Objective predictors

of subjective aesthetic ratings

# of web pages

by

Nicola Kim Williams

A thesis submitted to Keele University for the award of

Doctor of Philosophy

in Computer Science

June 2015

Keele University

Abstract

This research is concerned with the effect of visual stimulus on decision-making and opinions, what visual aspects of a page affect very early impressions of web sites, and how this relates to computational methods of prediction and evaluation of web pages.

The aim of this study was to discover whether there are identifiable visual attributes of web pages that can be used to predict subjective opinions.

This was explored through three separate studies. These consisted of two correlational studies and a categorisation task. Participants were gained through convenience and snowball sampling, and the materials reviewed were two distinct sets of web pages. Cards sorts, laddering and an online data collection tool were used to gather the information. Both qualitative and quantitative analysis was used to explore the information.

The visual attributes found to correlate with subjective opinions were inconsistent across the two correlational studies. Study One had a number of limitations that may have contributed to this inconsistency. Concrete findings were that levels of encouragement and discouragement influenced by web pages are on two distinct scales, as, although there is a negative correlation between them, a large number of pages were rated poorly on both scales. The similarity between the card sort and questionnaire results had consistent findings for predictors of low-rated web pages.

The findings from the cards sorts also show that users are able to make preference judgements of web pages without being able to understand the content. An application of the findings regarding prediction of low-rated pages would be to create web design optimisation system, enabling web pages to be reviewed computationally. Although this should never replace user testing, it may provide an economical alternative during the early stages of design.

Table of contents

List of figures

List of tables

Acknowledgements

Firstly, I would like to express my great appreciation to my supervisor, Dr. Gordon Rugg, for his encouragement and support throughout this research. His ability to provide a surreal, yet relevant, anecdote in almost every situation never ceases to amaze me. His guidance and reassurance has been invaluable.

I would also like to show my gratitude to Prof. Pearl Brereton, and my colleagues and peers at Keele University, for their time, advice, and assistance.

I wish to thank my parents for allowing me find my own path, knowing they would support me whatever choices I made. I am grateful to my husband, Tony, who has always believed in me and my ability, even when I did not.

To my close friends, you will never realise how much help you provided, through discussions, questions, distractions, and friendship.

Finally, I would like to thank everyone that has participated in my studies. Without your willingness to give up your time and thoughts, this work would not have been possible.

Chapter 1 Introduction

1.1 Introduction

For many organisations, their first interaction with a potential consumer is through their web site (Robins & Holmes, 2008). Decisions about whether or not to pursue content on a web site are made very quickly (Briggs, 2002; Lindgaard *et al.*, 2006; Weinschenk, 2011). It is important, therefore, to create the best first impression possible. If, at a glance, a user has a negative emotional response to a web site they will leave and find an alternative option (Robins & Holmes, 2008; Bilal, 2000). Users decide to stay or leave a page based on many different factors including the visual design (van Schaik & Ling, 2003; Nielsen, 1995; Lindgaard *et al.*, 2006; Tractinsky *et al.*, 2006; Kim & Fesenmaier, 2008; Papachristos & Avouris, 2011). It is important, therefore, to optimise the visual impression a web page display in order to retain users.

Evaluation of web pages is currently undertaken by developers and through end-user testing. This thesis raises the question of whether any of these processes can be automated (Reinecke *et al.,* 2013). An example of automated evaluation within accessibility is the use of validation websites to ensure accessibility guidelines have been followed (Mankoff *et al.*, 2005). In order to identify predictors of subjective opinions of web pages, first the patterns within these preferences must be identified and explored.

There has been much research into the usability of web sites, with less research into understanding the aesthetics of web sites (Michailidou *et al*., 2008). This is surprising, when firstly a user has to decide they wish to stay on a site, so optimising first impressions is just as important as usability (Papachristos & Avouris, 2011).

1.1.1 Overall research aims

The research aims of this thesis are as follows.

One theme is identifying which visual aspects of a web page contribute to a user's very early impression. Previous research into this topic has typically focused on first impressions or impressions after completion of tasks. Studies into first impressions have focused on viewing times of 500ms (Lindgaard *et al.*, 2006; Tractinsky *et al.*, 2006). This topic has been investigated using experiments, usually with a focus on a specific aspect such as colour, font, or screen design. A limitation of this method is that the set of sites used may not accurately reflect real sites. There are also examples of surveys, using existing sites as well as those created specifically for the study, and participants' opinions of the sites are gathered. In many cases values are gathered for sites, without an opportunity to ask the participants why they have given a particular opinion. This thesis examines this question via an online questionnaire survey and a card sorting study.

A second theme is identifying possible objective proxies for subjective impressions of a web page. Once trends between individual ratings of web pages have been identified, it is then possible to see if subjective opinions can be predicted by measuring a web page objectively. This would allow for evaluation without the need to show users the site (Ngo & Ch'ng, 2001). There is limited previous work in this area, however, the topic is receiving increasing attention, with a study by Reinecke *et al.* (2013) seeking to predict users' first impressions of websites. In order to identify possible proxies, the literature regarding a number of aesthetic preference domains will be reviewed. Links between the objective measures and the subjective ratings will be tested through a correlational study.

The literature underpinning the research topic can be found in Chapter 2, with additional, study specific, literature reviewed in the relevant chapter.

1.1.2 Objectives

- To determine whether it is possible to predict users' early impressions of websites through objective measurements of the pages.
- To determine whether encouragement and discouragement are on a single scale, in the context of web pages.
- To