributed. Appendix A shows the plots used to decide whether the data has homoscedasticity, which is when the dots in the plot is spread out. Appendix B shows the p-plots and histograms that were used assess whether the residuals in the data was normally distributed.

Relationships Between Aesthetics Ratings and Usability Scores

Bar charts were used to get an image of how the ratings of the different aesthetics components affect the usability score of the design. These charts showed the median usability score for each rating for each aesthetics component. The median was used because it is the middle usability value, and it is therefore not affected by any extreme or unusual values (Oates 2006). The charts will together with the simple linear regression help to give an answer to the first research question. Particularly, the charts will be used to look for trends in how the different components of the aesthetics affect the usability of the design.

Kruskal Wallis H Test

The Kruskal-Wallis H test can be used to determine if there are statistically significant differences between two or more groups of an independent variable on a continuous or ordinal dependent variable (Leard Statistics 2015b). This test will be used to find out if there are any differences between what the participants from different cultures believe about the different designs.

To run the Kruskal-Wallis H Test the data must meet the following assumptions (Leard Statistics 2015b):

1. One dependent variable measured at the continuous or ordinal level.
2. One independent variable that consists of two or more categorical, independent groups.
3. The observations should be independent.
4. Determine whether the distribution of scores for each group of the independent variable have the same or a different shape.

The first three assumptions are possible to check without doing any analysis of the data collected. In this research, the test will be run twice for each design and twice for each cluster, once for the aesthetics score and once for the usability score. Because the dependent variable is either the mean aesthetics or usability score, the first assumption is met. The independent variable is the participant's cultural group. In this research, six cultural groups, called clusters, are used, all of which are categorical and independent of each other, which means that the second assumption is also met. The third assumption is about not having any relationship between the data in each cluster, or between the clusters. The assumption is met by having different participants in each group, and not letting any participants be a member of more than one group.

The fourth assumption needs some more analysis, and boxplots will be used to determine whether the assumption is met or not. The Kruskal-Wallis test will be used independently of whether this assumption is fulfilled, but the interpretation of the results depends on this assumption. When the assumption is met the median values can be used to interpret the results, while mean ranks will be used otherwise.

If the results show that there are statistically significant differences between the scores of the independent variable and the dependent variable, a post hoc analysis will be performed. When this is the case, pairwise comparisons will be carried out using a procedure with a Bonferroni correction for multiple comparisons (Dunn 1964).

1.3.4 Tools

Some tools were used to create and share the questionnaire. Django and Bootstrap were used to make the web shop designs, and a website with information about the research and a link to the questionnaire, www.mariaremoy.no. SurveyMonkey was used to make the questionnaire and the evaluation because it is a good and easy tool for creating and editing questionnaires and it is easy to download all the answers in appropriate files. Moreover, it is possible not to register IP addresses on the participants, which makes it possible to create a anonymous questionnaire. SPSS Statistics was used to perform the tests in the statistical analysis.

1.4 Scope and Limitations

The goal of this thesis is to discuss and give an answer to the questions of whether users' perception of beauty affects their perception of usability, and whether a user thinks the web

shop designed to fit their culture is more beautiful or user-friendly than the other designs. Some of the work used in this master thesis was carried out in a preliminary study by the researcher (Remøy 2015). Parts of this is in the Motivation section of this chapter and chapter 2 Related Work.

The research in this masters' thesis is all performed by me, which means that there are some limitations on this work. Firstly, there is the issue of time. I only had 20 weeks to carry out the research and report it, which means that the amount of designs used in the study, the quality of the designs and the number of participants I was able to collect throughout this period was limited. The second limitation is that I designed all the web shops myself. I have very limited design experience, and with a set time frame it is limited how many iterations I could have on the designs. Employees at NTNU evaluated all the designs, but the number of evaluations I was able to get within the time that I had, was also limited. Thirdly, the designs in the questionnaire are all in English. Normally, when websites are localised, they are at least translated. Because the designs in this survey are not translated the participants that are used to websites in another language than English can feel like all the sites were foreign and therefore not made for them. However, the focus of the research is on the layout, colours and images of the design, and not the language. Therefore, I decided that because I only know two languages, Norwegian and English, I would make all the designs in English like the rest of the questionnaire.

1.5 Outline of the Thesis

The thesis is split into three parts.

Part 1: The first part include an introduction to the report, the research methodology, and theory and related work relevant for the thesis.

Part 2: The second part is the empirical study and includes the design of the study and the results of the study.

Part 3: The third and last part include a discussion of the results and a conclusion of the research questions, in addition to suggestions for future research.

2 | Related Work

2.1 Aesthetics and Usability

Because users tend to avoid using websites that are hard to use, one of the most important aspects of websites, in general, is whether it is perceived as easy to use. Therefore, it is important to design a usable website, particularly if the users can choose whether or not to use it. The definition of usability in this report is that usability refers to"the capability of the software product to be understood, learnt, used, and attractive to the user when used under specified conditions" (Bevan 2001). These conditions are in this case when the users want to purchase or browse for products online.

Another important factor for having a website that users choose to use is the aesthetics. In this thesis the aesthetics of a design means the different visible components of the design, like images, colours and icons. As mentioned in section 1.1 Motivation, because so many sites exist, the aesthetics has become one of the most important factors when users determine whether or not they want to use a particular site. Therefore, if the target audience finds the website visually appealing, it has a better chance at appealing to the users. Moreover, if the website is considered aesthetically pleasing and easy to use, then the users are more likely to use this site again.

As mentioned in section 1.1 Motivation, some studies have been able to show a relationship between the aesthetics of a design and the perceived usability of that design. One such study was performed by Kurosu and Kashimura (1995). In this study, the participants should evaluate the beauty and usability of an ATM based on images of the interface. Similar experiments were later performed by Tractinsky (1997), and Tractinsky et al. (2000) but with participants from a different country. The results of these experiments showed that how beautiful a user thought the interface was significantly influenced that participant's perceived ease of use of the entire system. Tractinsky (1997) also states that "first impressions often influence attitude formation to a large extent." In other words, what the users think about the website after looking at it for the first time, can affect how they experience it later. Subsequently, several experiments have indicated this same link between aesthetics and perceived usability of a system (Lindgaard et al. 2006, Hartmann et al. 2008, Lavie and Tractinsky 2004, Tractinsky et al. 2006). Moreover, as discussed in section 1.1 Motivation, the study in this thesis is also inspired by Kurosu and Kashimura (1995) and Tractinsky (1997). However, some experiments have not found this link (Hassenzahl 2004, van Schaik and Ling 2009, Tuch et al. 2012).

One study that failed to show such a link was a study performed by Tuch et al. (2012). In this study, 80 participants used an online shop to find products and then rate the shop on the aesthetics and usability before and after using the system. The online shop had four different versions that differed in the aesthetics and the usability, and each participant used one of these four systems. The results of this experiment were that the aesthetics did not affect the perceived usability, but that the shops with low usability could significantly lower a participant's opinion about the aesthetics.

van Schaik and Ling (2009) performed two experiments to investigate how users form aesthetic impressions and other judgements about websites. The study was carried out

by letting users evaluate websites both with and without a context in which to make their observations. There were four goals of these experiments. The first goal was to investigate how the aesthetic perception of a web page could differ from when presented with a context of use as opposed to without context. The second goal was to assess whether attractive pages are rated higher than unattractive pages in the right context. The third goal was to investigate whether classically aesthetic pages are rated as more attractive than expressively aesthetic pages. The fourth and last, objective of these experiments were to examine a statement made by Tractinsky et al. (2000), which is that 'what is beautiful is usable'.

Through these experiments, they found that context, aesthetic design and experience of using a product were significant for the user's perception of a website. However, the principle of that 'what is beautiful is usable' was not confirmed as more factors than just the aesthetics were influencing the usability of the website.

2.2 Cultural Differences

Another interesting aspect of designing websites is that the users of the sites can be from all over the world. Because of this, it is quite common for websites to offer a set of translated versions to help the users understand the website better and to increase the usability of the site for international users. However, to only translate a website without doing anything with the rest of the design is often considered as insufficient when designing for a cross-cultural audience. Different cultures often use websites for different purposes and in various ways. Therefore, people from different cultures might have differences of opinion on what is user-friendly or beautiful. Two experiments that show this was conducted by O'Keefe et al. (2000) and Chau et al. (2002). In these experiments, participants from Hong Kong and the USA were asked about their Internet habits. The results show that Hong Kong subjects used websites as a social communication device while the American subjects used them more for product information search purposes. Therefore, to design cross-cultural user-friendly websites, it is important to try to understand how the system will be used and what the users perceive as user-friendly.

Another reason designing websites for a cross-cultural audience can be difficult is because users from different countries may have different opinions about what is beautiful or good design. In earlier research, many websites have been analysed to find differences that may be caused by culture. Moreover, many researchers have designed and tested systems based on those kinds of analyses to determine if the culture does, in fact, create differences in preferences. Below follow two different examples of this.

Callahan (2006) conducted an analysis on University websites from Malaysia, Austria, the United States, Ecuador, Japan, Sweden, Greece, and Denmark. The goal of the experiment was to define graphical elements on 20 websites per country and investigate whether these elements could be measured using Hofstede's dimensions. The websites were analysed in their native language version, and each site was examined by two criteria: organisation and graphical design. The result of this analysis was that the design of the websites varied across cultures, and that, even though the correlations were weaker than expected, it is possible to relate the differences to Hofstede's cultural model.

In “Improving Performance, Perceived Usability, and Aesthetics with Culturally Adaptive User Interfaces” Reinecke and Bernstein (2011) reports that “it is unreasonable to design one common Website for everyone, and yet expect to attract an international audience”. Based on this claim, they made a system that automatically adapts itself to fit the user’s cultural background based on data about the user, and tested this system in an experiment. Their results showed that when the users were using the adapted version of the system they performed better and also thought that the system was easier to use. These results show that to create culturally adapted websites, or to ’localise’ the websites is a good idea.

2.2.1 Localization of User Interfaces

Localization is defined by Cyr and Trevor-Smith (2004) as “the process of adapting a product or service to a particular language, culture, and desired local ‘look-and-feel’.” The objective of localization is to make a website that appears to have been developed in the local culture. As mentioned before, many websites have different versions for different languages and even though that is a step towards localization, it is not enough when localising a website. To identify localization elements that could be generalised for different cultures, Barber and Badre (1998) defined the term ‘cultural markers’. In more detail, cultural markers are interface design elements that are common and possibly preferred within a particular cultural group. To identify these cultural markers can help designers and researchers to figure out how to localise designs so that they fit their audience, and whether localization of the designs is, in fact, necessary. In this thesis, cultural markers defined by other researchers in previous studies have been the basis for the guidelines used to design the web shops.

Another important aspect of localization is to define what is meant by different cultures. This thesis uses the definition by Hofstede (2011). He defines culture as “the collective programming of the mind that distinguishes the members of one group or category of people from others.” To be able to separate the different groups from each other and explain the differences between the groups of cultures, Hofstede developed a set of dimensions, described below. Subsection 2.2.3 Cultural Clusters, contains how these dimensions can be used to create different cultural clusters. These clusters will form the basis for the web shops designed for the questionnaire in this thesis. Finally, in subsection 2.2.4 Guidelines for Localization, a set of guidelines for localising designs will be explained. These guidelines are based on Hofstede’s dimensions, and they will be the requirements for the web shop designs in this thesis.

2.2.2 Dimensions of Culture

Hofstede’s dimensions are a way of representing differences and similarities in cultures at a country level. Although six dimensions now exist, in this research only five of them are used as many countries do not yet have a score at the newest dimension. Figure 2.1 gives an overview of these five dimensions.

[1]Retrieved from: https://globalandco.wordpress.com/2014/07/24/hofstedes-5-cultural-dimensions/

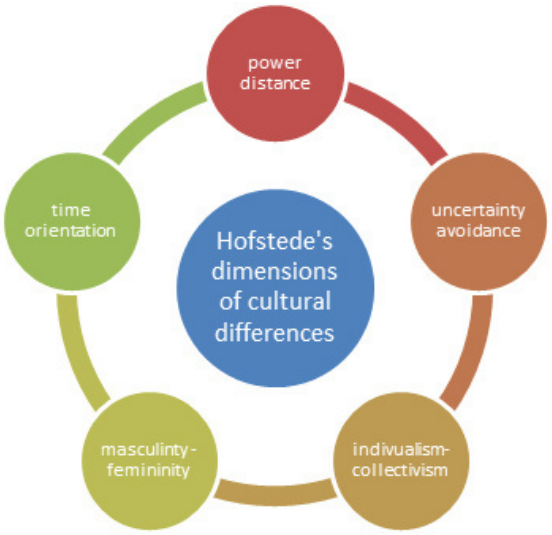


Figure 2.1: Hofstede's dimensions[1]

Power Distance

The first dimension is Power Distance (PD). This dimension is defined by Hofstede (2011) as "the extent to which the less powerful members of organisations and institutions (like the family) accept and expect that power is distributed unequally". Cultures with large PD scores see hierarchy as important, and the roles of individuals are strict both in organisations and in people's everyday life. Moreover, in cultures with small PD scores, the hierarchy is flat, and people have more similar roles than in cultures with high scores.

Uncertainty Avoidance

Uncertainty Avoidance (UA) "indicates to what extent a culture programs its members to feel either uncomfortable or comfortable in unstructured situations" (Hofstede 2011). In other words, cultures with strong uncertainty avoidance try to avoid situations that are new, surprising or different from what they are used to while cultures with weak uncertainty avoidance are more welcoming to these situations.

Individualism vs. Collectivism

Individualism vs Collectivism (IC) "is the degree to which people in a society are integrated into groups" (Hofstede 2011). People in an individualistic culture do not depend on others to take care of them and value alone time and privacy in their life. People from a collectivistic culture, however, are integrated into groups from their birth. These groups are supportive of each other and the members of the groups, like family, are loyal to each other.

Masculinity vs. Femininity

Masculinity vs Femininity (MF) refers to the distribution of values between the genders (Hofstede 2011). The study of Hofstede (2011) showed that "(a) women's values differ less among societies than men's values; (b) men's values from one country to another contain a dimension from very assertive and competitive and maximally different from women's values on the one side, to modest and caring and similar to women's values on the other." In other words, in cultures with a high masculinity score, the differentiation between the genders is much bigger than in feminine cultures.

Long-Term vs. Short-Term Orientation

Long-Term vs Short-Term Orientation (LTO) is a dimension that strongly correlates with recent economic growth, and the values of the two poles are defined by Hofstede (2011) as follows: The long-term pole consists of values like perseverance, thrift, ordering relationships by status and having a sense of shame. At the short-term pole the important values are reciprocating social obligations, respect for tradition, protecting one's 'face', and personal steadiness and stability.

Criticism of Hofstede's Dimensions

Even though Hofstede's framework is the most commonly used cultural framework in business research (Baack and Singh 2007), and it has been used several times to try to combine cultural differences to the design of graphical websites, researchers regularly criticise the framework. The main critic points of the framework are that the data collection took place more than 40 years ago and that the data collecting method was a survey (Baack and Singh 2007, Callahan 2006).

The fact that the data collection for Hofstede's dimensions happened more than 40 years ago has made many researchers claim that the data is too old, and can therefore not be considered valid anymore. As an answer to this, Hofstede argued that cultures not change rapidly, so the data do not become old fast (Callahan 2006).

Also, Hofstede argued that even though surveys are not an optimal way of studying cultures, it is a method commonly used in sociological research, and it should therefore not be ignored (Callahan 2006).

2.2.3 Cultural Clusters

Each of the web shops is designed to a particular group of countries, called a cluster. The countries are divided into these clusters according to their scores in Hofstede's dimensions. This thesis uses the clusters as defined by Wursten and Fadrhonc (2012).

Contest Cluster: The countries in the Contest cluster are typically the Anglo-Saxon countries descended from the British Empire (Robbins and Stylianou 2003, Wursten and Fadrhonc 2012). The countries have high PD and MF dimension scores, and low IC, UA and LTO dimension scores (Wursten and Fadrhonc 2012).

Network Cluster: The countries in the Network cluster typically share the influence of the Hanseatic League, and include countries like the Scandinavian countries and the Netherlands (Robbins and Stylianou 2003). These countries have a high IC dimension score, a low to middle UA dimension score, and low MF, PD and LTO dimension scores.

Well-Oiled Machine Cluster: The countries in the Well-Oiled Machine cluster are countries that speak German (Robbins and Stylianou 2003). The characteristics of the countries in this cluster are their high UA dimension score and their low PD dimension score (Wursten and Fadrhonc 2012). The rest of the scores can vary.

Solar Cluster: The countries in the Solar cluster are countries that typically descended from the Roman Empire (Robbins and Stylianou 2003). The cluster typically has high PD, UA and IC dimension scores. The MF and LTO dimension scores can vary.

Family Cluster: The countries in the Family cluster are countries that typically share the influence of the Chinese majority (Robbins and Stylianou 2003). The countries in the Family cluster are countries that have a high PD dimension score and low IC and UA dimension scores (Wursten and Fadrhonc 2012). The scores for the MF and LTO dimensions can vary.

Pyramid Cluster: The countries in this cluster have high PD and UA dimension scores and a low IC dimension score. The scores for the MF and LTO dimensions can vary.

2.2.4 Guidelines for Localization

This section will present the guidelines that were used to design the web shops for the survey in this thesis. As mentioned before, the guidelines are based on Hofstede's dimensions, which makes it easier to decide how to localise the designs for the different cultural clusters.

PD - Hierarchy:

Marcus and Gould (2000) expects that the interfaces from high power distance cultures have tall hierarchical structures because they are used to a hierarchical structure from other aspects of their lives. On the other hand, the interfaces from the low power distance cultures have shallow hierarchical structures.

IC - Navigation:

In individualistic countries the navigation used is a lot more textual than the navigation in collectivistic countries. This idea is supported by Cyr and Trevor-Smith (2004) and Reinecke and Bernstein (2011). Cyr and Trevor-Smith (2004) found that Japan, a collectivistic country, was twice as likely to prefer symbolic navigation tools than was Germany or the United States, individualistic countries. Moreover, Japan used symbols for links significantly more than Germany and the United States. Reinecke and Bernstein (2011) found that collectivistic countries use more image icons in the header menu of the website.

IC - Colorfulness:

In individualistic cultures, the colours of the UI tend to be homogenous while collectivistic cultures use more colours Reinecke and Bernstein (2011). Moreover, Cyr (2008) found that users from collectivistic countries have a strong preference for visuals, whereas users from more individualistic cultures prefer logical and structured layout.

IC - Images:

In individualistic countries, importance is given to individuals, young people and action while in collectivistic cultures relationships, history, older more experienced leaders and states of being that are thought to be wise are more important (Marcus and Gould 2000). One way of showing this difference is through images. Würtz (2006) found that images on the websites in individualistic cultures portray individuals and their lifestyle while collectivistic cultures show products or merchandise being used by people. Also, in individualistic countries, success is demonstrated through materialism and consumerism, and the motivation is based on personal achievements. On the other hand, in collectivistic cultures success is demonstrated through attainment of social-political agendas and the motivation is on group achievements (Marcus and Gould 2000).

MF - Attention:

In masculine cultures, the attention can be gained through games and competitions, while the attention in feminine cultures should be gained through poetry, visual aesthetics and appeals to unifying values (Marcus and Gould 2000). This statement was supported by Tractinsky (1997) who found that the feminine Israeli participants perceived ease of use and design aesthetics to be more closely related than the masculine Japanese participants.

MF - Saturation:

Reinecke and Bernstein (2011) found that feminine cultures use more pastel colours, and masculine cultures use more highly contrasting and bright colours. This guideline can be explained with that the masculine countries prefer more loud designs, with their attention being drawn in the direction of the colours or specific images. The feminine countries, on the other hand, prefer the aesthetically beautiful design with colours that is in harmony with each other and do not take too much attention from the information at the site.

UA - Information Quantity:

Marcus and Gould (2000)expect the interfaces in cultures with high uncertainty avoidance to be simple, contain a limited amount of choices and information, and to have clear metaphors while cultures with low uncertainty avoidance are complex with a lot of content and choices. Cyr and Trevor-Smith (2004) found that countries with high uncertainty avoidance used more point form and paragraphs in the text than countries with low uncertainty avoidance. Moreover, in the personalised interfaces for cultures with high uncertainty avoidance Reinecke and Bernstein (2011) hid all unnecessary information so that the user could concentrate on the important information. On the other hand, for cultures with low uncertainty avoidance, when users enter a dialogue, all other information

in the interface are still visible and accessible. Moreover, cultures with high uncertainty avoidance want the elements presented to give certainty about the information because they want to avoid taking risks (Cyr 2008). To avoid taking risks is not necessary for cultures with low uncertainty avoidance, where taking a risk is not considered as dangerous as in cultures with high uncertainty avoidance.

LTO - Information Density:

Also, while long-term oriented cultures strive for patience in achieving results and goals, short-term oriented cultures desire for immediate results and achievement of goals (Marcus and Gould 2000). Reinecke and Bernstein (2011) used this by making a complex interface version that showed all information at first sight and color-encoded information with large icons for long-term oriented cultures, and a simpler interface that showed only small units of information at first sight for short-term oriented cultures.

Table 2.1 shows a summary of the guidelines.

	High	**Low**
PD	Tall hierarchical structure of the website	Shallow hierarchical structure of the website
IC	Use textual navigation more	Use symbolic navigation more
	The user interface is homogenously coloured	Uses many different colours and images in the user interface
	Images showing individuals and their lifestyles	Images showing products or merchandise being used by people
MF	Attention gained through games and competitions.	Attention gained through poetry, visual aesthetics, and appeals to unifying values.
	Highly contrasting colours	Pastel colours
UA	Simplicity, with clear metaphors, limited choices, and restricted amounts of data	Complexity with maximal content and choices
LTO	Patience in achieving results and goals.	Desire for immediate results and achievement of goals.

Table 2.1: Guidelines for localization

II Empirical Study

3 | Description of the Survey

This chapter will describe the web shop designs used in the survey, including which countries that are in each cluster, the requirements used when creating the web shops and the final design of the web shops.

3.1 Problem Description

The goal of this survey is to collect data from people all over the world about their perception of different web shop designs. This data will then be used to answer the research questions proposed in section 1.2 Research Questions. Each of the web shops is designed to a particular cluster, as explained in Chapter 2 Related Work. The countries of the participants in this survey are divided into these clusters according to their scores in Hofstede's dimensions. The countries included in each cluster is listed below.

Contest Cluster: Australia, Canada, New Zealand, UK, and the USA.

Network Cluster: Denmark, Norway, and the Netherlands.

Well-Oiled Machine Cluster: Austria, Estonia, Germany, Hungary, Latvia, Lithuania, Luxembourg, and Switzerland.

Solar Cluster: Belgium, Czech Republic, France, Italy, Poland, Slovakia, and Spain.

Family Cluster: India, Nepal, and the Philipines.

Pyramid Cluster: Brazil, Bulgaria, Chile, Colombia, Croatia, Ethiopia, Ghana, Greece, Iran, Mexico, Morocco, Nigeria, Pakistan, Portugal, Romania, Russia, Saudi Arabia, Slovenia, Turkey, United Arab Emirates, and Uruguay.

Table 3.1 show a summary of the dimensional scores of the clusters as defined by Hofstede et al. (2010).

Cluster	PD	IC	MF	UA	LTO
Contest Cluster	Low	High	High	Low	Low
Network Cluster	Low	High	Low	Low/Medium	Low
Well-Oiled Machine Cluster	Low	Can vary	Can vary	High	Can vary
Solar Cluster	High	High	Can vary	High	Can vary
Family Cluster	High	Low	Can vary	Low	Can vary
Pyramid Cluster	High	Low	Can vary	High	Can vary

Table 3.1: Dimension scores for the clusters (Hofstede et al. 2010)

3.2 Requirements for the Web Shops

The designs of the web shops are based on the guidelines from the literature review I did in my specialisation project at NTNU and the cultural dimension scores of each cluster. Some of the guidelines, like the attention guideline, has been transformed to be more suitable and recognisable in a picture of a web shop.

PD - Hierarchy:

The web shops of the clusters with a high PD score will have a menu with a tall hierarchical structure. The web shops of the clusters with a low PD score will have a menu with a shallow hierarchical structure.

IC - Navigation:

The web shops with a high IC score will mostly use text as a navigational tool, which means that links and buttons will use text to tell the user what it does. On the other hand, the web shops with a low IC score will use icons in addition to text on the buttons and links.

IC - Colours:

The web shops of the clusters with a high IC score will be homogeneously coloured while the clusters with a low IC score will have many different colours in their design.

IC - Images:

The web shops of the clusters with a high IC score will include images of individuals and their lifestyle while the web shops of the clusters with a low IC score will include images of products or merchandise being used by people.

MF - Attention on sales:

The web shops of the clusters with a high MF score will have visible sales pictures on the site while the web shops of the clusters with a low MF score will not have any visible sales pictures.

MF - Saturation:

The colours on the web shops of the clusters with a high MF score will be highly contrasting and bright while the colours on the web shops of the clusters with a low MF score will be pastel colours.

UA - Information Quantity:

The web shops of the clusters with a high UA score will be simple, with a limited amount of information available and not many ways to search for products. The web shops of the

clusters with a low UA score will have much visible information, and several different ways to search for and find products.

LTO - Information Density:

The web shops of the clusters with a high LTO score will show much information at first sight. The web shops of the clusters with a low LTO score will only show a minimal amount of information at first sight. Different types of information include links to pages containing more information and information about which products that the web shop contains.

Table 3.2 show a summary of the requirements for the high and low dimensional scores.

	Category	**High**	**Low**
PD	Hierarchy	Tall hierarchy	Shallow hierarchy
IC	Navigation	Textual navigation	Symbols
	Colours	Homogeniously coloured	Many different colours
	Images	Individuals and their lifestyle	Products or merchandise being used by people
MF	Attention	Focus on sales	No focus on sales
	Saturation	Highly contrasting and bright colours	Pastel colours with little saturation
UA	Information Quantity	Simple interface with limited amounts of data and choices	Complex interface with a lot of choices and data
LTO	Information Density	The website show all information at first sight	The website provide minimal information at first sight

Table 3.2: Requirements for the web shops

3.3 First Version of the Survey

This section will explain the first version of the designs of the web shops, which questions the survey included, and the results of the first pilot test and the first evaluation of the designs.

3.3.1 Design of the Web Shops

General Layout For The Web Shops

For the design of the web shops, a general layout of what the web shops should look like was made. The reason for this was to have a limited amount of choices for how to meet every requirement. Figure 3.1 shows the general layout of the web shops.

All the web shops have the same name and logo, Shopping.com. This name is at the same position and has the same text font in every design. The area marked with the number 1 is the area for the search bar, the log-in link and the link for the shopping cart. The area marked with the number 2 is the space for the menu, and the area marked with the number 3 can be used for sales posters, showing products or more information. What the third area contains and the placements of the different parts can vary depending on the requirements.

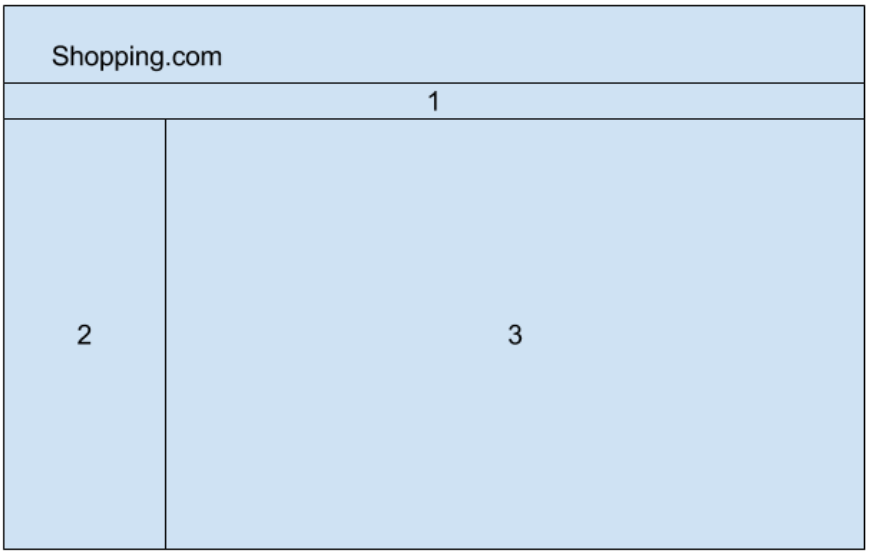


Figure 3.1: General web shop layout

The Contest Cluster Design

The characteristics of the Contest cluster is that it has low PD, UA, and LTO scores and high IC and MF scores. Therefore, the web shop should have a shallow hierarchy, a complex interface, textual navigation, and a focus on sales, be homogeneously coloured with highly contrasting and bright colours, show all information at first sight and include images of individuals and their lifestyle.

In the first design version of the Contest cluster, the hierarchy of the web shop is only visible on the menu to the left, which has only three visible levels. The links mostly only use text, except the shopping cart and the links with information about the web shop. The blue colour of the search bar is there to add a bright colour to the shop, at the same time as the overall colour scheme is homogeneous. The front page includes two sales posters, and the images on the posters are of hot air balloons and a girl jumping. These images

were chosen to give the user a positive and sporty impression of the shop. Another reason why the images were chosen was that they are blue, to keep the colouring of the shop homogeneous. The complexity of the interface is meant to be visible in the many ways that it is possible to search for products. The possibilities are to search for products by choosing categories in the menu, using the search bar to search for keywords, and by using the popular keywords links. In the search bar, it is also possible to choose a specific category to limit the search. Moreover, to minimise the visible information of the design, there are no product pictures.

Table 3.3 shows what was done in the design of the Contest cluster web shop to meet the requirements. Figure 3.2 shows what the design looks like.

	Requirement	**Solution**
PD	Menu with shallow hierarchy	One menu with three visible levels
IC	Textual navigation	Mostly textual links and buttons
	Homogeniously coloured	Only one additional colour used
	Images of individuals and their lifestyle	Images of hot air balloons and a girl jumping
MF	Focus on sales	Two sales posters
	Highly contrasting and bright colours	Colours used: • bright blue (#98DAFC)
UA	Complex interface	Three ways to search for products: • choosing categories in the menu • searching for keywords in the search bar • using the popular keywords links
LTO	Minimal information at first sight	No visible products

Table 3.3: The first design solution of the Contest cluster web shop

The Network Cluster Design

The Network Cluster has low PD, MF and LTO scores, a low to medium UA score, and a high IC score. Therefore, the web shop should have a shallow hierarchy, a somewhat complex interface, textual navigation, not focus on sales, be homogeneously coloured with pastel colours, show all information at first sight, and include images of individuals and their lifestyle.

The menu, the use of text in navigation, and the different ways to search for products in this web shop is the same as in the Contest cluster. However, there are no sales posters in this shop, and the middle section is used to display products while the right-hand side of

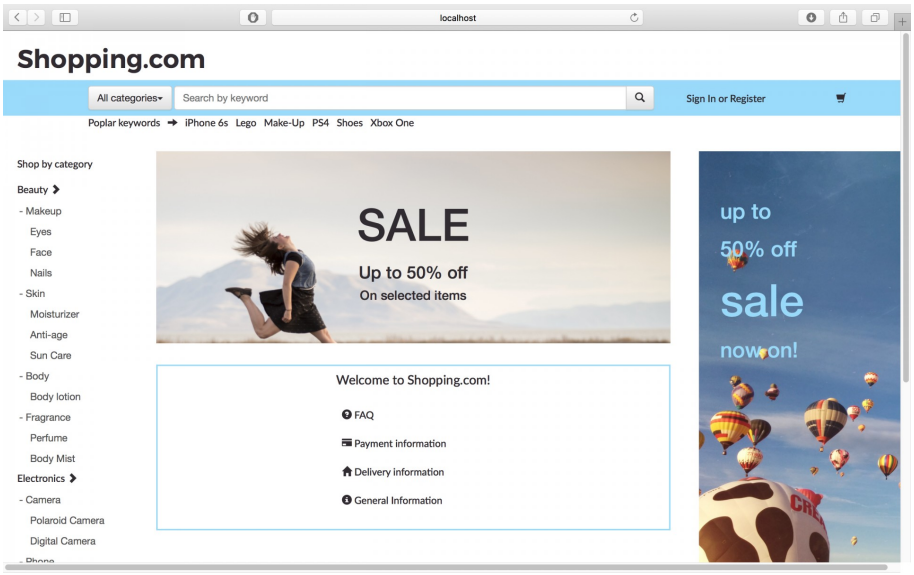


Figure 3.2: The first web shop design for the Contest cluster

the shop is used for the information links and an image of a girl jumping. The image is used as a decoration of the page and does not have any function other than making the web shop look a bit more alive. Even though the web shop should be homogeneously coloured, one light turquoise colour was added on different details to meet the requirement of using pastel colours. Also, the products are given much space, and the information links are moved to the right-hand side to make the user feel like there is not a lot of information visible, only products, which is expected.

Table 3.4 shows what was done in the design of the Network cluster web shop to meet the requirements. Figure 3.3 shows what the design looks like.

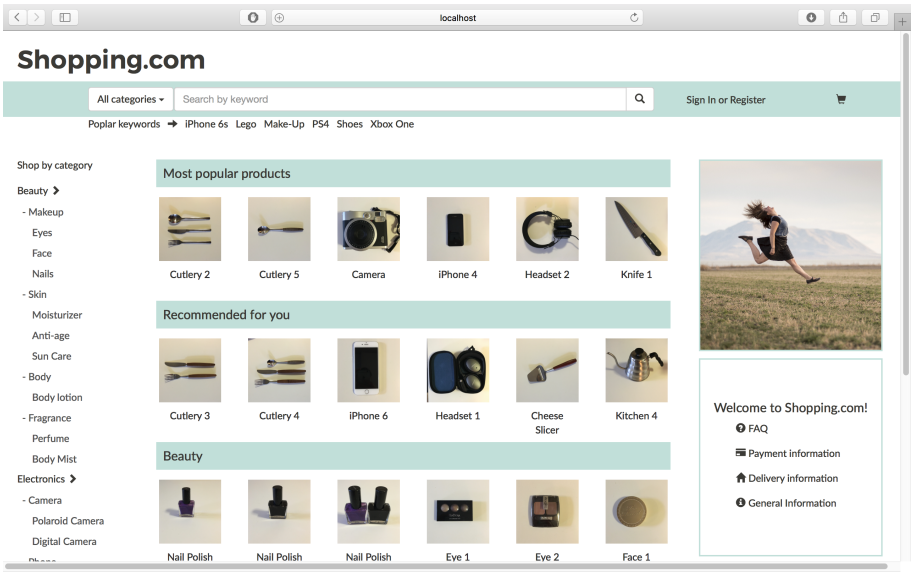


Figure 3.3: The first web shop design for the Network cluster

	Requirement	**Solution**
PD	Shallow hierarchy	One menu with three visible levels
IC	Textual navigation	Mostly textual links and buttons
	Homogeniously coloured	Only one additional color used
	Images of individuals and their lifestyle	Image of girl jumping
MF	No focus on sales	No sales posters
	Pastel colours with little saturation	Colours used: • pastel green (#C0DFD9)
UA	Complex interface	Three ways to search for products: • choosing categories in the menu • searching for keywords in the search bar • using the popular keywords links
LTO	Minimal information at first sight	Focus on products in stead of information to the user

Table 3.4: The first design solution of the Network cluster web shop

The Well-Oiled Machine Cluster Design

The Well-Oiled Machine cluster has a low PD score and high UA score. Because the rest of the scores can vary, there are only a few requirements for this design. This web shop should have a shallow hierarchy and have a simple interface.

The menu in this design has only one visible level, and the products that are visible on the front page are ordered by category. Also, to make things easy for the users of the web shop, some links to more information was placed next to the first line of products on the front page.

Table 3.5 shows what was done in the design of the Well-Oiled Machine cluster web shop to meet the requirements. Figure 3.4 shows what the design looks like.

	Require-ment	**Solution**
PD	Shallow hierarchy	One menu with one visible level
UA	Simple interface	Simple menu and the area with the information links is blue to make it easy to spot for the user

Table 3.5: The first design solution of the Well-Oiled Machine cluster web shop

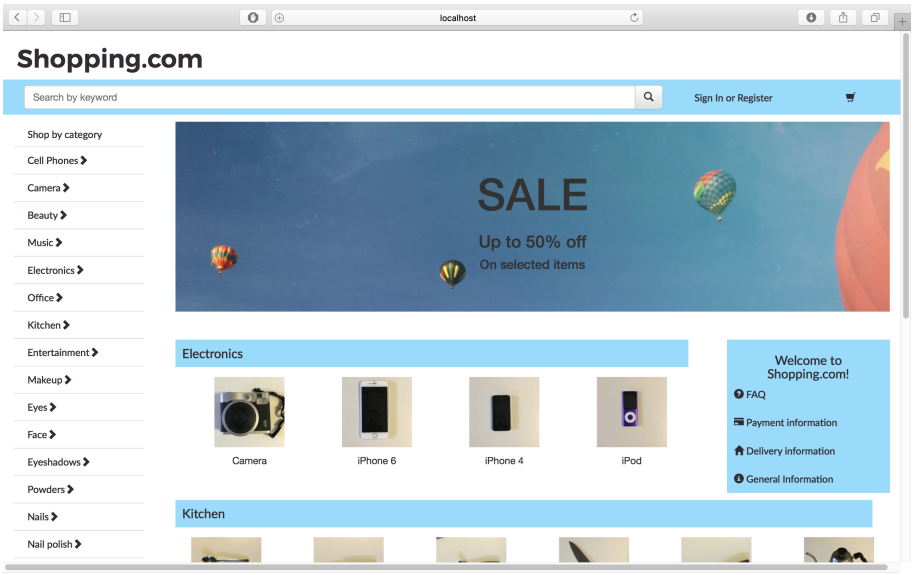


Figure 3.4: The first web shop design for the Well-Oiled Machine cluster

The Solar Cluster Design

The Solar cluster has high PD, IC and UA scores, while the MF and LTO scores can vary. Therefore, the web shop should have a simple interface, a tall hierarchy, textual navigation, be homogeneously coloured, and have images portraying individuals and their lifestyle.

The menu of this web shop has three visible levels, which indicates a certain hierarchy of the web shop. The upper half of the third area in the design is used for sales poster and links to information to keep the web shop simple even with the higher hierarchy. The lower area is used for showing products ordered by popularity and recommendations. One of the reasons for having a sales poster is because the web shop should contain an image that portrays individuals and their lifestyle. The hot air balloon picture is used because it implies that the shop has a sporty and fun attitude. Even though the web shop was supposed to be homogenously coloured, an additional colour was added to some of the details to make the web shop feel more alive. The extra colour was selected because it fits the image, and it does not make the web shop too much like the web shop for the Well-Oiled Machine cluster.

Table 3.6 shows what was done in the design of the Solar cluster web shop to meet the requirements. Figure 3.5 shows what the design looks like.

The Family Cluster Design

The Family cluster is characterised by its low IC and UA scores, and high PD score, while the rest of the values are undecided. Therefore, the interface should have a tall hierarchy, use icons in the navigation, be complex with much visible information and different choices for searching, contain many different colours, and have pictures of products or merchandise being used by people.

	Requirement	Solution
PD	Tall hierarchy	One menu with three visible levels
IC	Textual navigation	Mostly textual links and buttons
	Homogeniously coloured	One additional colour
	Images of individuals and their lifestyle	Image of hot air balloons
UA	Simple interface	Sales poster and information links at the same row and the information link area is in the additional colour to be easy to spot.

Table 3.6: The first design solution of the Solar cluster web shop

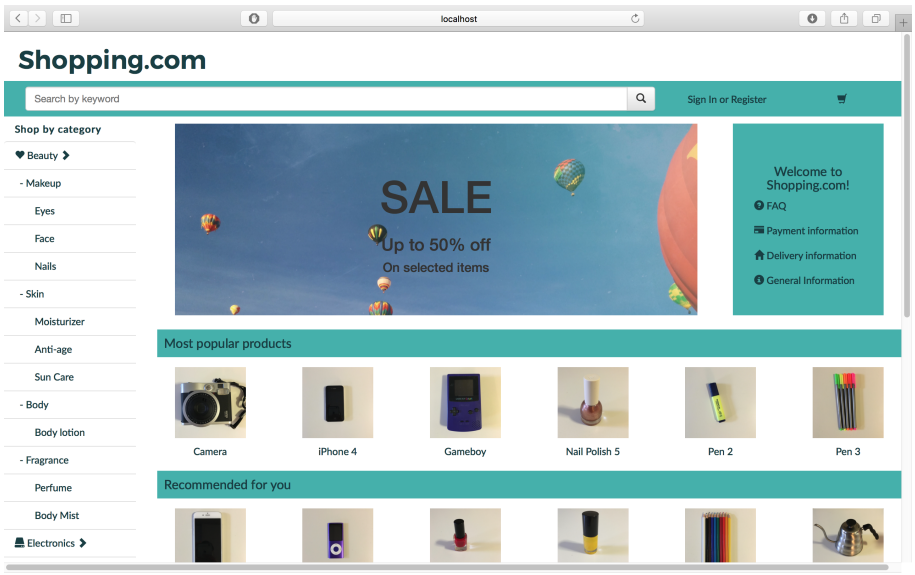


Figure 3.5: The first web shop design for the Solar cluster

This web shop is the one that is the most complex. It contains two menus, where one of them has three expanded levels while the other one is more simple, two different categories of products per line of product images, and a search field that can be used to search for products. Also, this web shop uses more icons on the links and in the menus than any of the other web shops. Still, a text is being used on almost all the links to avoid confusion. In this web shop, two different colours are used in the details, and the sales image is an image of someone using a phone.

Table 3.7 shows what was done in the design of the Family cluster web shop to meet the requirements. Figure 3.6 shows what the design looks like.

	Requirement	Solution
PD	Tall hierarchy	One menu with three expanded levels and one tab menu. Hierarchy shown in header above products
IC	Symbolic navigation	Symbols on almost all links and buttons
	Many different colours	Two additional colors
	Images of products or merchandise being used by people	Image of a phone being used by a person
UA	Complex interface	Two menus and two colums of product pictures per row

Table 3.7: The first design solution of the Family cluster web shop

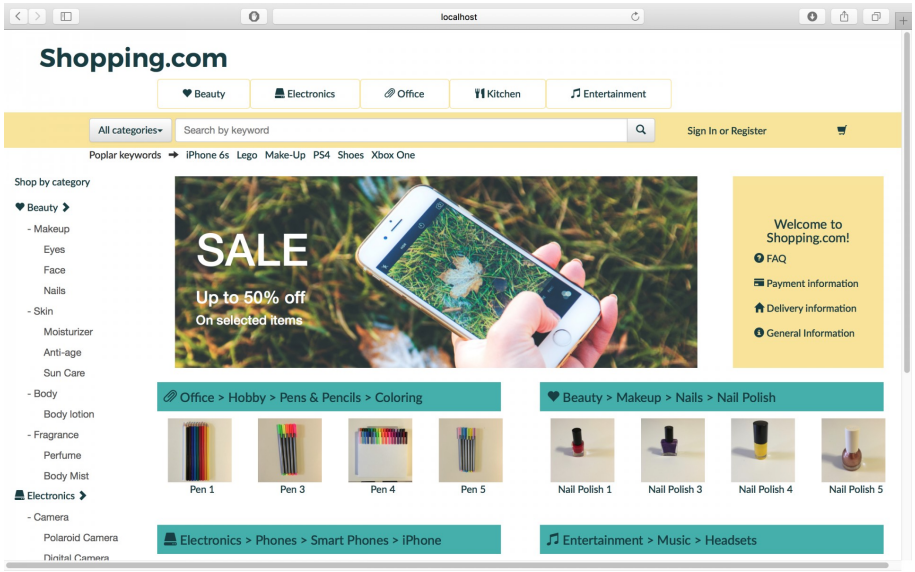


Figure 3.6: The first web shop design for the Family cluster

The Pyramid Cluster Design

The Pyramid cluster has the same scores as the Family cluster, except its high UA score. Therefore, the web shop should have a tall hierarchy, use icons in the navigation, be simple with little information and few different choices for searching for products, contain many different colours, and have pictures of products or merchandise being used by people.

In this web shop, there are only two visible levels in the menu. However, the headers over the product pictures show that the hierarchy has more levels than what is visible on the menu. Also, some icons are used in the navigation in addition to text. The colour scheme is different than for the Family cluster to avoid making them look too much alike, but there is still two distinct colours used on details and most of the product pictures include several colours. Also, a colourful sales poster is included to add more colour to the web shop. The

different methods for searching for products are to search by keyword in the search field, choose categories in the menu, or choose one of the products on the front page. There are no images of products being used by people, but there are many images of products. Also, to make the interface look simple, the area with the information links, and the search bar and product headers, has colourful backgrounds in two different colours to separate them and make them easy for the user to see.

Table 3.8 shows what was done in the design of the Pyramid cluster web shop to meet the requirements. Figure 3.7 shows what the design looks like.

	Requirement	**Solution**
PD	Tall hierarchy	One menu with two expanded levels. Hierarchy shown in header above products
IC	Symbolic navigation	Symbols on outer menu category, information links and shopping chart
	Many different colours	Two additional colors
	Images of products or merchandise being used by people	Image of dots in different colours
UA	Simple interface	Different kinds of information is grouped and highlighted with different colours

Table 3.8: The first design solution of the Pyramid cluster web shop

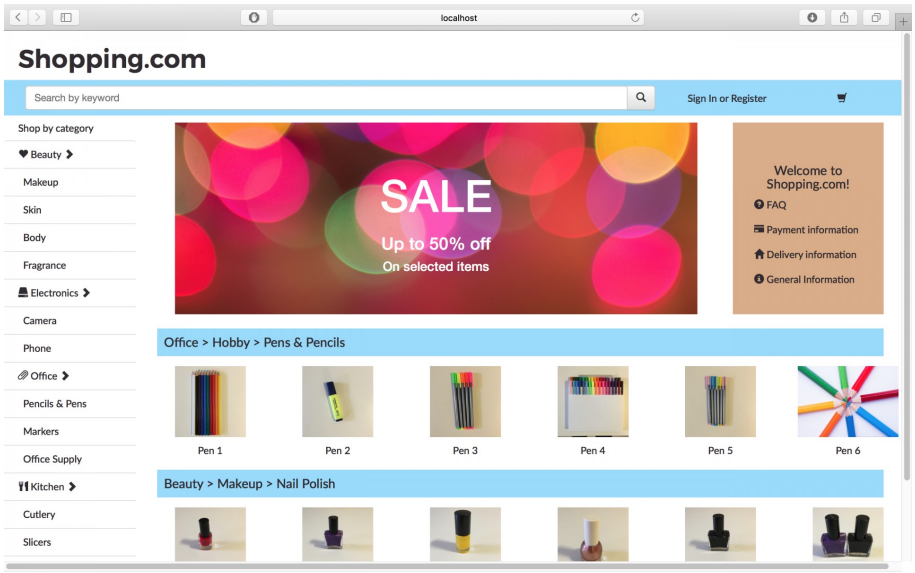


Figure 3.7: The first web shop design for the Pyramid cluster

3.3.2 Structure of the Survey

This survey was split into two parts. In the first part, the participants were asked to look at an image of one of the web shops and answer a question. Every web shop had two questions linked to it, but only one question was visible at a time. The first questions asked the participants to rate the design based on its beauty. The scale was a 7-point Likert scale that went from "not beautiful" (1) to "very beautiful" (7). The second question asked the participants to rate the design based on how easy they thought the site was to use. The scales used for this question was also a 7-point Likert scale, from "not easy" (1) to "very easy" (7). The second part of the questionnaire asked about the participants age, gender and where they were born.

3.3.3 Pilot-Test

In this pilot test, nine Norwegian students completed the questionnaire and then answered some simple questions about their experience of taking the questionnaire. The objectives of the pilot-test were to find out if the instructions of the questionnaire were sufficient, if any questions were hard to understand or ambiguous, how long it took to finish the questionnaire and to identify any possible problem areas in the questionnaire.

During the pilot-testing phase, two major problem areas were discovered. The first was about the designs of the web shops and the second was the questions being asked.

The problem with the designs of the web shops was that they were too similar. All the participants of the pilot-test commented that they thought many of the web shops were too much alike and therefore felt like they needed to give the different web shops the same ratings. This attitude was also obvious in the results of the questionnaire, which showed little variance in the ratings of each participant.

The questions in the survey were the other problem area. The participants did not have any problems understanding the questions, but some of them commented on that it was hard to give a score for the beauty and usability based on only one question for each. Also, to try to get a better understanding of what the participants think about the different web shops in more detail, more questions were needed. Moreover, to add more questions is also beneficial when analysing the responses.

3.3.4 Evaluation of the Web Shops

The first version of the web shop designs were all quite similar, and it was quite obvious from the pilot test that they all needed to be changed. However, it was not obvious what this change is. Therefore, an evaluation of the requirements of the designs was performed, with employees at NTNU as the participants. The categories that the participants evaluated the designs on was the hierarchy, the complexity, choices, information quantity, use of colours, saturation, focus on sales, navigation. Appendix C contains all the questions in their entirety.

The Contest Cluster

Table 3.9 show the evaluation scores of the contest cluster. As can be seen in this table, some of the values do not fit with the scores that it was designed for. In more detail, the hierarchy is too tall, the colours used in the design are not bright enough, and the web shop is not complex enough. Also, the images used are considered to be neutral, while they should be portraying individuals and their lifestyle.

Dimension	Requirement	Dimension score	Evaluation score
PD	Hierarchy	Low	High
IC	Navigation	High	High
	Colorfulness	High	High
	Images	High	Medium
MF	Attention	High	High
	Saturation	High	Low
UA	Information Quantity	Low	High
LTO	Information Density	Low	Low

Table 3.9: First evaluation of the Contest cluster

The Network Cluster

Table 3.10 shows the evaluation scores of the Network cluster. This table shows that the requirements about the colorfulness, information quantity, navigation, image and saturation in this web shop are met. However, the hierarchy is too tall, there is too much focus on sales, and too much information is visible to the user at first sight.

Dimension	Requirement	Dimension score	Evaluation score
PD	Hierarchy	Low	High
IC	Navigation	High	High
	Colorfulness	High	High
	Images	High	High
MF	Attention	Low	High
	Saturation	Low	Low
UA	Information Quantity	Low/Medium	Low
LTO	Information Density	Low	High

Table 3.10: First evaluation of the Network cluster

Well-Oiled Machine Cluster

Table 3.11 shows the evaluation scores of the Well-Oiled Machine cluster. This table shows that the two requirements that are set for this cluster each has gotten a medium

score. These scores indicate that the hierarchy is too tall, and the web shop is not simple enough. Therefore, the entire design of this web shop should be simplified.

Dimension	Requirement	Dimension score	Evaluation score
PD	Hierarchy	Low	Medium
UA	Information Quantity	High	Medium

Table 3.11: First evaluation of the Well-Oiled Machine cluster

The Solar Cluster

Table 3.12 shows the evaluation scores of the Solar cluster. In this web shop, the navigation and colorfulness are all right, but the hierarchy of the menu is too shallow, the image needs to portray individuals and their lifestyle more clearly, and the web shop is not simple enough.

Dimension	Requirement	Dimension score	Evaluation score
PD	Hierarchy	High	Low
IC	Navigation	High	High
	Colorfulness	High	High
	Images	High	Medium
UA	Information Quantity	High	Low

Table 3.12: First evaluation of the Solar cluster

The Family Cluster

Table 3.13 shows the evaluation scores of the Family cluster. This table shows that all the requirements, except for one are met, which is the navigation. The navigation should use more icons compared to the web shops with more textual navigation.

Dimension	Requirement	Dimension score	Evaluation score
PD	Hierarchy	High	High
IC	Navigation	Low	High
	Colorfulness	Low	Low
	Images	Low	Low
UA	Information Quantity	Low	Low

Table 3.13: First evaluation of the Family cluster

The Pyramid Cluster

Table 3.14 shows the evaluation scores of the Family cluster. This table shows that all the requirements except the one about information quantity are met. Therefore, the web shop design needs to be simpler overall.

Dimension	Requirement	Dimension score	Evaluation score
PD	Hierarchy	High	High
IC	Navigation	Low	Low
	Colorfulness	Low	Low
	Images	Low	Low
UA	Information Quantity	High	Low

Table 3.14: First evaluation of the Pyramid cluster

3.3.5 Changes Needed in the Survey

Based on the results from both the pilot-test and the evaluation performed on the first version of the survey, a second version of the survey was designed. Because the designs in the first round were found too similar to each other, new designs were made which focused on trying to separate the design more, while also fulfilling the requirements that were not met in the first version. Also, the number of questions for each design was increased to capture the participants opinions about the different aspects of the design better.

3.4 Second Version of the Survey

3.4.1 Design of the Web Shops

The Contest Cluster Design

In this version of the web shop for the contest cluster, several changes have been made. Firstly, in the first version of the Contest cluster design, the hierarchy was too tall. Because of this, the menu in this version of the design is more shallow than the menu in the first design.

Secondly, in the first design, the images were not considered to show individuals and their lifestyle clearly enough, and the colours of the website were not regarded as bright enough. To solve these problems, the two sales posters were replaced by one sales poster that both was showing people having fun, and that did not bring pastel colours into the design. By choosing this new picture, the colour in focus was only the bright blue colour, which makes the entire design more bright. However, because one of the sales posters were removed, product pictures were inserted to make the user feel like the web shop has a focus on sale.

Thirdly, the interface was not complex enough in the first design. Because pictures of products are introduced in this design, it immediately seems more complex. However,

there is also a requirement that says that there should only be minimal information visible at first sight. Because of this, the popular keywords links and the categories drop-down next to the search bar are removed to make the interface seem less cluttered, but not that much simpler.

Table 3.15 shows what was done in the design of the Contest cluster web shop to meet the requirements. Figure 3.8 shows what the design looks like.

	Requirement	**Solution**
PD	Shallow hierarchy	One menu with two visible levels
IC	Textual navigation	Mostly textual links and buttons
	Homogeniously coloured	Only one additional colour used
	Images of individuals and their lifestyle	Images of people lying on a car
MF	Focus on sales	One sales poster
	Highly contrasting and bright colours	Colours used: • bright blue (#98DAFC)
UA	Complex interface	Two ways to search for products: • choosing categories in the menu • searching for keywords in the search bar
LTO	Minimal information at first sight	Some products are shown

Table 3.15: The second design solution of the Contest cluster web shop

The Network Cluster Design

The main problem with the first version of this design was that there was too much information visible in the shop. Therefore, the user could feel like there is much focus on sales and the products even though the requirements say that there should be a minimal focus on sales. To solve this issue, some changes were made.

First of all, the products being visible to the user when entering the web shop should be smaller. Therefore, a bigger version of the image used as decoration on the right side of the web shop in the first version was placed on the row under the search field. This picture makes the user look more at the image and less at the products. Smaller changes made includes that the icons on the information links, the category dropdown next to the search field, and the links for quick searching on popular keywords were removed. These changes were made to make the design less cluttered and minimise the visible information without making the web shop seem too simple.

Another problem was that the hierarchy of the menu was too tall. This issue was changed by simplifying the menu only to have one visible level.

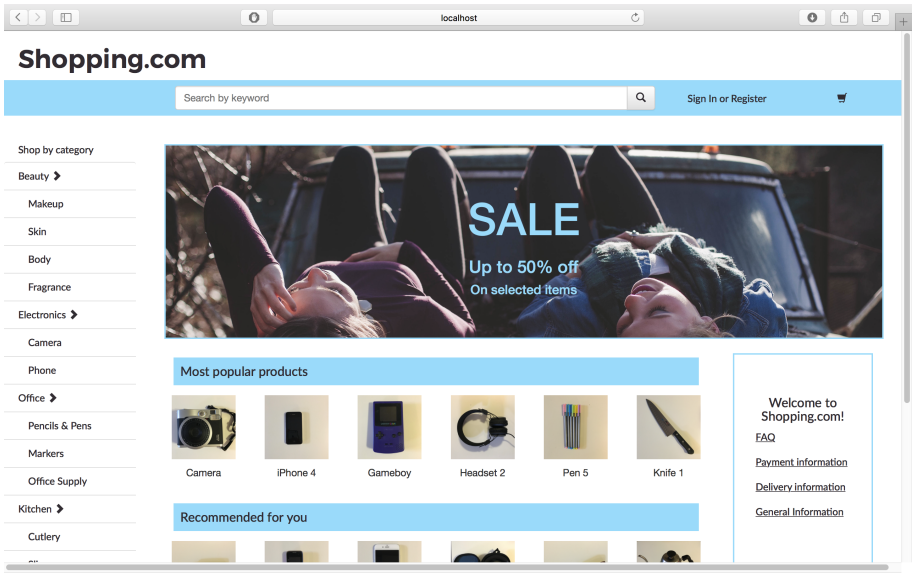


Figure 3.8: The second web shop design for the Contest cluster

Table 3.16 shows what was done in the design of the Network cluster web shop to meet the requirements. Figure 3.9 shows what the design looks like.

	Requirement	**Solution**
PD	Shallow hierarchy	One menu with one visible level
IC	Textual navigation	Mostly textual links and buttons
	Homogeniously coloured	Only one additional color used
	Images of individuals and their lifestyle	Image of girl jumping
MF	No focus on sales	No sales posters
	Pastel colours with little saturation	Colours used: • pastel green (#C0DFD9)
UA	Complex interface	Two ways to search for products: • choosing categories in the menu • searching for keywords in the search bar
LTO	Minimal information at first sight	Few products visible to the user, focus on the picture of the jumping girl

Table 3.16: The second design solution of the Network cluster web shop

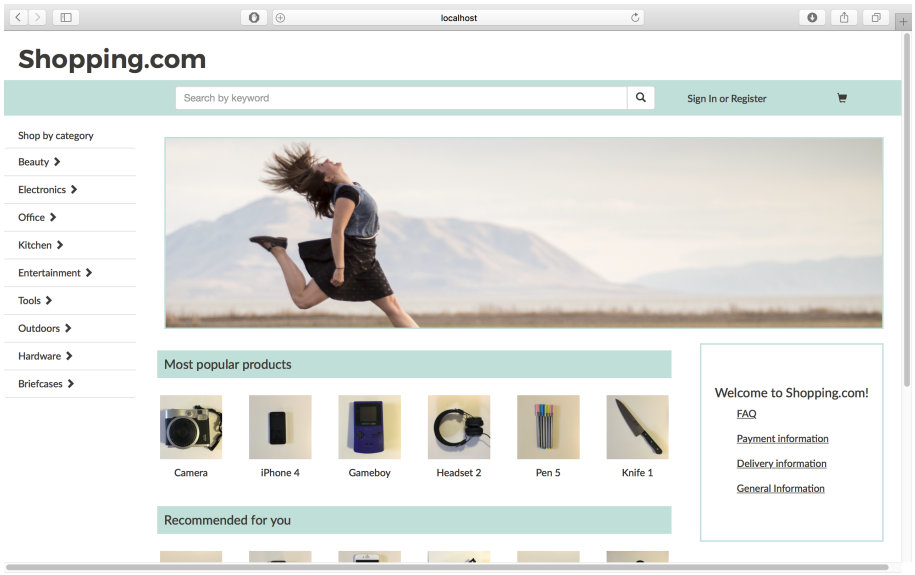


Figure 3.9: The second web shop design for the Network cluster

Well-Oiled Machine Cluster Design

The main focus when designing the second version of this web shop was to simplify the web shop as much as possible. The main change in this design was that the entire search bar was removed, and the menu was moved to the top of the page. The menu was also simplified by only showing the main categories of the web shop. Also, the information links were hidden in a drop down menu at the top of the page, next to the shopping chart and the sign in or register links. The rest of the page stayed the same, only with more space between the products on the page.

Table 3.17 shows what was done in the design of the Well-Oiled Machine cluster web shop to meet the requirements. Figure 3.10 shows what the design looks like.

	Requirement	**Solution**
PD	Shallow hierarchy	Small menu only showing a few categories in one level
UA	Simple interface	No search bar or information links. Few visible products

Table 3.17: The second design solution of the Well-Oiled Machine cluster web shop

The Solar Cluster Design

The main focus in the redesign of this web shop is to create a simpler interface, and at the same time make the hierarchy seem taller and put a bigger focus on lifestyle rather than products. This design is much simpler than the first version because the product pictures have been removed and replaced by the information links. These changes also make the users attention go towards the balloon picture instead of the products. Also, the search field is removed, and the menu only has two levels. However, the menu had '+' icons on each category on the second level, which indicates that the menu is extensible and to not give the impression of a shallow hierarchy.

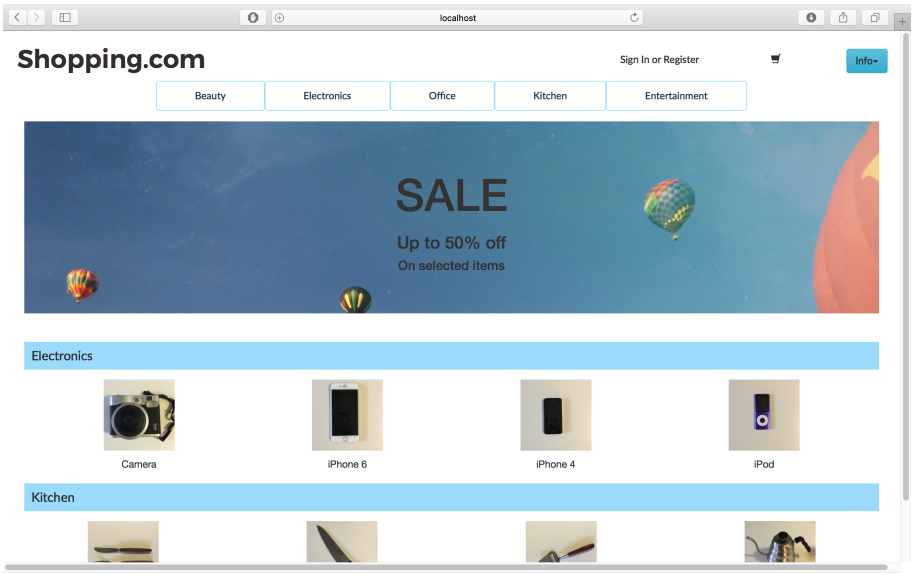


Figure 3.10: The second web shop design for the Well-Oiled Machine cluster

Table 3.18 shows what was done in the design of the Solar cluster web shop to meet the requirements. Figure 3.11 shows what the design looks like.

	Requirement	**Solution**
PD	Tall hierarchy	One menu with two visible levels
IC	Textual navigation	Mostly textual links and buttons
	Homogeniously coloured	One additional colour
	Images of individuals and their lifestyle	Image of hot air balloons
UA	Simple interface	No product pictures or search field.

Table 3.18: The second design solution of the Solar cluster web shop

The Family Cluster Design

On this design, only minimal changes were made. The requirements specify that this web shop should contain many colours, so a third colour was introduced. This colour was then used on the top menu and as a background for the information links. Also, the top menu was made larger by adding more categories to make the shop more complex to the user. Also, lines and '+' icons were added in the left-side menu to separate the different elements more, make it look tidier, and to add more icons. Lastly, the 'Sign In or Register' link also got an icon.

Table 3.19 shows what was done in the design of the Family cluster web shop to meet the requirements. Figure 3.12 shows what the design looks like.

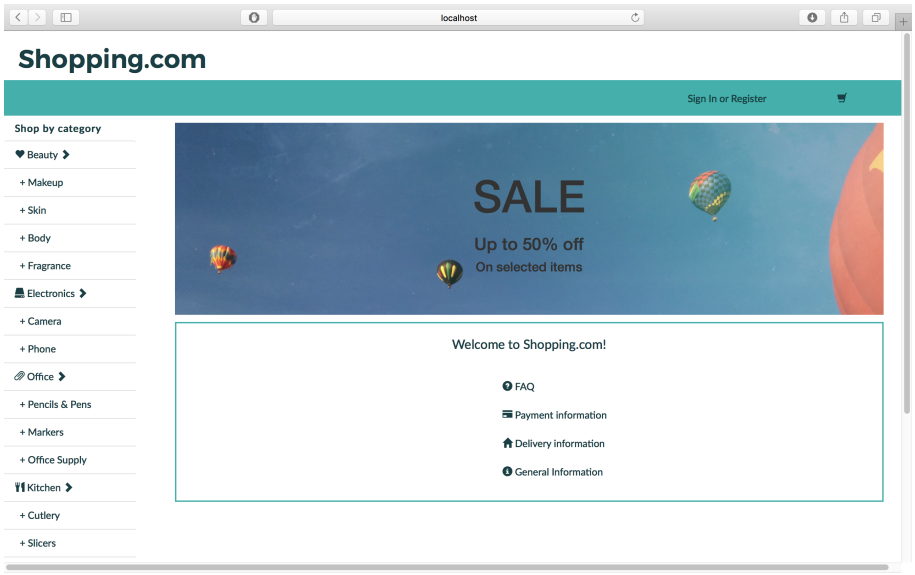


Figure 3.11: The second web shop design for the Solar cluster

	Requirement	Solution
PD	Tall hierarchy	One menu with three expanded levels and one tab menu. Hierarchy shown in header above products
IC	Symbolic navigation	Symbols on almost all links and buttons
	Many different colours	Three additional colours
	Images of products or merchandise being used by people	Image of a phone being used by a person
UA	Complex interface	Two menus and two columns of product pictures per row

Table 3.19: The second design solution of the Family cluster web shop

The Pyramid Cluster Design

The main focus in the redesign of this web shop was to make it simpler. Therefore, all the product pictures were hidden and replaced with links with the category names. These links were highlighted in the bright blue colour used for details on this website. Also, the entire search field, the shopping cart, the login and register links and the information links were removed, and a drop down menu was inserted instead. Also, a sales poster was added, and to add some more colour, the text on this new sales poster was made red. Lastly, the levels of the menus were made clearer, and '+' icons were added to help separate between the different labels, and to show that the menu is extensible.

Table 3.20 shows what was done in the design of the Pyramid cluster web shop to meet the requirements. Figure 3.13 shows what the design looks like.

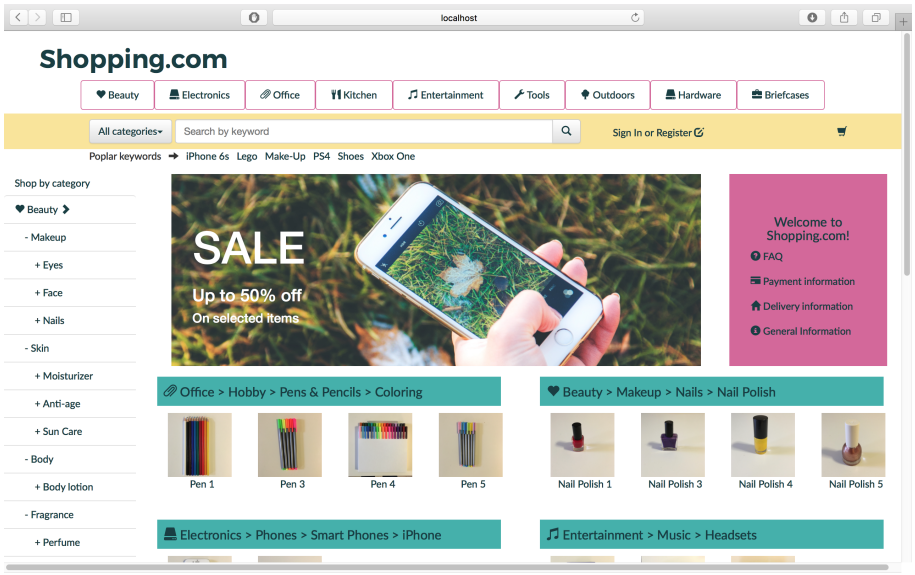


Figure 3.12: The second web shop design for the Family cluster

	Requirement	Solution
PD	Tall hierarchy	One menu with two expanded levels.
IC	Symbolic navigation	Only symbols on first level in the menu
	Many different colours	Two additional colors
	Images of products or merchandise being used by people	Image of dots in different colours
UA	Simple interface	No search field or shopping cart, register, or information links.

Table 3.20: The second design solution of the Pyramid cluster web shop

3.4.2 Structure of the Survey

Background Information

To be able to compare how people from various cultures differ in their opinion about the aesthetics and usability of the website designs, some background information is needed. The most important background information of the participants in this questionnaire is where they are from and where they have lived because this is what decide their cultural cluster. To be able to place the participants in the right cultural cluster, three questions were asked: “Where are you from?”, “in which country do you currently live?”, and “in which country have you lived most of your life?”. When the participants are divided into different clusters, their birthplace is the most important. However, if someone has spent most of his or her life and is still living in a country that belongs to a different cluster, the participant will be placed in that cluster. The other background information questions were

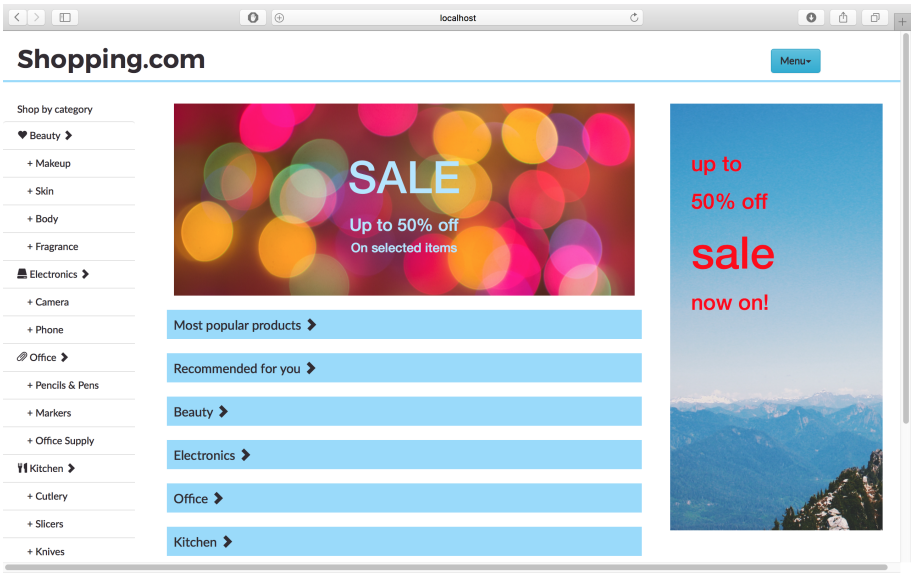


Figure 3.13: The second web shop design for the Pyramid cluster

about the age, gender, educational level, Internet use, and online shopping experiences of the participants. The questions can be found in Appendix D.

Web Shop Questions

In the second part of the survey, the six web shop prototypes will be evaluated. Every web shop has three pages with statements, one page for statements about the aesthetics of the web shop, one about the usability of the web shop, and one about whether the participant would consider using a web shop designed like this in real life. Each statement should be answered on a seven-point Likert scale from 'strongly disagree' (1) to 'strongly agree' (7).

Aesthetics:

The aesthetics of the design was split into three categories; design and style/atmosphere, presentation quality and design elements (Constantinides 2004). The design and style/atmosphere category is related to the actual design and style of the website as a whole and includes the user's first impression of the site. The presentation quality category is related to the total impression of the website after the first impression has settled. The last category, design elements, is about specific design elements that contribute to the web experience, and in this questionnaire, the focus is on the colours and the images and icons of the websites.

Inside each of these three categories, two additional dimensions of aesthetics were used, classical aesthetics and expressive aesthetics, as proposed by Lavie and Tractinsky (2004). Through four different studies they found that user's perceptions consist of two dimensions where the first dimension, classical aesthetics, refer to orderly, clean and clear design, and the second dimension, expressive aesthetics, refer to creativity and originality. Based on

this the questionnaire ended up with six statements about the aesthetics, which can be seen in Table 3.21.

#	Question	Source
1	The overall look of the website is visually appealing	Lindgaard et al. (2006), Kurosu and Kashimura (1995), Cyr et al. (2006), O'Brien and Toms (2010), Kirakowski et al. (1998), Tullis and Stetson (2004), Seo et al. (2016), Kim et al. (2003)
2	I think the design of this website is clear	Tullis and Stetson (2004), Kirakowski et al. (1998), Seo et al. (2016)
3	I think the design of this website is creative	Lavie and Tractinsky (2004), Seo et al. (2016), Seckler et al. (2015)
4	I think the layout appears professionally designed	Cyr et al. (2006), Seckler et al. (2015)
5	I like the icons and images used on this website	Cyr et al. (2006), O'Brien and Toms (2010), Kim et al. (2003)
6	I like the use of colors on this website	(Seckler et al. 2015)

Table 3.21: Questions about aesthetics

Usability:

To measure what the participants thought about the perceived usability of the websites the statements were split into two categories; convenience and site navigation, information architecture and search facilities/search processes (Constantinides 2004). These two categories were used because they do not require the participant to use the website to answer the questions. The first category, convenience, is about the ease of use of the website. To have good convenience means that the website is easy and fast to use for the tasks that the website should be able to perform, like information browsing and shopping. The other category, site navigation, information architecture and search facilities/search processes, is as the name implies about the navigation of the site and the search processes available on the website. Based on this the questionnaire ended up with five statements about the perceived usability, shown in Table 3.22.

#	Question	Source
1	This website seems easy to use	Kurosu and Kashimura (1995), Tullis and Stetson (2004), Seo et al. (2016)
2	The structure of the website seems logical to me	Christine Roy et al. (2001), Tullis and Stetson (2004), Seckler et al. (2015)
3	It would be easy to learn how to use this website	Tullis and Stetson (2004), Seo et al. (2016)
4	It seems easy to search for and find products on this website	Kirakowski et al. (1998), Christine Roy et al. (2001), Tullis and Stetson (2004)
5	It is easy to understand what the goal of this website is	Christine Roy et al. (2001)

Table 3.22: Questions about usability

Purchasing and Revisiting:

To measure whether the participants would consider using a web shop designed like the ones in the questionnaire two questions were asked, which can be seen in Table 3.23.

#	Question	Source
1	I would like to purchase a product from this website.	Cyr et al. (2006), Doney and Cannon (1997)
2	I would like to visit this website again.	Cyr et al. (2006), Tullis and Stetson (2004)

Table 3.23: Questions about purchasing or revisiting

3.4.3 Pilot-Test

In this round of pilot-testing eight Norwegian students participated. The tests were performed in the same manner as the last pilot-test, namely by letting the participants do the survey and then answer another questionnaire with simple questions about the survey. The objective of this pilot-test was like in the last pilot-test to find out if the instructions of the questionnaire were sufficient, if any questions are hard to understand or ambiguous, how long it takes to finish it and to identify any possible problem areas.

The results of this pilot test were overall positive. All the participants understood the instructions and thought the survey was easy to answer. Also, no one had any problems understanding the different questions, and most of the participants thought the number of questions was ok. In this round of pilot-testing, no major problem areas were discovered,

but some minor changes will be made. Not all the countries in the world have been given scores in Hofstede's dimensions, and it is, therefore, unnecessary to have a list of all the countries in the world in the survey. Instead, the list will only contain the countries that have a score in Hofstede's dimensions, and a text box was placed underneath the list of the participants that are from or have lived in countries that are not on the list.

3.4.4 Evaluation of the Web Shops

Before sharing the survey, a new evaluation of the new website designs was performed. The purpose of this assessment was to get an idea about whether the requirements of the designs were met. The participants of the evaluation were employees at NTNU, but not the same ones as in the first evaluation. Also, based on feedback from the first assessment, some of the questions were changed to make them easier to understand. The categories that were asked about was the hierarchy, information density, choices, patience, colorfulness, saturation, images, focus on sales, and navigation. Appendix E contains all the questions in their entirety.

The Contest Cluster

Table 3.24 show the evaluation scores of the contest cluster. In this assessment, it is clear that the hierarchy of the web shop is still too tall. The hierarchy is, however, the only requirement that was not met and it is, therefore, easy to correct because the only item that should be changed is the menu. This change will not affect any of the other requirements.

Dimension	Requirement	Dimension score	Evaluation score
PD	Hierarchy	Low	High
IC	Navigation	High	High
	Colorfulness	High	High
	Images	High	High
MF	Attention	High	High
	Saturation	High	High
UA	Information Quantity	Low	Low
LTO	Information Density	Low	Low

Table 3.24: Second evaluation of the Contest cluster

The Network Cluster

Table 3.25 show the evaluation scores of the Network cluster. This design was very similar to the previous design because the original design almost met its requirements. From the results of this evaluation, one can see that the hierarchy and attention scores have gotten lower, the information quantity and information density scores have gotten higher, and the rest is unchanged. Therefore, this version of the design meets all its requirements.

Dimension	Requirement	Dimension score	Evaluation score
PD	Hierarchy	Low	Low
IC	Navigation	High	High
	Colorfulness	High	High
	Images	High	High
MF	Attention	Low	Low
	Saturation	Low	Low
UA	Information Quantity	Low/Medium	Medium
LTO	Information Density	Low	Low

Table 3.25: Second evaluation of the Network cluster

Well-Oiled Machine Cluster

Table 3.26 show the evaluation scores of the Well-Oiled Machine cluster. The main focus of the redesign of this web shop was to make it a lot simpler. As can be seen in Table 3.26, the second design of the web shop meets its requirements.

Dimension	Requirement	Dimension score	Evaluation score
PD	Hierarchy	Low	Low
UA	Information Quantity	High	High

Table 3.26: Second evaluation of the Well-Oiled Machine cluster

The Solar Cluster

Table 3.27 show the evaluation scores of the Solar cluster. When re-designing this web shop the goal was to get a web shop that would be evaluated as having high scores in the PD, IC, and UA dimensions. In the first version of this web shop, the evaluation showed that the hierarchy was too shallow, the images were too neutral and that the interface was overall too complex. In this second version, all of this was corrected, and the requirements are now met.

Dimension	Requirement	Dimension score	Evaluation score
PD	Hierarchy	High	High
IC	Navigation	High	High
	Colourfulness	High	High
	Images	High	High
UA	Information Quantity	High	High

Table 3.27: Second evaluation of the Solar cluster

The Family Cluster

Table 3.28 show the evaluation scores of the Family cluster. The only requirement that was not met in the first version of the web shop was the one about navigation, and whether icons or text was being used. However, in the second design, a couple of more changes was made to create greater differences between the designs. These changes did not affect the evaluation of the design negatively, and this design can be used in the survey.

Dimension	Requirement	Dimension score	Evaluation score
PD	Hierarchy	High	High
IC	Navigation	Low	Low
	Colorfulness	Low	Low
	Images	Low	Low
UA	Information Quantity	Low	Low

Table 3.28: Second evaluation of the Family cluster

The Pyramid Cluster

Table 3.29 show the evaluation scores of the Pyramid cluster. In this design, the image requirement and the information quantity requirement is not met. When removing the pictures of the products and adding another sales poster without any products, the image score showed that the participants got an impression that the web shop was focused on individuals and their lifestyle rather than on products or products being used by people. Also, the interface is still not simple enough.

Dimension	Requirement	Dimension score	Evaluation score
PD	Hierarchy	High	High
IC	Navigation	Low	Low
	Colorfulness	Low	Low
	Images	Low	High
UA	Information Quantity	High	Low

Table 3.29: Second evaluation of the Pyramid cluster

3.4.5 Final Changes in the Survey

After the second round of pilot testing and evaluation of the designs, only a few changes needed to be made to the questionnaire. These changes include both some adjustments on the background questions in the survey, and some corrections on the web shop designs.

Background Questions

The feedback on the questions in the second round of pilot testing was good, so no changes were done on the questions in the questionnaire. However, because not all countries have received a score in Hofstede's dimensions, the drop-down list of countries was changed only to contain the countries that have a score. Also, a textbox was placed under the menu for participants from other countries. This was done to make it easier to filter out the participants that can not be placed in a cluster so that they are easily identifiable. The reason why they can still be participants in this study is that their data is useful when analysing the research question that is about whether a user's perception of beauty affects their perception of usability.

Web shop designs

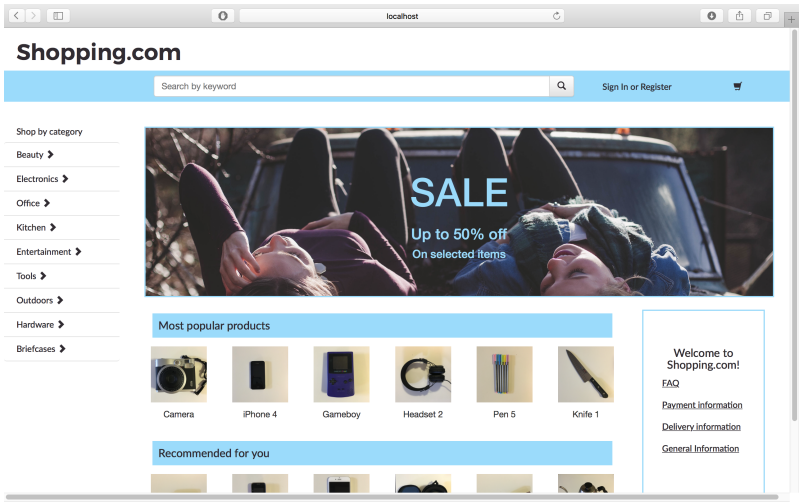
After the second evaluation, most of the designs met all their requirements and will, therefore, be used as they are in the survey. For two of the designs, some minor changes needed to be made before they could be used in the questionnaire. These designs were the designs for the Contest cluster and the Pyramid cluster.

For the Contest cluster design, the hierarchy of the menu needed to be lower, so the menu was replaced with a new menu with only one visible level. For the Pyramid cluster design, the images were supposed to show more of products being used by people, so the sales images were replaced by one big image of someone using a phone. Because the sales poster at the right side of the web shop was removed, the menu and new sales poster were made longer to take up space. Figures 3.14a and 3.14b show the final design of respectively the Contest cluster web shop and the Pyramid cluster web shop.

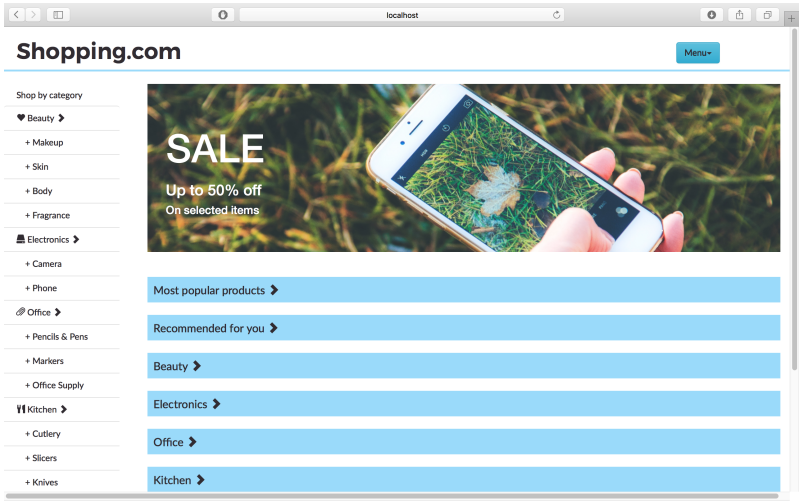
3.5 Ethical Considerations

To ensure that the research conducted is considered ethical, two steps have been taken. The first step is the participants right not to participate, to give informed consent and to withdraw. The survey is advertised and shared through social media, and it is entirely voluntary to participate. Moreover, it is possible to visit the website and read about the research and how the data will be handled and used before deciding whether or not to take the survey. Moreover, the participants get information about that they can leave the survey whenever they want before starting the survey.

The second step is anonymity. This research is collecting demographics and information about the participants, and their opinions on different images. This data can not be linked back to the respondent at any point, and no IP address is stored or linked to any answer. Therefore, it is impossible to know who responded to the survey and whether the same person has answered it several times.



(a) The final design for the Contest cluster



(b) The final design for the Pyramid cluster

Figure 3.14: Final designs of the Contest and Pyramid cluster designs

3.6 Participants

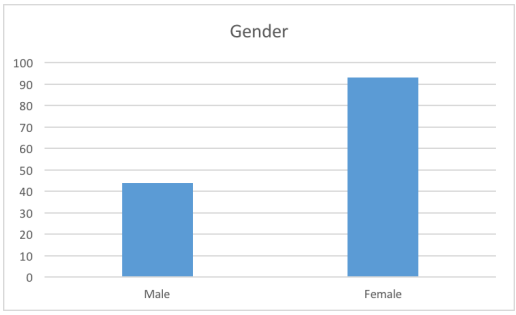
To be able to get participants from a lot of different countries the questionnaire was shared on Facebook in groups with members from many different countries, and some groups with members mostly from one country. Most of the groups were groups of international students at various universities, cities or countries, and some groups were groups of people with the same interests. Appendix F contains a full list of the groups that were used to share the questionnaire.

Through the process of collecting data 167 people started the questionnaire and 100 people finished it entirely. In this research, every submission that evaluated one web shop or more are counted as participants, so the number of participants in the analysis was 138. Both genders, all the age groups, all educational levels and all the clusters are represented, although the participants are not evenly divided between the groups. This division was anticipated because of the sampling technique, and the only requirement to the data was

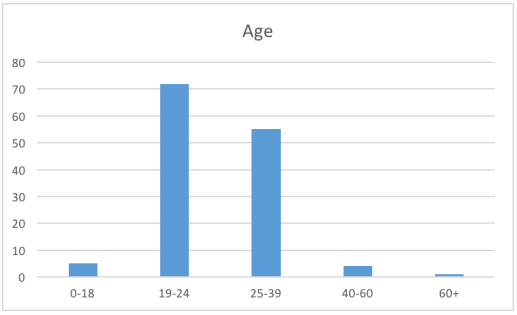
that all clusters should have at least ten participants that have evaluated at least one web shop.

Figure 3.15a show the gender distribution and Figure 3.15b show the age distribution of the participants. From these figures it is clear that the majority of the participants in this study are female and between 19 and 39 years old. Moreover, in Figure 3.15c we can see that the majority of the participants have a bachelor's degree or graduate degree. These characteristics are not surprising because the questionnaire was mostly shared on groups that include a lot of international students. Moreover, the fact that the division between the education group and age group is uneven could result in that the results are not generalizable to the entire population, including the older generation without education. However, because the younger adults are the ones that use web shops the most, it makes sense to focus on this age group. Moreover, because the participants are from a lot of different countries, and there are participants from all age groups and educational groups in the survey, the results give a good indication on the opinions of web shop users around the world.

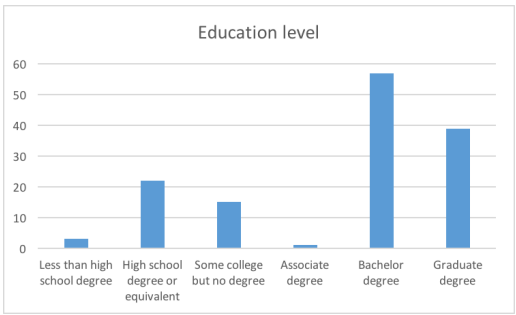
Figure 3.15d show how the participants are distributed into the clusters. This distribution is more even than the others, even though some clusters have a lot more participants than others. This is a bit unfortunate, and a more even distribution would be preferable, but at least all of the clusters have data from more than ten participants. The main reason why some clusters have more participants than others come from how the clusters are defined and how many countries that exist in each cluster. The definition of the Contest cluster and the Networking cluster is more strict in what values the countries can have than the rest of the clusters. Also, especially many countries in this research fall under the Pyramid cluster.



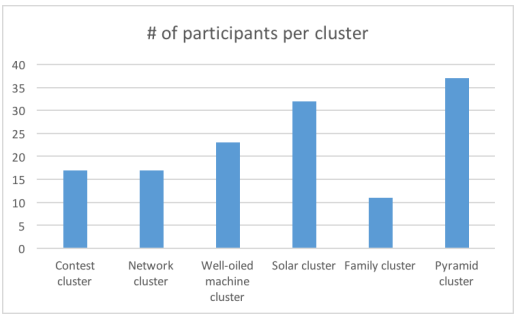
(a) Gender distribution



(b) Age distribution



(c) Education level distribution



(d) Cluster distribution

Figure 3.15: Distributions of the participants

4 Results

4.1 Reliability

To assess whether the questionnaire does, in fact, measure usability and aesthetics Cronbach's alpha was used as explained in Section 1.3.3 Cronbach's Alpha. Because the aesthetics and usability were expected to be different both between clusters and between designs, the alpha was calculated by using the answers from the participants of the Pyramid cluster on each design independently. This cluster was chosen because it has the most participants, which will give a more reliable result than any of the other clusters can.

Five questions measured the usability, and six questions measured the aesthetics. Both scales had a high level of internal consistency, with alphas that ranged from 0.866 to 0.940 for the usability and 0.815 to 0.953 for the aesthetics. These results show that the questionnaire is reliable.

Design	Usability alpha	Aesthetics alpha
Contest	0.889	0.922
Network	0.866	0.815
Well-Oiled Machine	0.940	0.935
Solar	0.939	0.950
Family	0.909	0.952
Pyramid	0.916	0.916

Table 4.1: Results for Cronbach's alpha

4.2 Aesthetics and Usability

Linear regression was used to investigate a possible relationship between users perception of aesthetics and usability. The regression was run once without distinguishing between the different designs, and on each design independently. Scatterplots of mean usability scores against mean aesthetics scores were plotted for all designs together, and each design individually. Visual inspection of these plots all indicated a linear relationship between the variables.

As mentioned in Section 1.3.3 Simple Linear Regression the Durbin-Watson statistics were used to assess whether or not there is an independence of observations. Because of the design of the survey, it is unlikely that the observations are not independent. The reason is both related to the data collection method, with participants living in many different countries, and to the fact that there has not been any interaction between the researcher and any of the participants about the questionnaire before their participation. However, the Durbin-Watson statistics also gives a good indication whether the observations are independent or not, so this is also used this to determine whether a simple linear regression would be appropriate to use.

Table 4.2 shows the Durbin-Watson statistics for every run of the linear regression. The statistic can range from 0 to 4, and the value one wants to come close to is 2. For our data, the values range from 1.628 to 2.026 which are all great values, and it indicates that the observations are independent.

Design	**Durbin-Watson Statistics**
All designs	1.628
Contest	1.870
Network	2.026
Well-Oiled Machine	1.890
Solar	1.816
Family	1.979
Pyramid	1.929

Table 4.2: Durbin-Watson statistics for each design

4.2.1 All designs

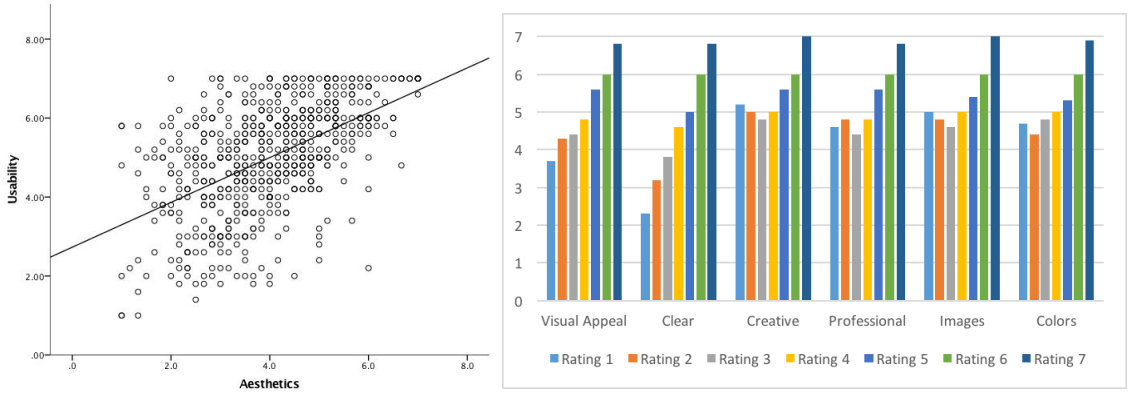
When the data from all the designs were combined, one outlier was found in the data set, but it was kept after assessing that there were minimal differences in the results with and without it. The prediction equation is: $usability\_score = 2.730 + (0.568 \times aesthetics\_score)$. The average aesthetics score accounted for 28.6% of the variation in the usability score with adjusted $R^2 = 28.5\%$, a medium size effect Field (2013). The average aesthetics score statistically significantly predicted the usability score, $F(1,662) = 265.459, p < .0005$.

Figure 4.1a presents a scatterplot that shows the results of the linear regression, by showing the dependent usability score, against the independent aesthetics score. Moreover, a line of best fit is also shown in the scatterplot.

Figure 4.1b shows the median of the usability score for every rating of every component. The median usability score for all the designs combined is 5.0. This figure clearly indicates that visual appeal and clarity of the design is related to the usability score because increasing ratings of these components result in a higher usability score. The clarity seems to have the most effect on the usability because low ratings of the clarity give a lower usability rating than any of the other components.

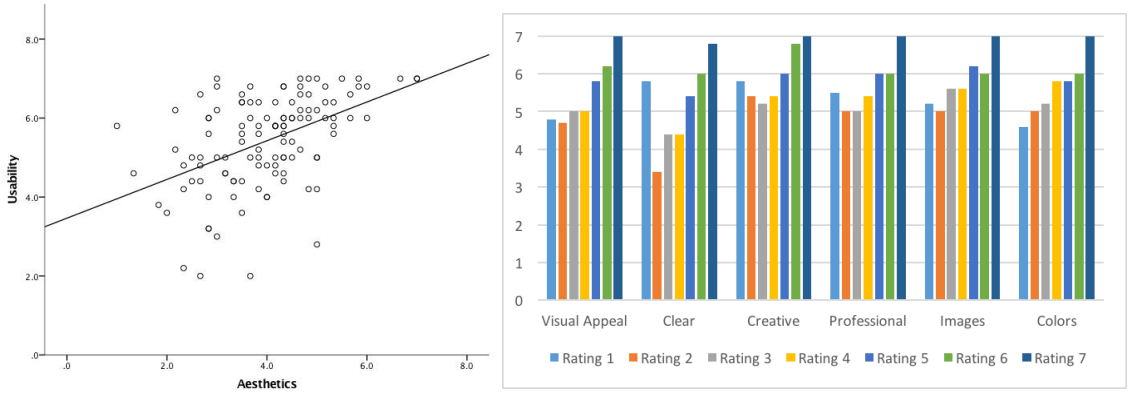
4.2.2 Contest Cluster Design

The data set for the Contest cluster design contained two outliers. After assessing whether there were any significant differences between running the regression with and without these values, the outliers were kept in the data because the changes in the results were minimal. The prediction equation is: $usability\_score = 3.464 + (0.491 \times aesthetics\_score)$. The average aesthetics score accounted for 22.7% of the variation in the usability score with adjusted $R^2 = 22.0\%$, a small to medium effect size (Field 2013). The average aesthetics score statistically significantly predicted the usability score, $F(1,113) = 33.240, p < .0005$.



(a) (b)

Figure 4.1: Relationship between usability and aeshtetics for all designs

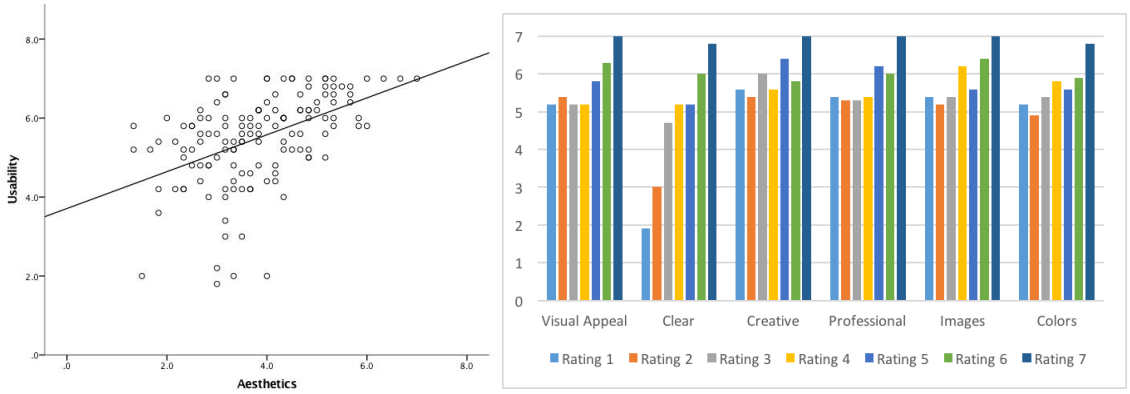


(a) (b)

Figure 4.2: Relationship between usability and aesthetics for the Contest cluster design

Figure 4.2a presents a scatterplot that shows the results of the linear regression, by showing the dependent usability score, against the independent aesthetics score. Moreover, a line of best fit is also shown in the scatterplot.

Figure 4.2b shows the median of the usability score for every rating of every component in the Contest design. The median usability score for all the components in this design is 5.8. Generally, for most of the ratings the usability score gets better as the rating increase, but mostly these differences are small. However, even though the differences are small, a pattern for how the usability score gets better when the participants like the aesthetics are easy to spot in the visual appeal, colours and images components. This pattern can be interpreted to mean that even though there is no clear relationship between low ratings and low usability scores, there is a relationship between high ratings and high usability scores. Also, the graph for the clarity component shows a relationship where the usability score increases quite much with growing ratings, except for the lowest rating which resulted in a quite high usability score.



(a) (b)

Figure 4.3: Relationship between usability and aeshtetics for the Network cluster design

4.2.3 Network Cluster Design

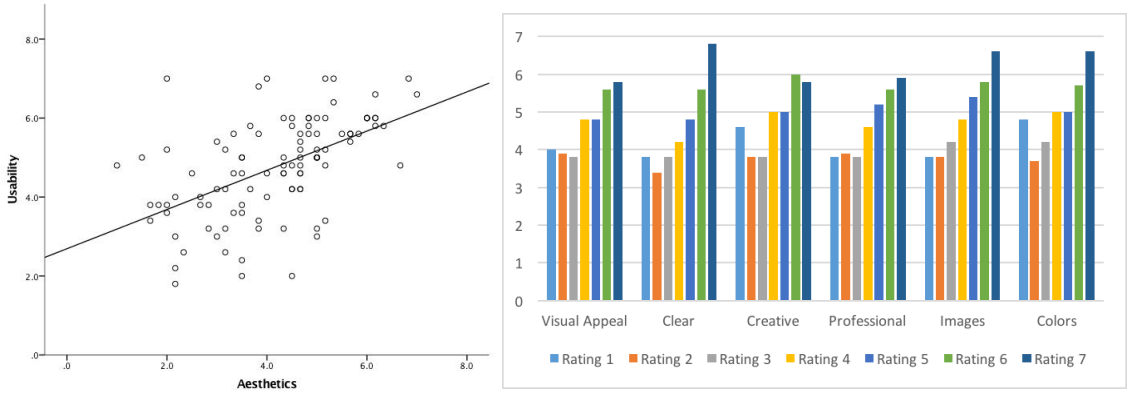
The data set of the Network cluster design contained three outliers. After assessing whether there were any significant differences between running the regression with and without these values, the outliers were kept because the changes in results were minimal. The prediction equation is: $usability\_score = 3.710 + (0.467 \times aesthetics\_score)$. The average aesthetics score accounted for 22.2% of the variation in the usability score with adjusted $R^2 = 21.7\%$, a small to medium effect size (Field 2013). The average aesthetics score statistically significantly predicted the usability score, $F(1,137) = 39.135, p < .0005$.

Figure 4.3a presents a scatterplot that shows the results of the linear regression, by showing the dependent usability score, against the independent aesthetics score. Moreover, a line of best fit is also shown in the scatterplot.

Figure 4.3b shows the median of the usability score for every rating of every component in the Network design. Overall for this design, the median usability score is 5.8. From the figure, we can see that for all the components, except the clarity component, the usability score is quite high independent of the rating. Also, the usability score increases with higher ratings of the components, but the differences between the scores for the different ratings are small. This result is not surprising as the median usability score is 5.8 and the linear regression shows that only 21.7% of the usability score is accounted for by the aesthetics ratings. However, some of the usability scores should be related to the aesthetics ratings, and this can be seen at the clarity component. Here the usability score increases with increasing ratings of this component and a low rating of the clarity component leads to a much lower usability score than a higher rating of the component. This pattern shows that for this design, the clarity of the design was the aesthetics component that had the biggest impact on how user-friendly the participants perceived the design to be.

4.2.4 Well-Oiled Machine Cluster Design

The data set for the Well-Oiled Machine cluster design contained one outlier, but it was kept after assessing that there were minimal differences in the results with and



(a) (b)

Figure 4.4: Relationship between usability and aeshtetics for the Well-Oiled Machine cluster design

without it. The prediction equation for this design is: $usability\_score = 2.692 + (0.496 \times aesthetics\_score)$. The average aesthetics score accounted for 28.5% of the variation in the usability score with adjusted $R^2 = 27.8\%$, a medium effect size (Field 2013). The average aesthetics score statistically significantly predicted the usability score, $F(1,100) = 39.838, p < .0005$.

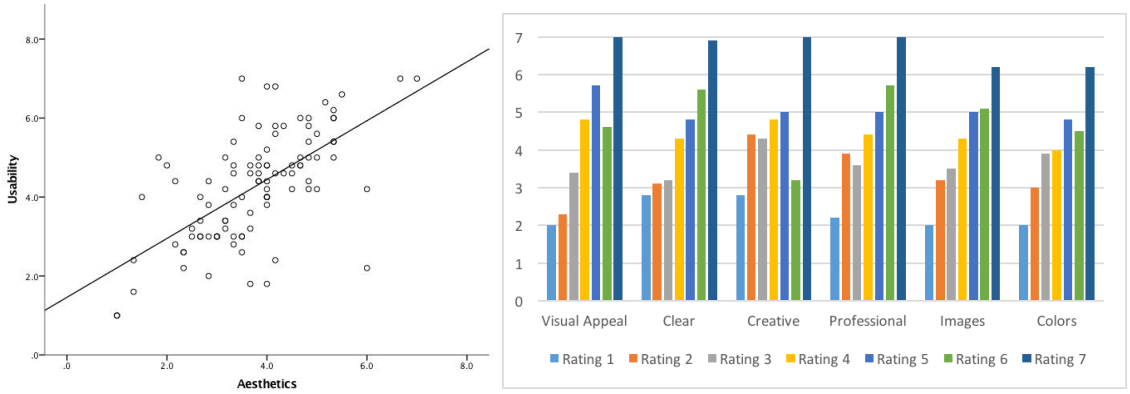
Figure 4.4a presents a scatterplot that shows the results of the linear regression, by showing the dependent usability score, against the independent aesthetics score. Moreover, a line of best fit is also shown in the scatterplot.

Figure 4.4b shows the median of the usability score for every rating of every component in the Well-Oiled Machine design. Overall the median usability score for this design is 4.9. This figure shows a quite similar relationship between all the different components and the usability score, with higher aesthetics ratings giving higher usability scores. One noticeable aspect of this figure is that only the higher ratings of the clarity, images, and colours components result in noticeable higher usability scores.

4.2.5 Solar Cluster Design

The data set for the Solar cluster design contained one outlier. After assessing whether there were any significant differences between running the regression with and without this value, the outlier was kept because the changes in the results were small. The prediction equation is: $usability\_score = 1.461 + (0.745 \times aesthetics\_score)$. The average aesthetics score accounted for 39.8% of the variation in the usability score with adjusted $R^2 = 39.1\%$, a medium effect size (Field 2013). The average aesthetics score statistically significantly predicted the usability score, $F(1,97) = 64.014, p < .0005$.

Figure 4.5a presents a scatterplot that shows the results of the linear regression, by showing the dependent usability score, against the independent aesthetics score. Moreover, a line of best fit is also shown in the scatterplot.



(a) (b)

Figure 4.5: Relationship between usability and aeshtetics for the Solar cluster design

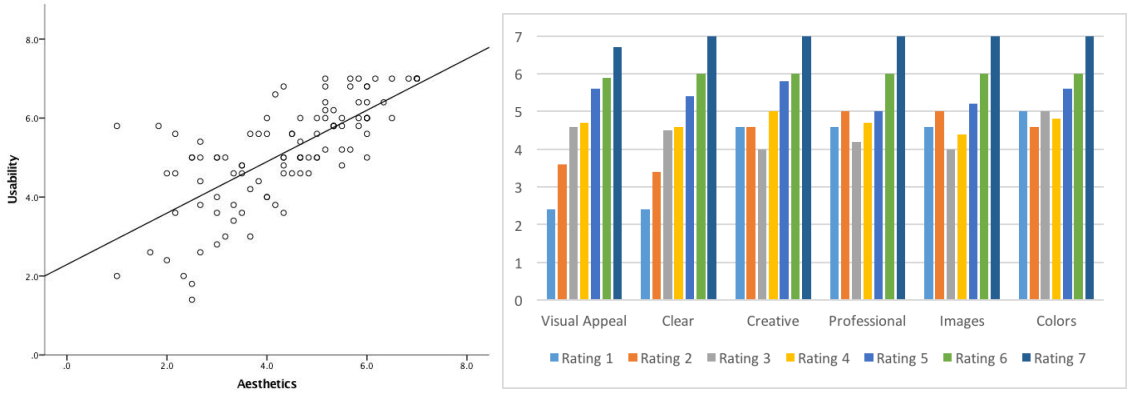
Figure 4.5b shows the median of the usability score for every rating of every component in the Solar design. The median usability score for this design is 4.4. So far, this is the design where the aesthetics have accounted for the biggest percentage of the usability score, which is also visible in the figure. Overall, low aesthetics ratings give low usability scores and as the aesthetics rating increase, so does the usability score. There are some deviations in the graphs, some aesthetics ratings that result in surprisingly low or high usability scores, but the general tendency shows a clearer relationship between the aesthetics ratings and the usability score in this design than in the previous designs. Moreover, in this design, all the components show a similar relationship which means that all the components seem to have some influence on the aesthetics score.

4.2.6 Family Cluster Design

The data set for the Family cluster design contained one outlier. After assessing whether there were any significant differences between running the regression with and without this value, the outlier was kept because the changes in the results were small. The prediction equation for this design is: $usability\_score = 2.290 + (0.652 \times aesthetics\_score)$. The average aesthetics score accounted for 51.5% of the variation in the usability score with adjusted $R^2 = 51.0\%$, a large effect size (Field 2013). The average aesthetics score statistically significantly predicted the usability score, $F(1, 101) = 107.187, p < .0005$.

Figure 4.6a presents a scatterplot that shows the results of the linear regression, by showing the dependent usability score, against the independent aesthetics score. Moreover, a line of best fit is also shown in the scatterplot.

Figure 4.6b shows the median of the usability score for every rating of every component in the Family design. This design is the one where the aesthetics score accounts for the greatest percentage of the usability score. Also, the median usability score for all the components combined is 5.2. From the figure, we can see that there is a clear relationship between the visual appeal and the clarity components, and the usability score. When the components are rated low, the score is also low, and when the rating increase so does the usability score. These two components are, in this design, the components that have the



(a) (b)

Figure 4.6: Relationship between usability and aeshtetics for the Family cluster design

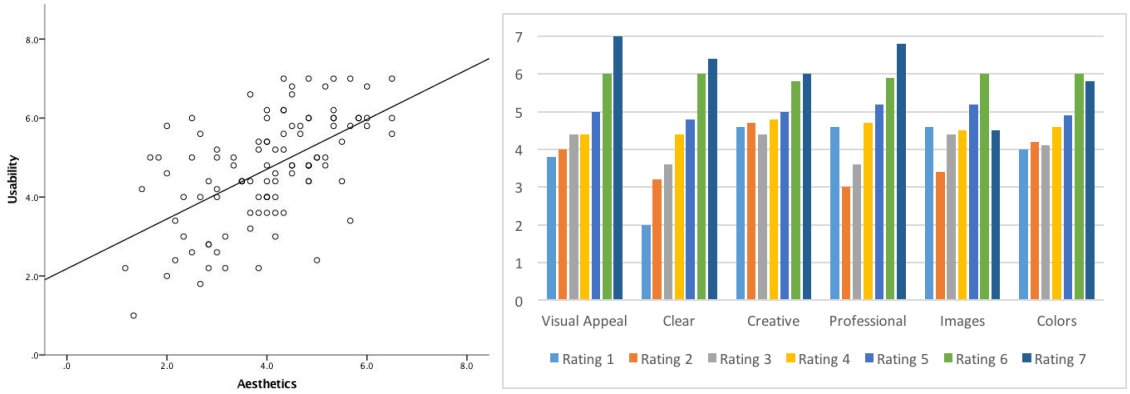
biggest influence on the usability and seems to be the main reason for why the aesthetics score accounts for over 50% of the usability score. For the other components a certain relationship is there, but this is mostly for the three highest ratings, as they give better usability scores increasingly. For lower ratings of these components, the usability score is about the same.

4.2.7 Pyramid Cluster Design

The data set of the Pyramid cluster design contained no outliers. The prediction equation for this design is: $usability\_score = 2.185 + (0.630 \times aesthetics\_score)$. The average aesthetics score accounted for 32.3% of the variation in the usability score with adjusted $R^2 = 31.7\%$, a medium effect size (Field 2013). The average aesthetics score statistically significantly predicted the usability score, $F(1,104) = 49.632, p < .0005$.

Figure 4.7a presents a scatterplot that shows the results of the linear regression, by showing the dependent usability score, against the independent aesthetics score. Moreover, a line of best fit is also shown in the scatterplot.

Figure 4.7b shows the median of the usability score for every rating of every component in the Pyramid design. The median usability score for all the aesthetics components in this design is 4.8. The two components that show a quite clear and positive relationship between the ratings of the aesthetics components and the usability scores are the visual appeal and clarity components. The visual appeal component shows a particularly strong relationship for higher ratings while the clarity shows a stronger relationship for the lower ratings. This relationship implies that the clarity is necessary to get a decent usability score and that the visual appeal is more important to get an even better usability score. For the professional component, the relationship is quite similar to the clarity component, except for the lowest rating that has a quite unexpectedly high usability score. Also, for the rest of the components, the usability score seem to be somewhat related to the ratings, as lower ratings give lower usability scores than higher ratings, but the differences in the usability scores are small.



(a) (b)

Figure 4.7: Relationship between usability and aeshtetics for the Pyramid cluster design

4.3 Cultural Differences

Kruskal-Wallis H test was used to investigate the second and third research questions. These research questions are about whether users prefer the designs developed for them, and if there are any differences in the preferences of the different clusters. The test was run on each design separately, and then on each cluster separately to investigate these questions. To determine whether the distribution of scores for each group of the independent variable have the same shape, or a different shape visual inspection of boxplots was performed. Appendix G show all the boxplots that were inspected. Because not all distributions were similar and it is easier to use the same type of values to illustrate the differences in all the cases, the mean ranks will be used for all the test runs.

4.3.1 Design by Design

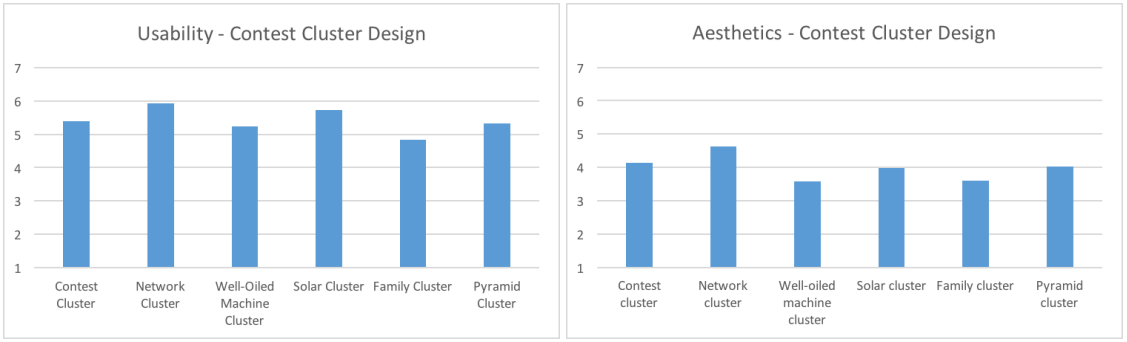
In this part, the Kruskal-Wallis H test is run on each design separately. The objective of these tests is to see if there are any statistically significant differences between the usability and aesthetics scores given by each cluster on the designs.

Contest Cluster Design

Figure 4.8 show the mean usability and aesthetics scores for the Contest cluster design. In both of these diagrams, one can see that there are differences between the scores, but the differences are small. The Kruskal-Wallis H test, which shows that there are differences, but the differences are not statistically significant, confirms these results with $\chi^2(5) = 10.197, p = 0.070$ for the usability scores and $\chi^2(5) = 9.674, p = 0.085$ for the aesthetics scores.

Even though the differences are not significant, both the mean scores and the results from the Kruskal-Wallis H test can be used to decide which clusters that think the design is most usable or beautiful. Table 4.3 show a summary of the results of the Kruskal-Wallis H test performed on the data from the Contest cluster design.

Figure 4.8a and the usability results in Table 4.3 shows that the usability scores increased from the Family cluster to the Well-Oiled Machine cluster, to the Pyramid cluster, to the Contest cluster, to the Solar cluster, to the Network cluster. Figure 4.8b and the aesthetics results in Table 4.3 shows that aesthetics scores increased from the Well-Oiled Machine cluster to the Family cluster, to the Pyramid cluster, to the Solar cluster, to the Contest cluster, to the Network cluster.



(a) Mean usability scores

(b) Mean aesthetics scores

Figure 4.8: Usability and aesthetics scores for the Contest cluster design

Cluster	Usability results	Aesthetics results	N
Contest	56.58	59.78	18
Network	72.63	77.93	15
Well-Oiled Machine	53.14	42.39	14
Solar	67.19	58.35	26
Family	35.06	46.00	9
Pyramid	53.20	57.59	33

Table 4.3: Kruskal-Wallis T test results for the Contest cluster design

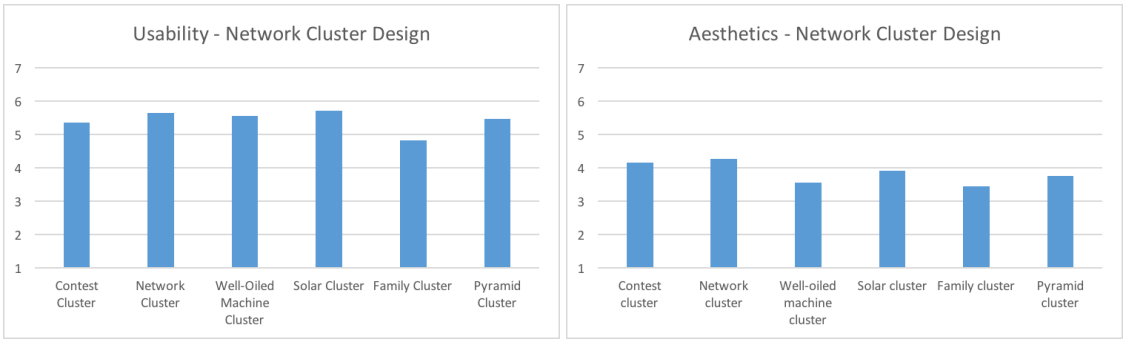
Network Cluster Design

Figure 4.9 show the mean usability and aesthetics scores for the Network cluster design. In both of these diagrams, one can see that there are differences between the scores, but the differences are small. The Kruskal-Wallis H test, which shows that there are differences, but the differences are not statistically significant, confirms these results with $\chi^2(5) = 7.666, p = 0.176$ for the usability scores and $\chi^2(5) = 7.913, p = 0.161$ for the aesthetics scores.

Even though the differences are not significant, both the mean scores and the results from the Kruskal-Wallis H test can be used to decide which clusters that perceive the design as most usable or beautiful. Table 4.4 show a summary of the results of the Kruskal-Wallis H test performed on the data from the Network cluster design.

Figure 4.9a and the usability results in Table 4.4 shows that the usability scores increased from the Family cluster to the Contest cluster, to the Pyramid cluster, to the Well-Oiled Machine cluster, to the Network cluster, to the Solar cluster. Figure 4.8b and the aesthetics

results in Table 4.3 shows that aesthetics scores increased from the Family cluster to the Well-Oiled Machine cluster, to the Pyramid cluster, to the Solar cluster, to the Contest cluster, to the Network cluster. In other words, the participants from the Network cluster perceived this design to be more user-friendly and beautiful than the participants from the other clusters.



(a) Usability scores

(b) Aesthetics scores

Figure 4.9: Usability and aesthetics scores for the Network cluster design

Cluster	Usability results	Aesthetics results	N
Contest	62.47	80.06	18
Network	77.12	85.21	17
Well-Oiled Machine	70.46	56.80	23
Solar	79.39	72.27	32
Family	44.18	55.27	11
Pyramid	67.80	66.88	37

Table 4.4: Kruskal-Wallis T test results for the Network cluster design

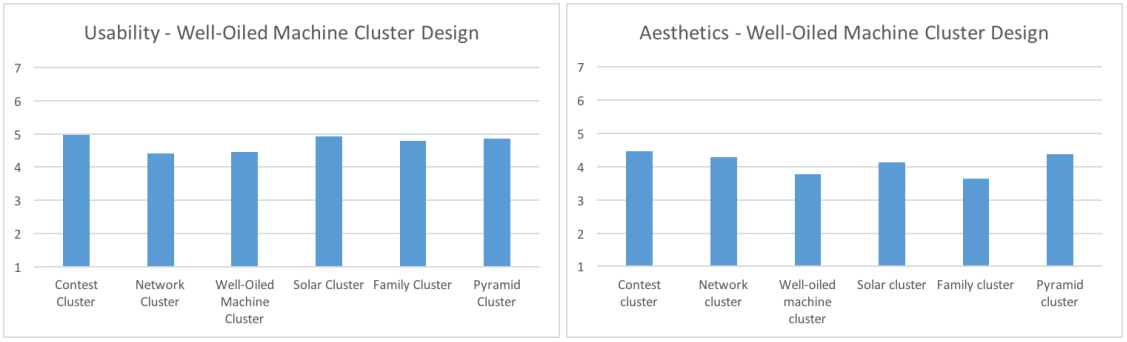
Well-Oiled Machine Cluster Design

Figure 4.10 show the mean usability and aesthetics scores for the Well-Oiled Machine cluster design. In both of these diagrams, one can see that there are differences between the scores, but the differences are small. The Kruskal-Wallis H test, which shows that there are differences, but the differences are not statistically significant, confirms these results with $\chi^2(5) = 3.044, p = 0.693$ for the usability scores and $\chi^2(5) = 3.818, p = 0.576$ for the aesthetics scores.

Even though the differences are not significant, both the mean scores and the results from the Kruskal-Wallis H test can be used to decide which clusters that perceive the design as most usable or beautiful. Table 4.5 show a summary of the results of the Kruskal-Wallis H test performed on the data from the Well-Oiled Machine cluster design.

Figure 4.10a and the usability results in Table 4.5 shows that the usability scores increased from the Network cluster to the Well-Oiled Machine cluster, to the Family cluster, to the Solar cluster, to the Pyramid cluster, to the Contest cluster. Figure 4.10b and the aesthetics results in Table 4.5 shows that aesthetics scores increased from the Family cluster to the

Well-Oiled Machine cluster, to the Solar cluster, to the Network cluster, to the Pyramid cluster, to the Contest cluster.



(a) Usability scores

(b) Aesthetics scores

Figure 4.10: Usability and aesthetics scores for the Well-Oiled Machine cluster design

Cluster	Usability results	Aesthetics results	N
Contest	56.38	57.97	16
Network	41.93	52.61	14
Well-Oiled Machine	45.14	44.41	11
Solar	54.87	48.89	23
Family	51.06	40.56	9
Pyramid	54.97	57.13	30

Table 4.5: Kruskal-Wallis T test results for the Well-Oiled Machine cluster design

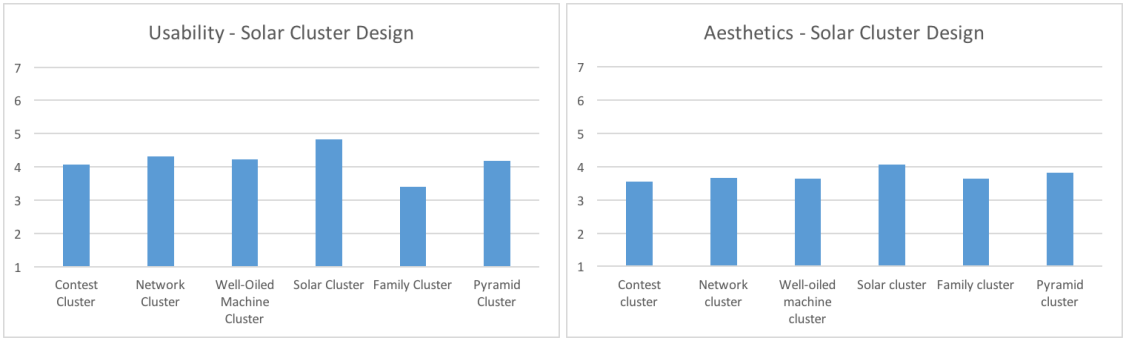
Solar Cluster Design

Figure 4.11 show the mean usability and aesthetics scores for the Solar cluster design. In both of these diagrams, one can see that there are differences between the scores, but the differences are small. The Kruskal-Wallis H test, which shows that there are differences, but the differences are not statistically significant, confirms these results with $\chi^2(5) = 8.413, p = 0.135$ for the usability scores and $\chi^2(5) = 2.926, p = 0.711$ for the aesthetics scores.

Even though the differences are not significant, both the mean scores and the results from the Kruskal-Wallis H test can be used to decide which clusters that think the design is most usable or beautiful. Table 4.6 show a summary of the results of the Kruskal-Wallis H test performed on the data from the Solar cluster design.

Figure 4.11a and the usability results in Table 4.6 shows that the usability scores increased from the Family cluster to the Contest cluster, to the Network cluster, to the Well-Oiled Machine cluster, to the Pyramid cluster, to the Solar cluster. Figure 4.11b and the aesthetics results in Table 4.6 shows that aesthetics scores increased from the Contest cluster to the Family cluster, to the Network cluster, to the Well-Oiled Machine cluster, to the Pyramid cluster, to the Solar cluster. For this design, even though the differences are not statistically

significant, it is worth to notice that the Solar cluster regards its design as being the most beautiful and user-friendly of all the clusters.



(a) Usability scores

(b) Aesthetics scores

Figure 4.11: Usability and aesthetics scores for the Solar cluster design

Cluster	Usability results	Aesthetics results	N
Contest	45.70	43.50	15
Network	49.35	46.69	13
Well-Oiled Machine	49.73	48.36	11
Solar	62.61	57.70	23
Family	30.19	45.50	8
Pyramid	49.82	52.25	30

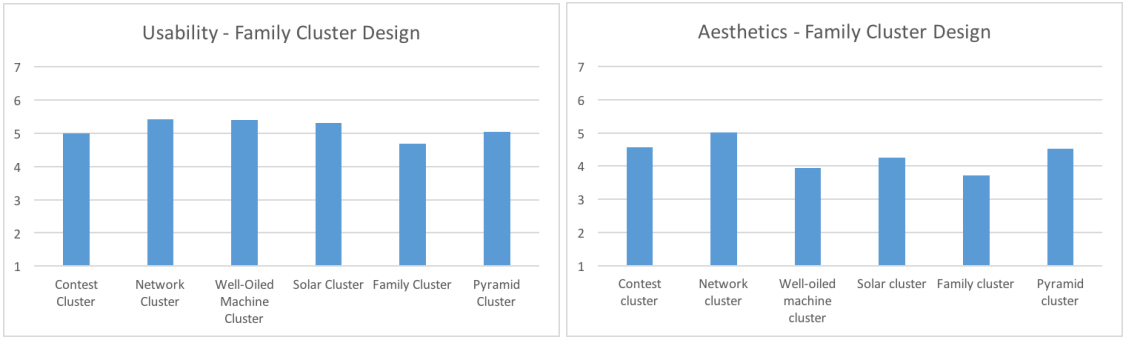
Table 4.6: Kruskal-Wallis T test results for the Solar cluster design

Family Cluster Design

Figure 4.12 show the mean usability and aesthetics scores for the Family cluster design. In both of these diagrams, one can see that there are differences between the scores, but the differences are small. The Kruskal-Wallis H test, which shows that there are differences, but the differences are not statistically significant, confirms these results with $\chi^2(5) = 4.177, p = 0.524$ for the usability scores and $\chi^2(5) = 5.902, p = 0.316$ for the aesthetics scores.

Even though the differences are not significant, both the mean scores and the results from the Kruskal-Wallis H test can be used to decide which clusters that think the design is most usable or beautiful. Table 4.7 show a summary of the results for the Kruskal-Wallis H test performed on the data from the Family cluster design.

Figure 4.12a and the usability results in Table 4.7 shows that the usability scores increased from the Family cluster to the Contest cluster, to the Pyramid cluster, to the Well-Oiled Machine cluster, to the Network cluster, to the Solar cluster. Figure 4.12b and the aesthetics results in Table 4.7 shows that aesthetics scores increased from the Family cluster to the Well-Oiled Machine cluster, to the Solar cluster, to the Network cluster, to the Pyramid cluster, to the Contest cluster.



(a) Usability scores

(b) Aesthetics scores

Figure 4.12: Usability and aesthetics scores for the Family cluster design

Cluster	Usability results	Aesthetics results	N
Contest	47.78	55.31	16
Network	57.46	64.43	14
Well-Oiled Machine	57.04	42.25	12
Solar	58.43	49.37	23
Family	37.89	39.83	9
Pyramid	50.72	55.73	30

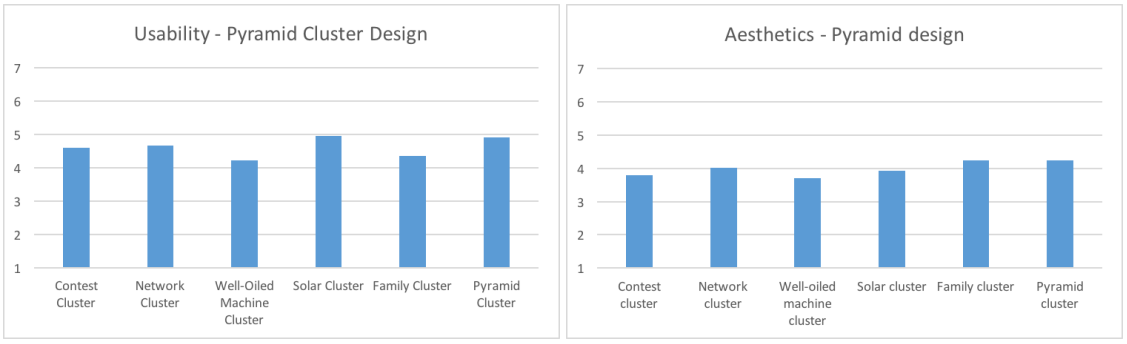
Table 4.7: Kruskal-Wallis T test results for the Family cluster design

Pyramid Cluster Design

Figure 4.13 show the mean usability and aesthetics scores for the Pyramid cluster design. In both of these diagrams, one can see that there are differences between the scores, but the differences are small. The Kruskal-Wallis H test, which shows that there are differences, but the differences are not statistically significant, confirms these results with $\chi^2(5) = 4.482, p = 0.482$ for the usability scores and $\chi^2(5) = 3.795, p = 0.579$ for the aesthetics scores.

Even though the differences are not significant, both the mean scores and the results from the Kruskal-Wallis H test can be used to decide which clusters that think the design is most usable or beautiful. Table 4.8 show a summary of the results for the Kruskal-Wallis H test performed on the data from the Pyramid cluster design.

Figure 4.13a and the usability results in Table 4.8 shows that the usability scores increased from the Family cluster to the Well-Oiled Machine cluster, to the Network cluster, to the Contest cluster, to the Solar cluster, to the Pyramid cluster. Figure 4.13b and the aesthetics results in Table 4.8 shows that aesthetics scores increased from the Well-Oiled Machine cluster to the Contest cluster, to the Network cluster, to the Solar cluster, to the Family cluster, to the Pyramid cluster.



(a) Usability scores

(b) Aesthetics scores

Figure 4.13: Usability and aesthetics scores for the Pyramid cluster design

Cluster	Usability results	Aesthetics results	N
Contest	51.12	49.69	17
Network	50.64	50.18	14
Well-Oiled Machine	45.04	46.27	13
Solar	59.54	51.46	24
Family	42.78	60.06	9
Pyramid	60.02	61.85	30

Table 4.8: Kruskal-Wallis T test results for the Pyramid cluster design

4.3.2 Cluster by Cluster

In this part, the Kruskal-Wallis H test was run on each cluster separately, with the different designs as independent variables. The objective of these tests is to see if there are any statistically significant differences between the usability scores and aesthetics scores given to each design and to figure out which designs each cluster likes the most and the least.

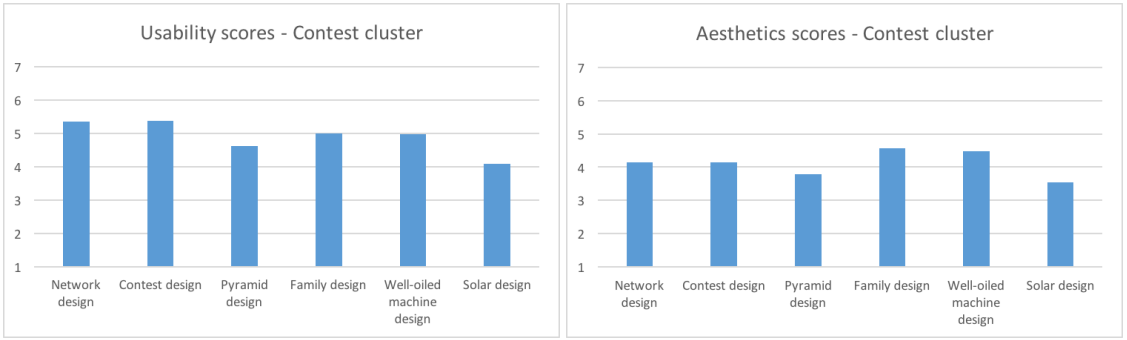
Contest Cluster

Figure 4.14 show the mean usability and aesthetics scores for the Contest cluster. These diagrams show that there are differences between which designs the participants from the Contest cluster thought were most user-friendly and beautiful, but the differences in the aesthetics scores are small. The Kruskal-Wallis H test, which shows that there are statistically significant differences between the usability scores, $\chi^2(5) = 12.380, p = 0.030$, but not between the aesthetics scores, $\chi^2(5) = 6.288, p = 0.279$, confirms these results.

Table 4.9 presents a summary of the results for the Kruskal-Wallis H test performed on the data from the Contest cluster. Figure 4.14a and the usability results in Table 4.9 shows that the usability scores increased from the Solar design to the Pyramid design, to the Well-Oiled Machine design, to the Family design, to the Contest design, to the Network design. Figure 4.14b and the aesthetics results in Table 4.9 shows that aesthetics

scores increased from the Solar design to the Pyramid design, to the Contest design, to the Network design, to the Family design, to the Well-Oiled Machine design.

To investigate which designs that had statistically significantly different usability scores the post hoc analysis were performed. In this case, the post hoc analysis revealed statistically significant differences in usability scores between the Solar design and the Network design ($p = 0.049$), but not any other design combinations.



(a) Usability scores

(b) Aesthetics scores

Figure 4.14: Usability and aesthetics scores for the Contest cluster

Design	**Usability results**	**Aesthetics results**	**N**
Contest	60.31	50.81	18
Network	60.53	52.17	18
Well-Oiled Machine	50.81	57.38	16
Solar	30.77	37.80	15
Family	53.69	59.88	16
Pyramid	43.62	44.32	17

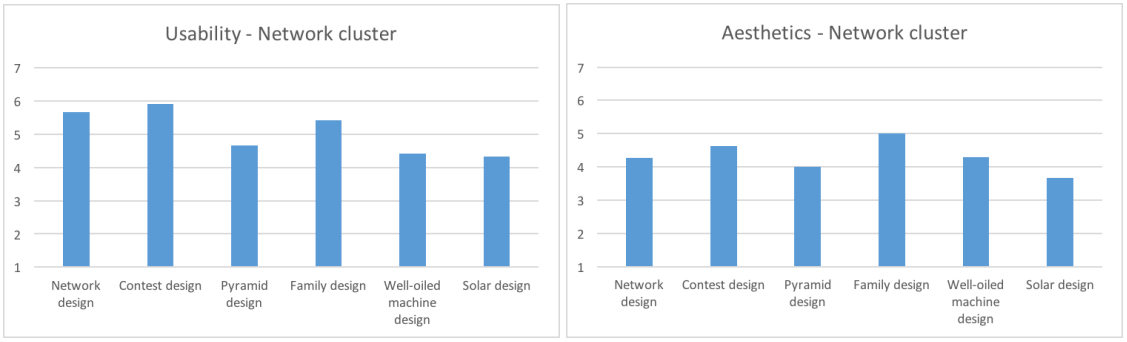
Table 4.9: Kruskal-Wallis T test results for the Contest cluster

Network Cluster

Figure 4.15 show the mean usability and aesthetics scores for the Network cluster. These diagrams show that there are differences between which designs the participants from the Network cluster thought were most user-friendly and beautiful. The Kruskal-Wallis H test, which shows that there are statistically significant differences both between the usability scores, $\chi^2(5) = 22.951, p < 0.001$, and between the aesthetics scores, $\chi^2(5) = 14.996, p = 0.010$, confirms these results.

Table 4.10 show a summary of the results for the Kruskal-Wallis H test performed on the data from the Network cluster. Figure 4.8a and the usability results in Table 4.3 shows that the usability scores increased from the Solar design to the Well-Oiled Machine design, to the Pyramid design, to the Family design, to the Network design, to the Contest design. Figure 4.8b and the aesthetics results in Table 4.3 shows that aesthetics scores increased from the Solar design to the Pyramid design, to the Network design, to the Well-Oiled Machine design, to the Contest design, to the Family design.

To investigate which designs that had statistically significantly different usability or aesthetics scores the post hoc analysis were performed. This analysis revealed differences in usability scores between the Solar and Contest designs ($p = 0.008$), the Solar and Network designs ($p = 0.046$), and the Well-Oiled Machine and Contest designs ($p = 0.011$).For the aesthetics scores, there were differences only between the Solar and Family designs ($p = 0.013$).



(a) Usability scores

(b) Aesthetics scores

Figure 4.15: Usability and aesthetics scores for the Network cluster

Design	Usability results	Aesthetics results	N
Contest	61.17	53.23	15
Network	55.59	43.09	17
Well-Oiled Machine	29.50	43.79	14
Solar	28.08	27.69	13
Family	49.89	59.93	14
Pyramid	34.93	34.64	14

Table 4.10: Kruskal-Wallis T test results for the Network cluster

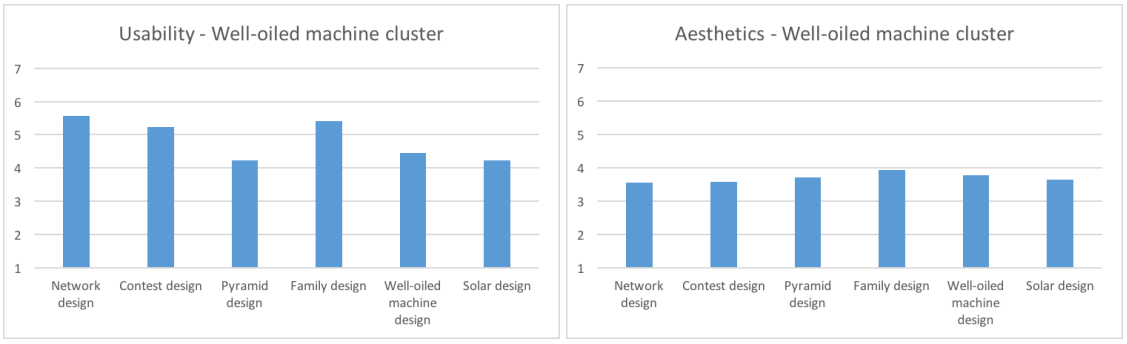
Well-Oiled Machine Cluster

Figure 4.16 show the mean usability and aesthetics scores for the Well-Oiled Machine cluster. In both of these diagrams, one can see that there are differences between the scores, but the differences are small. The Kruskal-Wallis H test, which shows that there are differences, but the differences are not statistically significant, confirms these results with $\chi^2(5) = 10.280, p = 0.068$ for the usability scores and $\chi^2(5) = 0.793, p = 0.977$ for the aesthetics scores.

Even though the differences are not significant, both the mean scores and the results from the Kruskal-Wallis H test can be used to decide which designs the cluster think is most usable or beautiful. Table 4.11 show a summary of the results of the Kruskal-Wallis H test performed on the data from the Well-Oiled Machine cluster design.

Figure 4.16a and the usability results in Table 4.11 shows that the usability scores increased from the Pyramid design to the Solar design, to the Well-Oiled Machine design, to the Contest design, to the Family design, to the Network design. Figure 4.16b and the aesthetics

results in Table 4.11 shows that aesthetics scores increased from the Network design to the Contest design, to the Solar design, to the Pyramid design, to the Well-Oiled Machine design, to the Family design.



(a) Usability scores

(b) Aesthetics scores

Figure 4.16: Usability and aesthetics scores for the Well-Oiled Machine cluster

Design	**Usability results**	**Aesthetics results**	**N**
Contest	46.68	41.04	14
Network	51.85	39.50	23
Well-Oiled Machine	34.41	44.73	11
Solar	32.27	42.59	11
Family	48.33	45.88	12
Pyramid	31.58	44.31	13

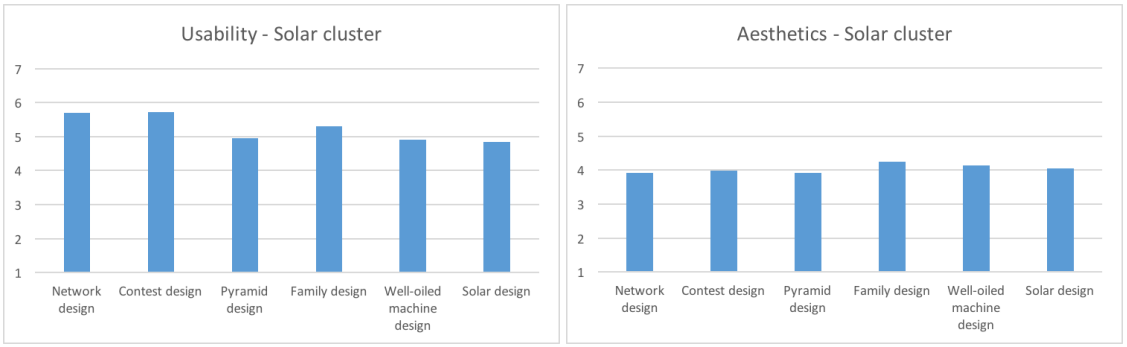
Table 4.11: Kruskal-Wallis T test results for the Well-Oiled Machine cluster

Solar Cluster

Figure 4.17 show the mean usability and aesthetics scores for the Solar cluster. These diagrams show that there are differences between which designs the participants from the Solar cluster thought were most user-friendly and beautiful, but the differences in the aesthetics scores are small. The Kruskal-Wallis H test, which shows that there are statistically significant differences between the usability scores, $\chi^2(5) = 17.228, p = 0.004$, but not between the aesthetics scores, $\chi^2(5) = 1.222, p = 0.943$ confirms these results.

Table 4.12 show a summary of the results of the Kruskal-Wallis H test performed on the data from the Solar cluster. Figure 4.17a and the usability results in Table 4.12 shows that the usability scores increased from the Well-Oiled Machine design to the Solar design, to the Pyramid design, to the Family design, to the Contest design, to the Network design. Figure 4.17b and the aesthetics results in Table 4.12 shows that aesthetics scores increased from the Pyramid design to the Solar design, to the Contest design, to the Network design, to the Well-Oiled Machine design, to the Family design.

To investigate which designs that had statistically significantly different usability scores the post hoc analysis were performed. In this case, the post hoc analysis revealed no statistically significant differences in the usability scores between any design combination.



(a) Usability scores

(b) Aesthetics scores

Figure 4.17: Usability and aesthetics scores for the Solar cluster

Design	Usability results	Aesthetics results	N
Contest	92.58	74.02	26
Network	92.81	74.44	32
Well-Oiled Machine	58.91	78.63	23
Solar	59.57	73.50	23
Family	79.76	84.13	23
Pyramid	64.15	72.31	24

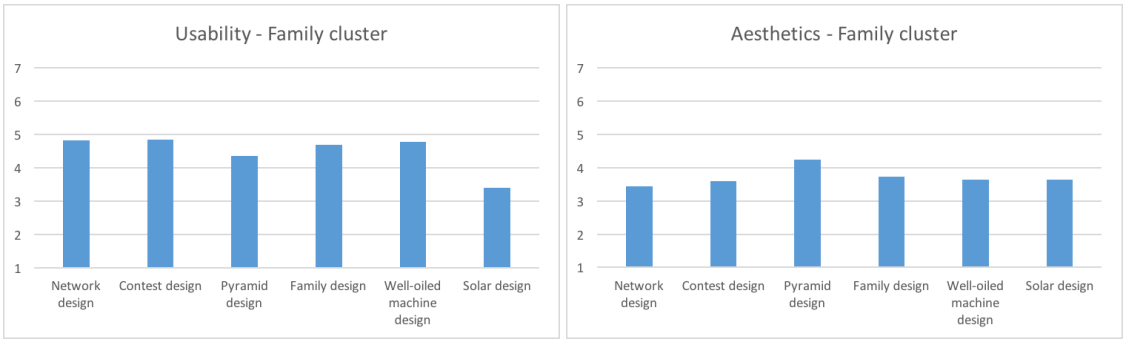
Table 4.12: Kruskal-Wallis T test results for the Solar cluster

Family Cluster

Figure 4.18 show the mean usability and aesthetics scores for the Family cluster. In both of these diagrams, one can see that there are differences between the scores, but the differences are small. The Kruskal-Wallis H test, which shows that there are differences, but the differences are not statistically significant, confirms these results with $\chi^2(5) = 10.544, p = 0.061$ for the usability scores and $\chi^2(5) = 3.258, p = 0.660$ for the aesthetics scores.

Even though the differences are not significant, both the mean scores and the results from the Kruskal-Wallis H test can be used to decide which cdesigns the cluster think is most usable or beautiful. Table 4.13 show a summary of the results of the Kruskal-Wallis H test performed on the data from the Contest cluster design.

Figure 4.18a and the usability results in Table 4.13 shows that the usability scores increased from the Solar design to the Pyramid design, to the Family design, to the Network design, to the Well-Oiled Machine and Contest designs. Figure 4.18b and the aesthetics results in Table 4.13 shows that aesthetics scores increased from the Network design to the Solar design, to the Contest design, to the Family design, to the Well-Oiled Machine design, to the Pyramid design.



(a) Usability scores

(b) Aesthetics scores

Figure 4.18: Usability and aesthetics scores for the Family cluster

Design	Usability results	Aesthetics results	N
Contest	33.50	27.33	9
Network	31.73	23.00	11
Well-Oiled Machine	33.50	29.44	9
Solar	13.00	25.88	8
Family	29.94	27.78	9
Pyramid	23.83	35.44	9

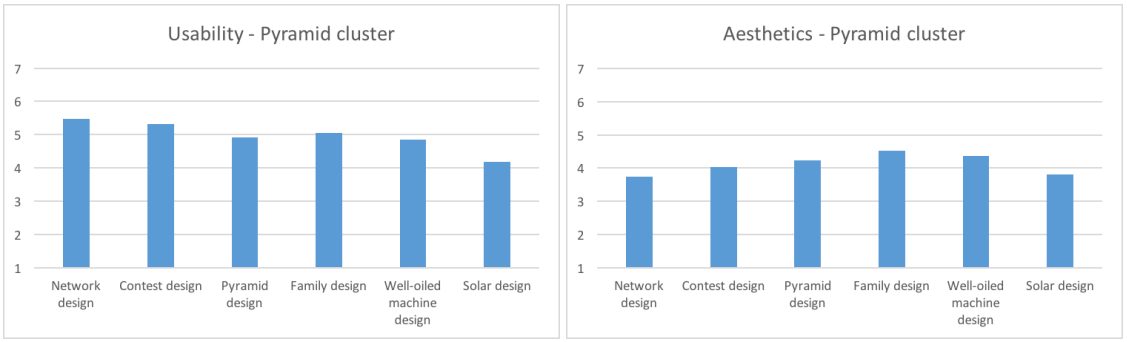
Table 4.13: Kruskal-Wallis T test results for the Family cluster

Pyramid Cluster

Figure 4.19 show the mean usability and aesthetics scores for the Pyramid cluster. These diagrams show that there are differences between which designs the participants from the Pyramid cluster thought were most user-friendly and beautiful, but the differences in the aesthetics scores are small. The Kruskal-Wallis H test, which shows that there are statistically significant differences between the usability scores, $\chi^2(5) = 16.729, p = 0.005$, but not between the aesthetics scores, $\chi^2(5) = 9.715, p = 0.084$ confirms these results.

Table 4.14 show a summary of the results for the Kruskal-Wallis H test performed on the data from the Pyramid cluster. Figure 4.19a and the usability results in Table 4.14 shows that the usability scores increased from the Solar design to the Well-Oiled Machine design, to the Pyramid design, to the Family design, to the Contest design, to the Network design. Figure 4.19b and the aesthetics results in Table 4.14 shows that aesthetics scores increased from the Network design to the Solar design, to the Contest design, to the Pyramid design, to the Well-Oiled Machine design, to the Family design.

To investigate which designs that had statistically significantly different usability scores the post hoc analysis were performed. In this case, the post hoc analysis revealed statistically significant differences in median usability scores between the Solar and Contest designs ($p = 0.023$), and the Solar and Network designs ($p = 0.002$), but not between any other design combinations.



(a) Usability scores

(b) Aesthetics scores

Figure 4.19: Usability and aesthetics scores for the Pyramid cluster

Design	**Usability results**	**Aesthetics results**	**N**
Contest	107.89	91.05	33
Network	115.05	78.51	37
Well-Oiled Machine	89.08	106.33	30
Solar	64.03	84.65	30
Family	97.57	112.32	30
Pyramid	93.57	104.55	30

Table 4.14: Kruskal-Wallis T test results for the Pyramid cluster

III Discussion, Conclusion and Future Research

5 Discussion

5.1 Aesthetics and Usability

The first research question in this thesis is about whether the aesthetics of a design affect the perceived usability of the system. Results from earlier research have both been able and not able to find such a link between aesthetics and usability. These results make it hard to decide whether or not to regard aesthetics as a part of the usability in the design process, and it is, therefore, interesting to investigate this even further.

The linear regression shows that, when the aesthetics scores go up, the usability scores also increases. Also, the effect aesthetics has on usability seems to be more accurate for the high aesthetics ratings. Therefore, the participants that liked the look of a design a lot also perceived the design as user-friendly. For lower aesthetics ratings the usability scores are more varied, but in many cases, the usability score does not suffer significantly from lower aesthetics ratings. This trend gives an indication that aesthetics can be used to provide the user with a sense of whether the design is user-friendly, but that this is mostly when the user likes the aesthetics of the design. Moreover, even though designs that are regarded beautiful are also seen as user-friendly, when the user does not like the aesthetics, they do not seem to automatically think that the system seems difficult to use.

Overall, the aesthetics score is expected to account for 28.5% of the variation of the usability score in the population. This relationship indicates that there is a link between the aesthetics and the usability when looking at all the designs combined, but this link is not very strong. Also, when having a closer look at each design individually, the results all show a statistically significant prediction of the usability score, but the percentage varies. For the Contest cluster design and Network cluster design the aesthetics score is expected to account for respectively 22.0% and 21.7% of the variation of the usability score in the population. These relationships indicate that the link between the aesthetics and the usability in these designs are not very strong. However, for the Pyramid cluster design, Well-Oiled Machine cluster design, Solar cluster design and Family cluster design the relationship is stronger with respectively 31.7%, 27.8%, 39.1% and 51.0%.

When analysing which aesthetics components that affect the usability score, the different designs all showed signs to that most of the components had a relationship that suggested that low ratings of the component resulted in lower usability scores and that the usability score increased with higher ratings of the components. All over, the visual appeal and clarity components showed a more significant relationship with the usability score than the other components, but how strong the relationships was and how the ratings affected the usability score, varied between the designs. The colours and images components also had a noticeable relationship with the usability score, but these relationships were not as strong as for the visual appeal and clarity. For the two last components, the creative and professional components, the relationship with the usability score varied more across the designs, and it is, therefore, hard to generalise how the usability might be linked to these components, and how to use them to achieve more user-friendly systems.

Even though the aesthetics can be said to be related to usability, and that it is important to design an aesthetically pleasing system to give the users a good first impression and a good

experience when looking at it, it is hard to anticipate how a system should be designed to get this effect. For the web shops in this study, each design had different relationships between the aesthetics rating and the usability score. Also, how the various components were related to the usability score was different for the different designs, without any obvious reason. Therefore, it was difficult to generate any guidelines for what to do or not to when designing a system because there are no clear results that were true for all the designs. What this could mean is that the overall perception of the aesthetics is more important than each component individually and that the composition of the different components is more important than each component individually.

From the results, it seems that the aesthetics of a design does, in fact, influence the perceived user-friendliness, but the relationships found in this study are not strong enough to only rely on the aesthetics when designing a user-friendly system. However, it is important to care about the aesthetics when designing a system because the aesthetics are what makes the first impression of a system. Therefore, what the system look like can give the users a positive or negative attitude towards the system, even before they have used the system. This attitude can then affect what the user think about the user friendliness even after having used the system for a while. In this research, this link seemed most relevant when the user liked the aesthetics of the system because high aesthetics ratings led to a relatively high or high usability score for all the components in all the designs.

5.2 Cultural Differences

The second and third research questions are about whether a person's cultural background influences what they think of the aesthetics and the usability of a system. This issue includes both whether people from different clusters prefer different designs, and whether users like the culturally adapted design better than they like the other designs.

When investigating whether participants from different clusters prefer different designs, the results showed that no statistically significant differences existed. In other words, people had the same opinion about all the designs, regardless of where they came from. This result suggests that a participant's cultural background does not have an impact on what designs they prefer, but that good or bad design is the same for people regardless of where they are from. However, this result does not say anything about whether participants from every cluster like some designs better than other.

When looking at how the different clusters rank the designs, some clusters have statistically significant differences between the rankings, but these differences are always between the top one and bottom one, and once also between the top two and bottom two ranks. In other words, the differences between each rank are not big enough to be significant in any cases, but some similarities or differences between the clusters opinions still exist.

How the designs are ranked according to their usability is very similar between the designs. All the clusters except the Family cluster, agree that the Solar cluster design, Pyramid cluster design and the Well-Oiled Machine cluster design are the least user-friendly and that the Family cluster design, Contest cluster design, and Network cluster design are the most user-friendly. For the Family cluster, the only difference is that the Family cluster

design and the Well-Oiled Machine cluster design has switched groups. Because there are no statistically significant differences between any designs within the top three or bottom three for any of the cluster, the fact that there are some differences in how the top three and bottom three are ranked in the different clusters does not make much difference. In other words, for the usability, all the clusters prefer and dislike the same designs, except for the Family cluster, which has minor differences in the ranking of the designs compared to the other clusters. However, this does not mean that the Family cluster prefer different designs than the other clusters because the Family cluster does not have any statistically significant differences in the usability for any of the designs.

For the aesthetics, there are some similarities as well. The only significant difference in aesthetics ranks is in the ranks of the Network cluster, which gives the Family cluster design a much better aesthetics rank than the Solar cluster design. Therefore, for the rest of the clusters, even though it is possible to rank the scores, there are only small differences, even between the top and bottom ranks. For the Contest cluster and the Solar cluster, the bottom three designs consist of the designs from the Solar, Pyramid, and Contest clusters, while the top three designs are from the Network, Family and Well-Oiled Machine clusters. For the Network cluster, the designs are ranked almost equally, but the Network cluster design is placed in the bottom three while the Contest cluster design is moved to the top three. For the Well-Oiled Machine cluster, the Family cluster, and the Pyramid cluster, the bottom three designs are from the Network, Solar, and Contest clusters while the top three are from the Pyramid, Well-Oiled Machine and Family clusters. These results indicate that, even though the differences between the rankings are small, the Well-Oiled Machine, Family and Pyramid clusters all thought that the designs that were localised to fit their culture were in the top three, but none of them placed their localised design as number one. On the other hand, the Contest cluster, Network cluster, and the Solar cluster placed their localised design in the bottom three, but not at the lowest position.

To summarise the results, there were not any statistically significant differences between what the different clusters think about the aesthetics or usability for the different designs. Also, when looking at how the different clusters rank the different designs, only a few minor differences were found. Therefore, the results of this research show that culture does not affect the users preferences for usability or aesthetics in websites. Moreover, this indicates that to localise the layout of a design to fit different cultures is not necessary.

However, even though the results showed that culture most likely does not affect our opinion about which designs that are the most user-friendly or beautiful, everyone has opinions about whether or not he or she think a website is easy to use or beautiful. However, these opinions may come from something that is not connected to culture, like familiarity. Often, what we are used to is seen as better than other options, regardless of whether it is the best choice. In this study, most of the participants were students living and studying abroad, mostly in Western countries. Even though they might not have been living in this country for very long, it is likely that they have been using the Internet a lot in their new country. This Internet use could have led them to get used to another type of websites than what they are used to in their countries of origin. Also, many people use international websites on a daily basis, and may be more accustomed to websites that were developed in countries and cultures otherwise foreign to them. This issue was not taken into consideration when

analysing the data, and it can be one of the reasons why cultural background does not seem to be as important for people's preferences in the designs.

6 | Conclusion and Future Research

6.1 Conclusion

During the research of this thesis I have looked at whether the aesthetics and the perceived user-friendliness of a system are related, and whether the cultural background of users affect their opinion about what a user-friendly or aesthetically beautiful system is. During this research, data has been gathered from people all over the world and it has therefore been possible to analyse the data both with and without the participants cultural background in mind.

6.1.1 Aesthetics and Usability

One research question was asked to investigate the link between aesthetics and usability.

> **RQ 1:** Does a user's perception of aesthetics affect their perception of usability?

The results of the survey showed that the aesthetics of the design does affect the perceived usability of a system, but the relationship is not very strong. Also, this relationship is most reliable when the aesthetics of the design is satisfying to the user. The aesthetics are, therefore, important to consider when designing a system because a well designed system makes the user perceive the system as more user-friendly. However, the aesthetics can not be used alone to design a user-friendly system as only a smaller percentage of the usability can be said to come from the aesthetics.

6.1.2 Cultural Differences

Two research questions were asked to investigate whether a user's cultural background affects their perception of user-friendliness and aesthetics.

> **RQ 2:** Does a user's cultural background affect which design they consider to be user-friendly?

> **RQ 3:** Does a user's cultural background affect which design they consider to be aesthetically beautiful?

The results showed that there were no significant differences between what the participants from the different clusters thought about the designs, both regarding usability and beauty. Also, even though there were some differences between what some clusters thought about each design, the differences were in general small. Therefore, based on the results, the conclusion for **RQ 2** and **RQ 3** is that a user's cultural background does not affect which designs they consider to be either user-friendly or beautiful.

6.2 Future Research

In this thesis, six different website designs have been used to investigate whether the aesthetics of a design affects the perceived usability of that design and whether our cultural background affects what designs we prefer. This study showed a link between the aesthetics and usability in general, but no link between a user's cultural background and its preferences for aesthetics or usability. However, both these questions should be researched more because this study is limited in both its number of participants and in the number of designs. Also, it is important to remember that in this thesis, all the results builds on participants first impression.

One possibility for further research is to do a more in-depth study on whether different aesthetics components affect the perceived usability of the design. It is interesting to see if some components have a stronger positive or negative effect on the perceived usability than others. To do a study about this could also help making generalisations and guidelines about how to design a system that is perceived as user-friendly, at least at first sight.

The second possibility for further research is to do a bigger version of the questionnaire with more participants and more designs. Also, the designs could be more diverse to make it easier for the participants to differentiate between the designs when taking the questionnaire. To do a study like this is interesting because even though this thesis did not find a relationship between the culture and the opinions on the usability or aesthetics on a system, that does not mean that a relationship like that does not exist. Also, the designs do not have to fit a particular culture, like in this study. They can just be designed as different websites.

A third possibility for further research is to do a similar study on websites, but also include language. This study could be interesting to do to see if the language of the website is important for the users, or if the translation is more important for people from some cultures than others.

A fourth possibility for further research is to investigate whether other factors than culture affects what users from different cultures prefer. Many experiments using participants from different countries connect all their findings to cultural differences. This kind of relation is also common in experiments where user interfaces or websites from different countries are analysed. However, it is worth looking into whether culture is the dominant factor to what users prefer, or if something completely different like familiarity is the most important factor. There is a possibility that many users are accustomed to a particular kind of website or user interface that are designed the way they are because of other reasons than culture, like familiarity. How familiarity can affect a user's perception of designs was discussed in Chapter 5 Discussion. Therefore, it is interesting to see whether all users would prefer the same kind of user interface, possibly only translated, as long as they are used to it. Alternatively, it would be interesting to investigate whether there are deeper rooted cultural reasons for differences in users preferences.

Bibliography

Baack, D. W. and Singh, N. (2007). Culture and web communications. *Journal of Business Research*, 60(3):181–188

Badre, A. N. (2001). The effects of cross cultural interface design orientation on world wide web user performance. Technical report, Georgia Institute of Technology.

Barber, W. and Badre, A. (1998). Culturability: The merging of culture and usability. In *Proceedings of the 4th Conference on Human Factors and the Web*.

Bevan, N. (2001). International standards for hci and usability. *International journal of human-computer studies*, 55:533–552.

Callahan, E. (2006). Cultural similarities and differences in the design of university web sites. *Journal of Computer-Mediated Communication*, 11:239–273.

Chau, P. Y. K., Cole, M., Massey, A. P., Montoya-Weiss, M., and O'Keefe, R. M. (2002). Cultural differences in the online behavior of consumers. *Communications of the ACM*, 45(10).

Christine Roy, M., Dewit, O., and Aubert, B. A. (2001). The impact of interface usability on trust in web retailers. *Internet research*, 11(5):388–398.

Constantinides, E. (2004). Influencing the online consumer's behavior: the web experience. *Internet research*, 14(2):111–126.

Cyr, D. (2008). Modeling web site design across cultures: Relationships to trust, satisfaction, and e-loyalty. *Journal of Management Information Systems*, 24(4):47–72.

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

Cyr, D. and Trevor-Smith, H. (2004). Localization of web design: An empirical comparison of german, japanese and united states web site characteristics. *Journal of the American Society for Information Science and Technology*, 55(13):1199–1208.

Doney, P. M. and Cannon, J. P. (1997). An examination of the nature of trust in buyer-seller relationships. *the Journal of Marketing*, pages 35–51.

Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics*, 6(3):241–252.

Fanelli, D. (2011). Negative results are disappearing from most disciplines and countries. *Scientometrics*, 90(3):891–904.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage, 4th edition.

Hartmann, J., Sutcliffe, A., and Angeli, A. D. (2008). Towards a theory of user judgment of aesthetics and user interface quality. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 15(4):15.

Hassenzahl, M. (2004). The interplay of beauty, goodness, and usability in interactive products. *Human-Computer Interaction*, 19(4):319–349.

Hofstede, G. (2011). Dimensionalizing cultures: The hofstede model in context. *Online Readings in Pshycology and Culture*, 2(11).

Hofstede, G., Hofstede, G. J., and Minkov, M. (2010). *Cultures and Organizations: Software of the Mind*. McGraw-Hill USA, 3rd edition.

Internet Live Stats (2015). Internet users. `http://www.internetlivestats.com/internet-users/#trend`. Accessed: 2016-05-19.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Med*, 2(8):e124.

Kim, S.-E., Shaw, T., and Schneider, H. (2003). Web site design benchmarking within industry groups. *Internet Research*, 13(1):17–26.

Kirakowski, J., Claridge, N., and Whitehand, R. (1998). Human centered measures of success in web site design. In *Proceedings of the Fourth Conference on Human Factors & the Web*.

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

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

Leard Statistics (2015a). Cronbach's alpha using spss statistics. *Statistical tutorials and software guides*. Retrieved from `https://statistics.laerd.com`.

Leard Statistics (2015b). Kruskal-wallis h test using spss statistics. *Statistical tutorials and software guides*. Retrieved from `https://statistics.laerd.com`.

Leard Statistics (2015c). Simple linear regression using spss statistics. *Statistical tutorials and software guides*. Retrieved from `https://statistics.laerd.com`.

Lindgaard, G., Fernandes, G., Dudek, C., and Brown, J. (2006). Attention web designers: You have 50 milliseconds to make a good first impression! *Behaviour & information technology*, 25(2):115–126.

Marcus, A. and Gould, E. W. (2000). Cultural dimensions and global web user-interface design. *Crosscurrents*, pages 32–46.

Oates, B. J. (2006). *Researching Information Systems and Computing*. SAGE.

O'Brien, H. L. and Toms, E. G. (2010). The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology*, 61(1):50–69.

O'Keefe, R. M., Cole, M., Chau, P. Y., Massey, A., Montoya-Weiss, M., and Perry, M. (2000). From the user interface to the consumer interface: results from a global experiment. *International Journal of Human-Computer Studies*, 53(4):611–628.

Reinecke, K. and Bernstein, A. (2011). Improving performance, perceived usabiltiy, and aesthetics with culturally adaptive user interfaces. *ACM Transactions on Computer-Human Interaction*, 18(2).

Remøy, M. A. (2015). User satisfaction and trust in cross-cultural web sites. Technical report, Norwegian University of Science and Technology.

Robbins, S. S. and Stylianou, A. C. (2003). Global corporate web sites: an empirical investigation of content and design. *Information & Management*, 40:205–212.

Seckler, M., Opwis, K., and Tuch, A. N. (2015). Linking objective design factors with subjective aesthetics: An experimental study on how structure and color of websites affect the facets of users' visual aesthetic perception. *Computers in Human Behavior*, 49:375–389.

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

Simon, S. J. (2001). The impact of culture and gender on web sites: An empirical study. *The DATA BASE for Advances in Information Systems*, 32(1).

Siu-Tsen Shen, Martin Woolley, S. P. (2006). Towards culture-centred design. *Interacting with Computers*, 18:820–852.

Tractinsky, N. (1997). Aesthetics and apparent usability: Empirically assesing cultural and methodological issues. In *CHI '97 Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*.

Tractinsky, N., Cokhavi, A., Kirschenbaum, M., and Sharfi, T. (2006). Evaluating the consistency of immediate aesthetic perceptions of web pages. *International journal of human-computer studies*, 64(11):1071–1083.

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

Tsikriktsis, N. (2002). Does culture influence web site quality expectations? an empirical study. *Journal of Service Research*, 5(2):101–112.

Tuch, A. N., Roth, S. P., HornbæK, K., Opwis, K., and Bargas-Avila, J. A. (2012). Is beautiful really usable? toward understanding the relation between usability, aesthetics, and affect in hci. *Computers in Human Behavior*, 28(5):1596–1607.

Tullis, T. S. and Stetson, J. N. (2004). A comparison of questionnaires for assessing website usability. In *Usability Professional Association Conference*, pages 1–12.

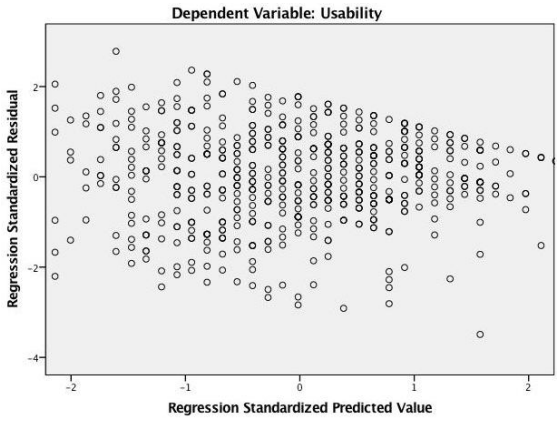
van Schaik, P. and Ling, J. (2009). The role of context in perceptions of the aesthetics of web pages over time. *International Journal of Human-Computer Studies*, 67(1):79–89.

Wikipedia contributors (2016). Cronbach's alpha. Retrived from `https://en.wikipedia.org/w/index.php?title=Cronbach%27s_alpha&oldid=719307038`. Accessed: 2016-04-11.

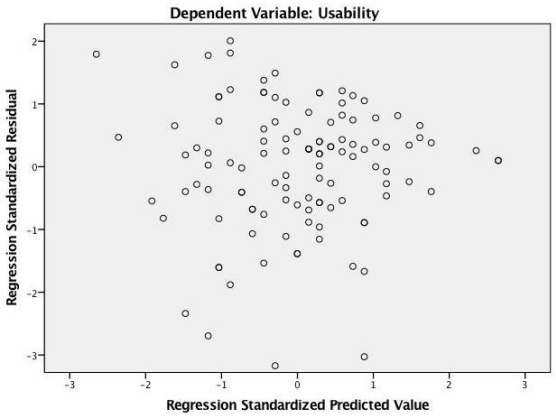
Wursten, H. and Fadrhonc, T. (2012). International marketing and culture. https://geert-hofstede.com/tl_files/Marketing_and_Culture_itim_International.pdf.

Würtz, E. (2006). Intercultural communication on web sites: A cross-cultural analysis of web sites from high-context cultures and low-context cultures. *Journal of Computer-Mediated Communication*, 11:274–299.

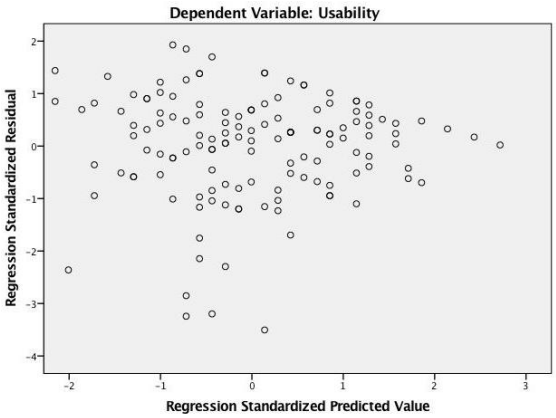
A Graphs for Homoscedasticity



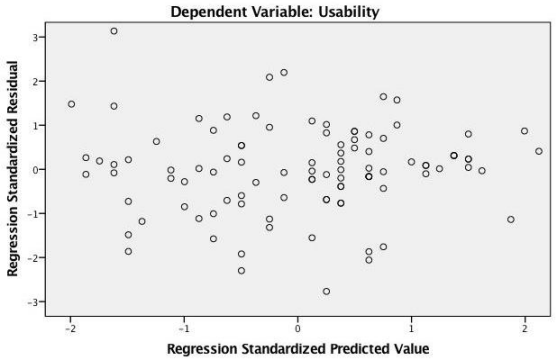
(a) All the designs combined



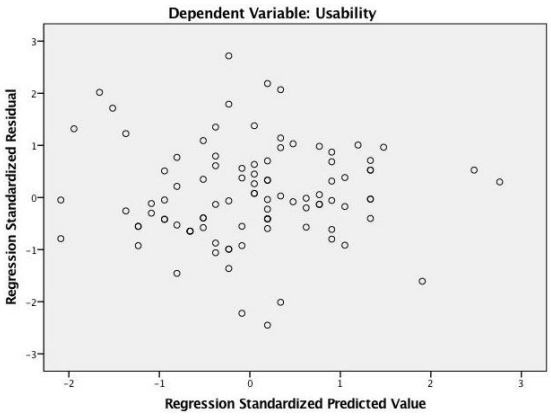
(b) Contest cluster design



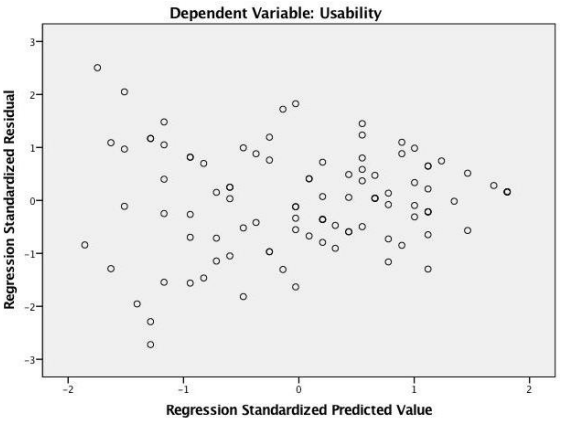
(c) Network cluster design



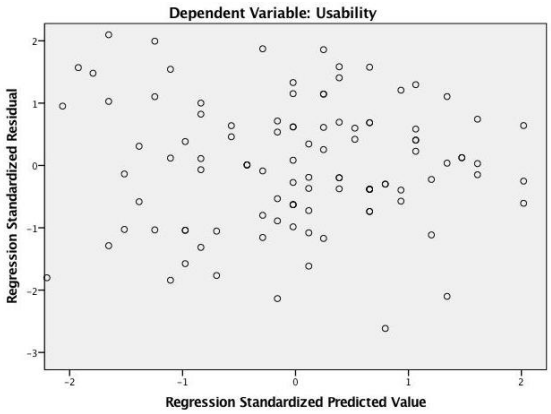
(d) Well-Oiled Machine cluster design



(e) Solar cluster design



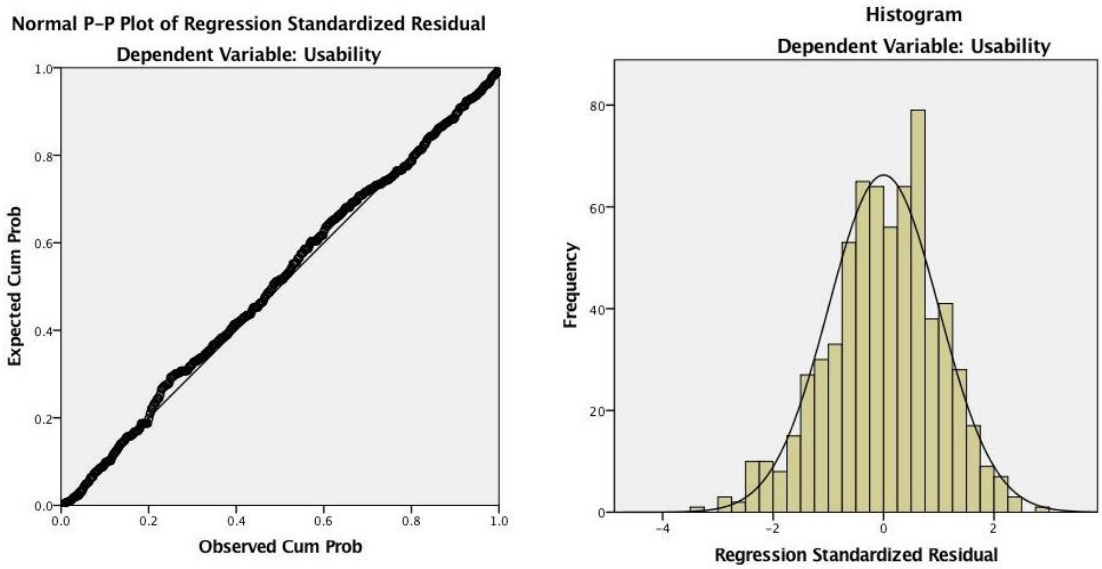
(f) Family cluster design



(g) Pyramid cluster design

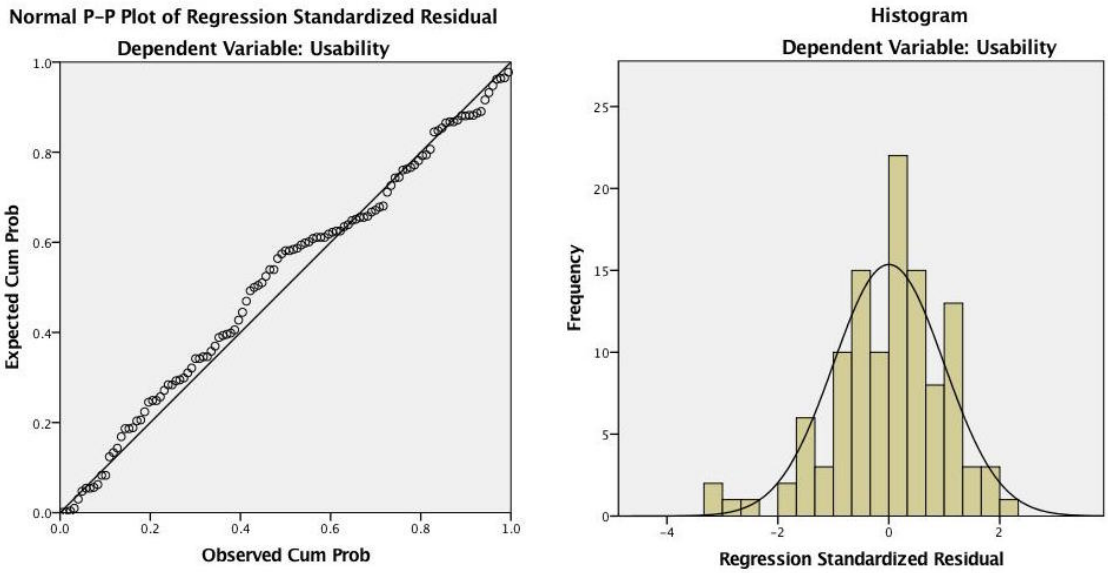
Figure A.1: Scatterplots for assessing homoscedasticity

B Graphs for Normality of Residuals



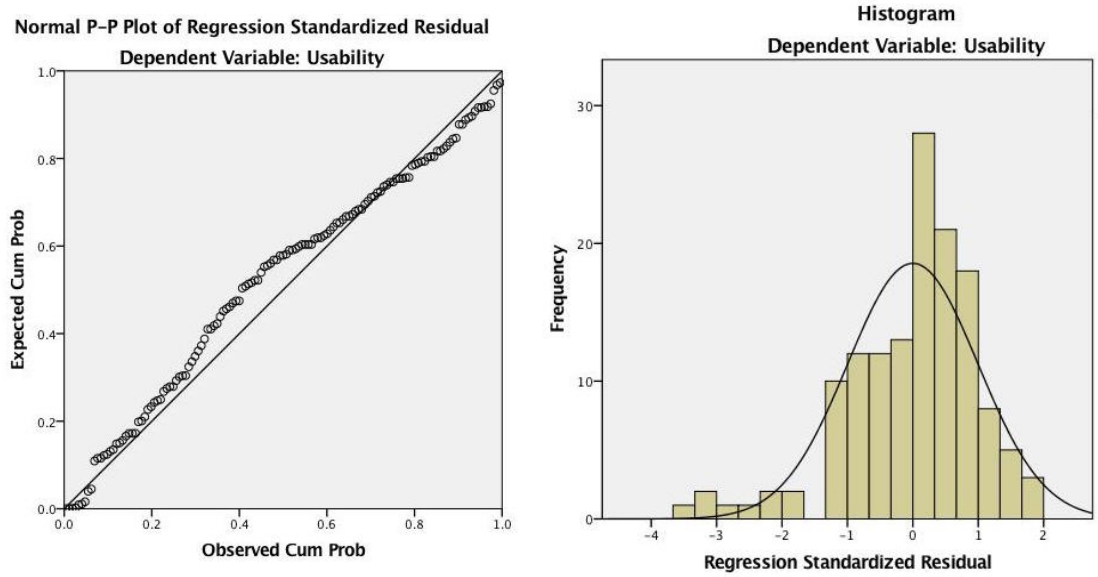
(a) (b)

Figure B.1: Normality of residuals in all the designs



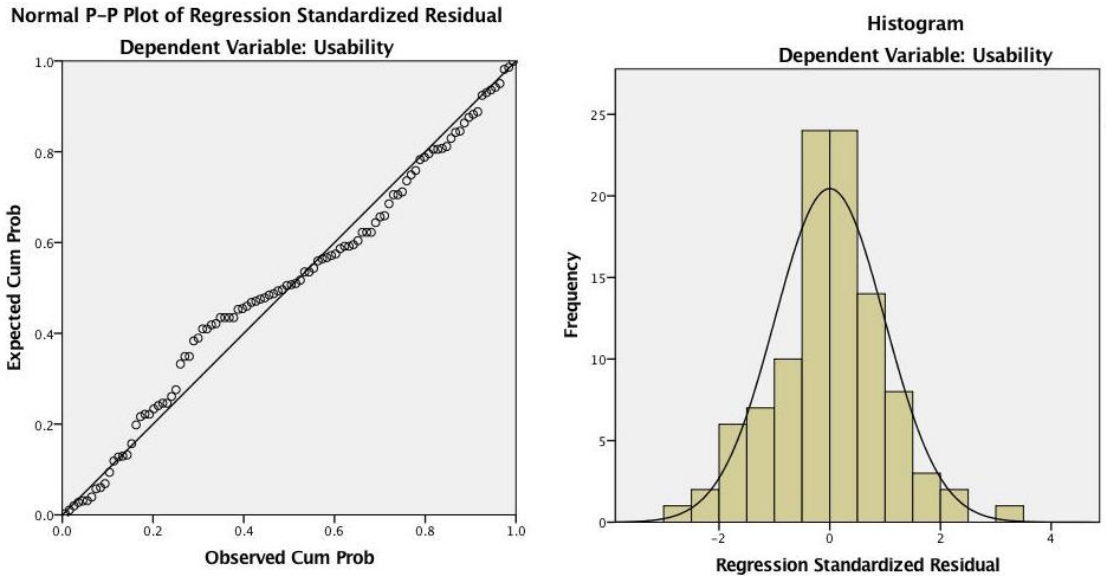
(a) (b)

Figure B.2: Normality of residuals in the Contest cluster design



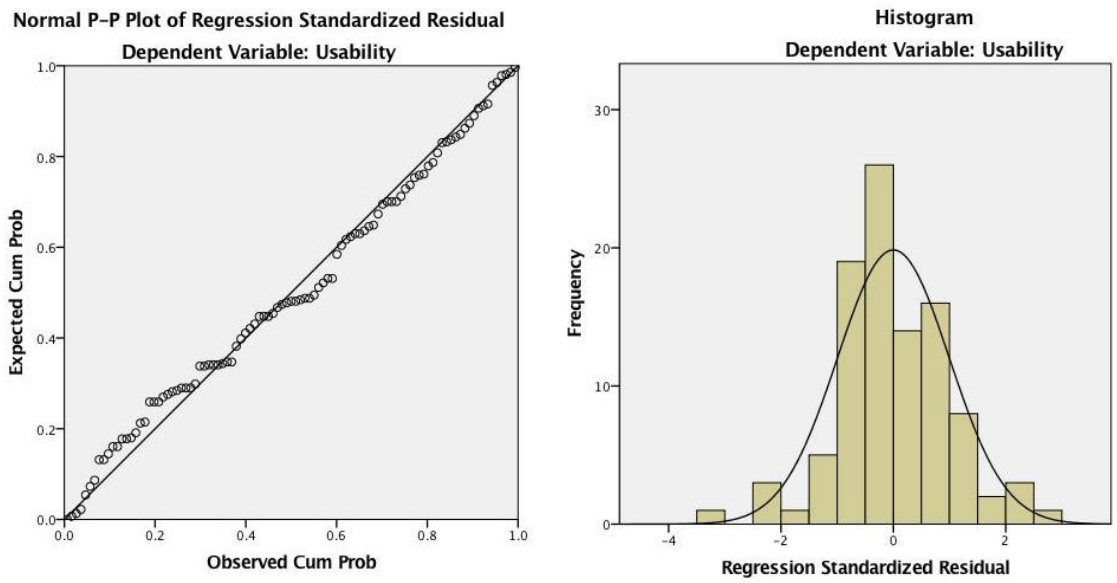
(a) (b)

Figure B.3: Normality of residuals in the Network cluster design



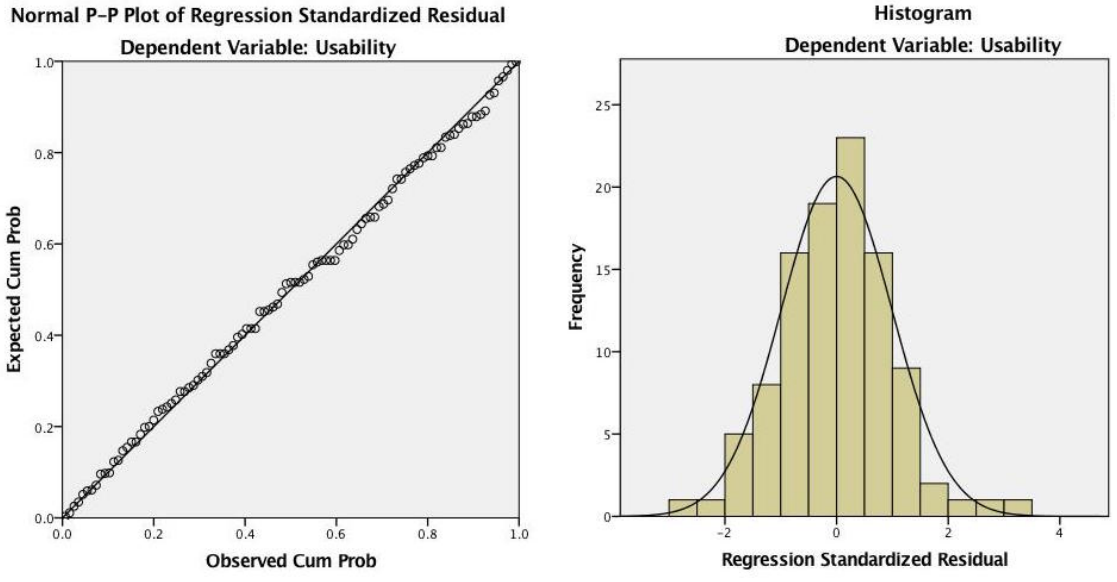
(a) (b)

Figure B.4: Normality of residuals in the Well-Oiled Machine cluster design



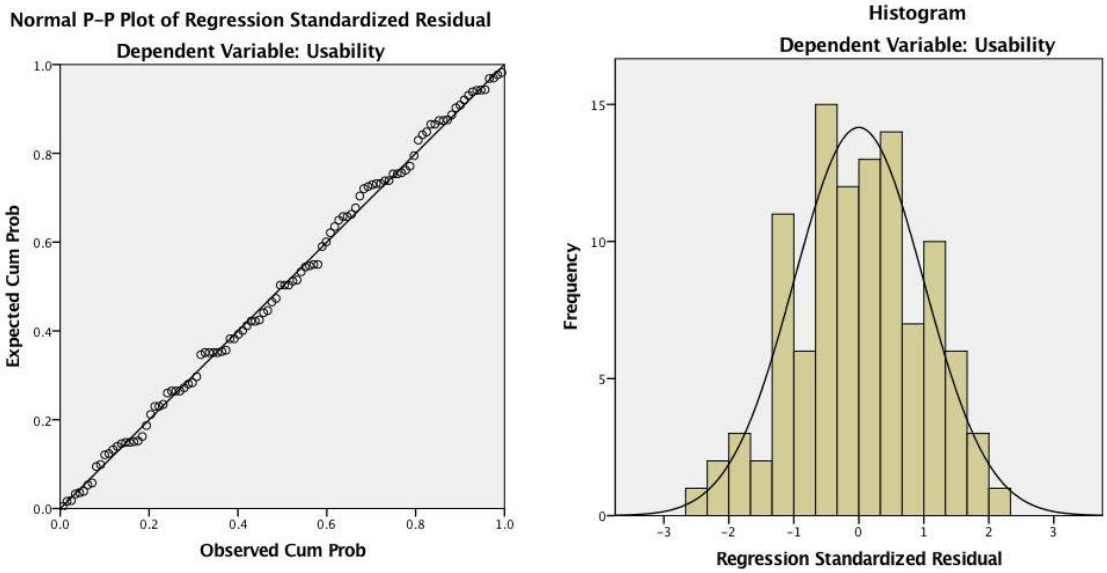
(a) (b)

Figure B.5: Normality of residuals in the Solar cluster design



(a) (b)

Figure B.6: Normality of residuals in the Family cluster design



(a) (b)

Figure B.7: Normality of residuals in the Pyramid cluster design

C Evaluation of the First Website Designs

*** 1. How tall is the hierarchy of this website?**

Shallow			Medium			Tall
○	○	○	○	○	○	○

*** 2. How complex is the structure of this website?**

Simple			Medium			Complex
○	○	○	○	○	○	○

*** 3. How many choices does this website provide?**

Few			Medium			Many
○	○	○	○	○	○	○

*** 4. How much information does this website provide?**

Minimal amount of information			Medium amount of information			A lot of information
○	○	○	○	○	○	○

*** 5. How would you describe the use of colors in this website?**

Homogenously coloured			Some colours			Many different colours
○	○	○	○	○	○	○

*** 6. How would you describe the acutal colors in this website?**

Highly contrasting and bright colours			Colours with medium saturation and contrast			Pastel colors with little saturation
○	○	○	○	○	○	○

*** 7. How would you describe the images in this web site (not the product images)?**

Individuals and their lifestyle	Products or merchandise being used by people	Neutral
○	○	○

*** 8. How much focus is it on sales on this website?**

Much focus			Some focus			No focus
○	○	○	○	○	○	○

*** 9. How would you describe the use of navigational icons on this website?**

No icons			Some icons			Many icons
○	○	○	○	○	○	○

D Questions about Background Information

*** 1. How old are you?**

- ◯ 0 - 18
- ◯ 19 - 24
- ◯ 25 - 39
- ◯ 40 - 60
- ◯ 60 plus

*** 2. What is your gender?**

- ◯ Male
- ◯ Female

*** 3. Where are you from?**

*** 4. In which country do you currently live?**

*** 5. In which country have you lived most of your life?**

6. What is the highest level of school you have completed or the highest degree you have received?

- ◯ Less than high school degree
- ◯ High school degree or equivalent (e.g., GED)
- ◯ Some college but no degree
- ◯ Associate degree
- ◯ Bachelor degree
- ◯ Graduate degree

*** 7. How often do you do these activities?**

	Daily	2 - 3 times a week	Once a week	2 - 3 times a month	Once a month	Under once a month	Not at all
Use the Internet?	◯	◯	◯	◯	◯	◯	◯
Browse for products in online stores?	◯	◯	◯	◯	◯	◯	◯
Purchase products from online stores?	◯	◯	◯	◯	◯	◯	◯

*** 8. Do you feel comfortable shopping from online stores?**

◯ Yes

◯ No

9. Why do you not feel comfortable with shopping online?

E Evaluation of the Second Website Designs

*** 1. How tall do you think the hierarchical structure of this website is?**

Shallow			Medium			Tall
○	○	○	○	○	○	○

*** 2. How much information do you think this website provides at first sight?**

A lot of information			Medium amount of information			Small amount of information
○	○	○	○	○	○	○

*** 3. How many choices do you think the user have when searching for products on this website?**

Many			Medium			Few
○	○	○	○	○	○	○

*** 4. How long do you think it would take to find a product on this site?**

Short period of time			Medium period of time			Long period of time
○	○	○	○	○	○	○

*** 5. How would you describe the use of colors in this website?**

Many different colours			Some different colours			Homogeneously coloured
○	○	○	○	○	○	○

*** 6. How would you describe the actual colours on the website?**

Pastel colours			Medium bright colours			Bright colours
○	○	○	○	○	○	○

*** 7. How would you describe the image(s) in this web sites (not the product images)?**

Products or merchandise being used by people			Neutral			Individuals and their lifestyle
○	○	○	○	○	○	○

*** 8. How much focus do you think there is on sales campaigns and discounts on this website?**

No focus			Some focus			Much focus
○	○	○	○	○	○	○

*** 9. How would you describe the use of navigational icons on this website?**

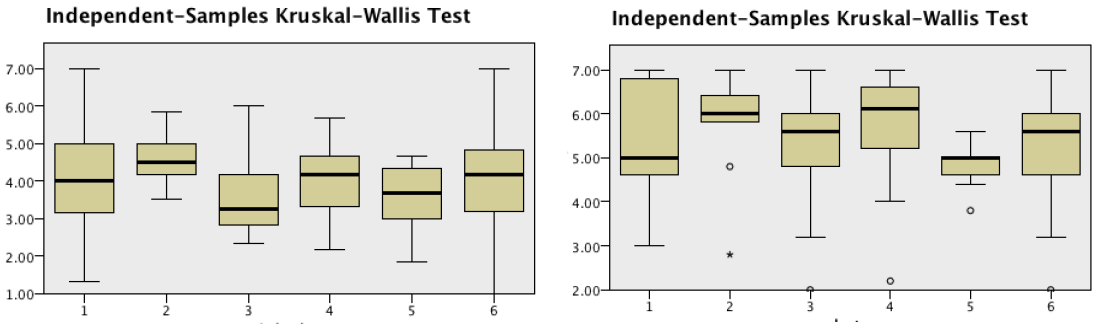
Many icons			Some icons			No icons
○	○	○	○	○	○	○

F | Facebook Groups for the Survey

2016 HU Berlin International Students
AIESEC WHOLE WORLD
Erasmus & Exchange Students Barcelona 2015-2016
Erasmus and International Students LYON 2015-2016
Erasmus ESN Leuven '15/'16
Erasmus Montpellier 2015 - 2016 (+ International Students) | ESN
ESN Dijon 2015 - 2016 Erasmus & International Students
Foothill International Students
Fuller International Students
Global AIESEC Conferences
Global Social Entrepreneurial Hub
IAESTE Switzerland Social
International Composers Network
International law students council
International networking J_I_S_C_P
International Students - Oxford Brookes University
International Students & Applicants
International Students at HiOA
International Students at the Faculty of Humanities - UiO
International Students at the University of Oslo
International Students at the University of Oslo
International Students ESC Rennes 2015/2016 (ESC Arrival)
International Students in Canada
International Students in Norway
International Students in Odense
International Students in Sydney
International Students Office - Concordia University
International Students Services - ISS Budapest
Milan International Group - Expats & Students
Montevideo International Students
NTNU International Students
NTNU Trondheim - Erasmus and International Students
Quest UQ
SBE International Business Students 2015-2018
TAMIU Student Network
UPEI International Students
Utrecht International Students
Yonsei I Spring 2016 I Exchange Students I Fall & Spring Semester

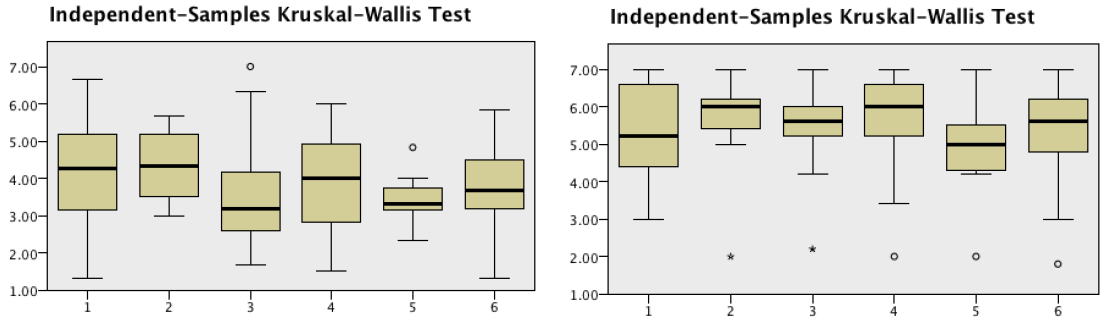
G Boxplots for Kruskal-Wallis T Test

G.1 Design by Design



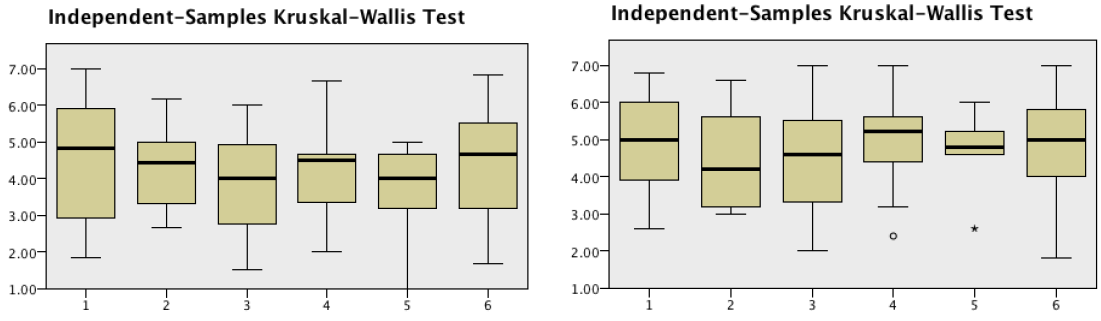
(a) Aesthetics (b) Usability

Figure G.1: Boxplots for the Contest cluster design



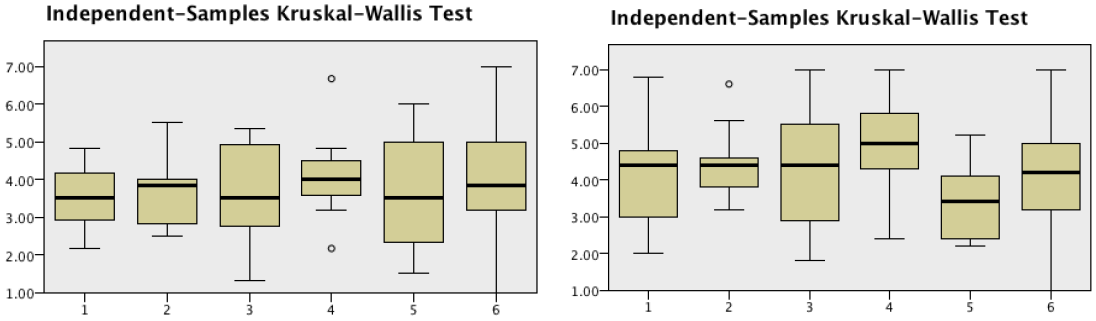
(a) Aesthetics (b) Usability

Figure G.2: Boxplots for the Network cluster design



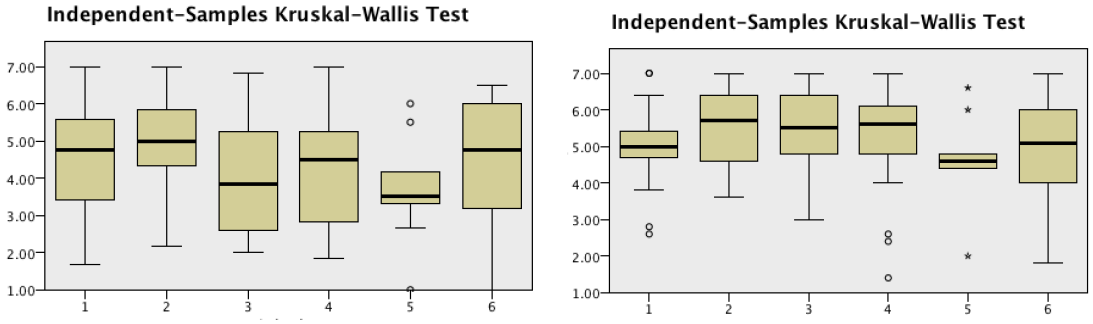
(a) Aesthetics (b) Usability

Figure G.3: Boxplots for the Well-Oiled Machine cluster design



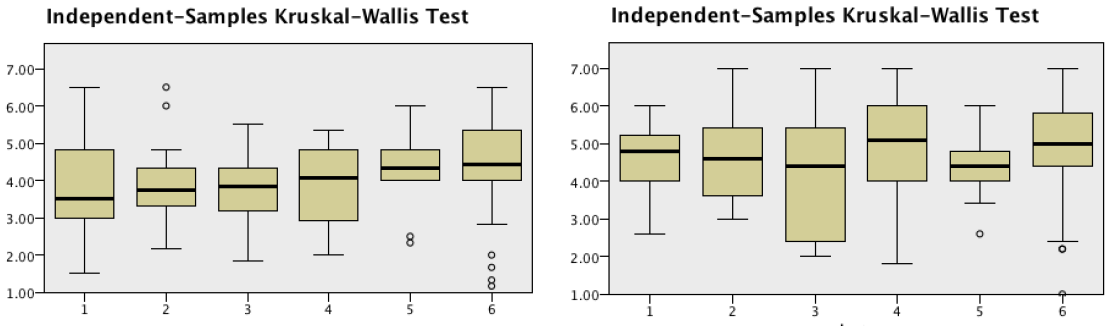
(a) Aesthetics (b) Usability

Figure G.4: Boxplots for the Solar cluster design



(a) Aesthetics (b) Usability

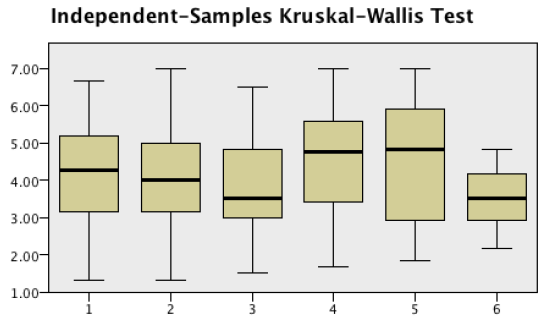
Figure G.5: Boxplots for the Family cluster design



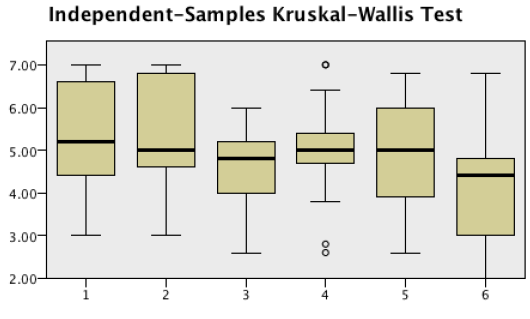
(a) Aesthetics (b) Usability

Figure G.6: Boxplots for the Pyramid cluster design

G.2 Cluster by Cluster

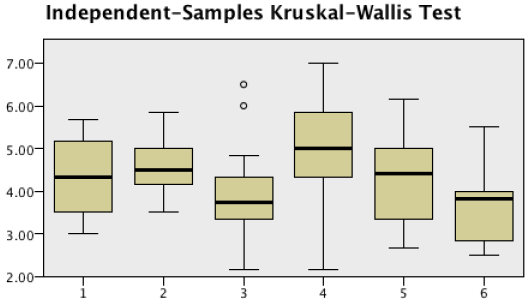


(a) Aesthetics

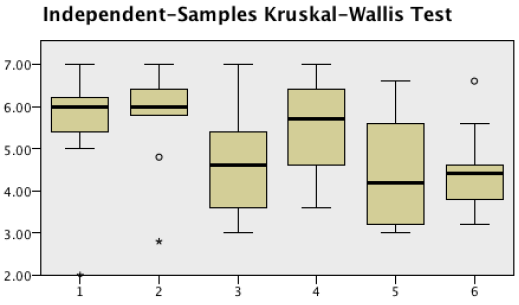


(b) Usability

Figure G.7: Boxplots for the Contest cluster

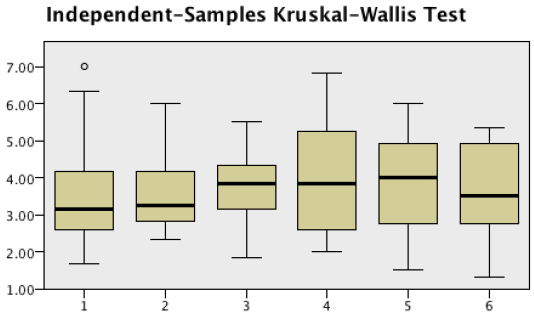


(a) Aesthetics

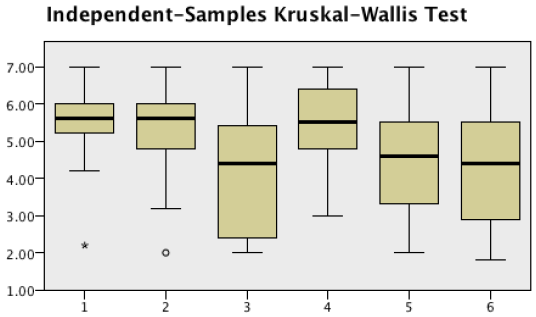


(b) Usability

Figure G.8: Boxplots for the Network cluster

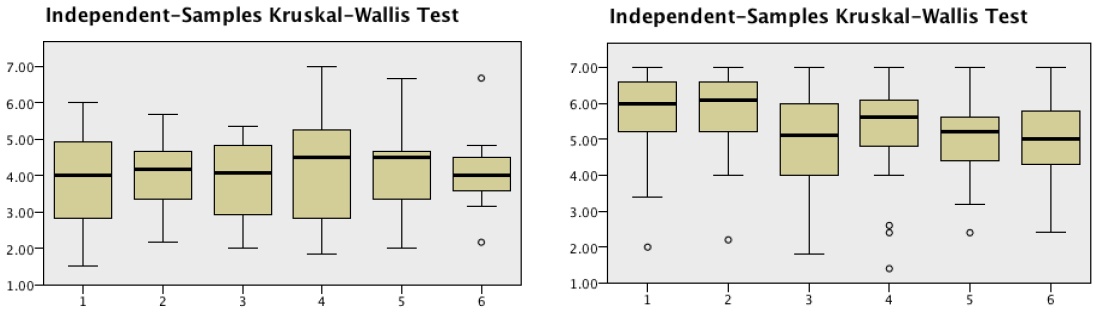


(a) Aesthetics



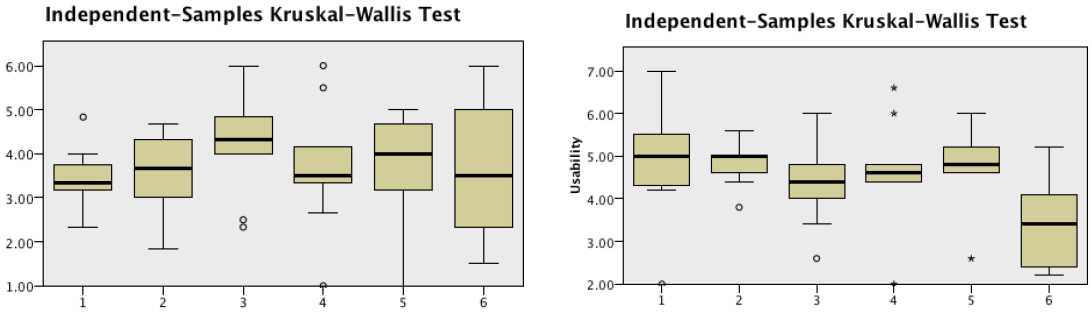
(b) Usability

Figure G.9: Boxplots for the Well-Oiled Machine cluster



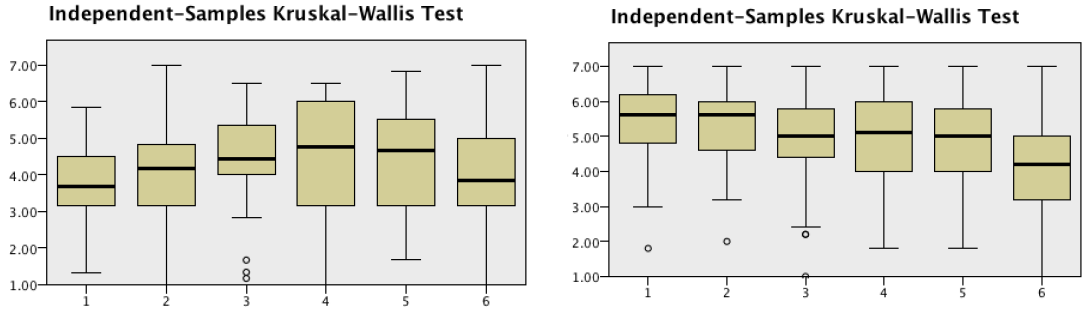
(a) Aesthetics (b) Usability

Figure G.10: Boxplots for the Solar cluster



(a) Aesthetics (b) Usability

Figure G.11: Boxplots for the Family cluster



(a) Aesthetics (b) Usability

Figure G.12: Boxplots for the Pyramid cluster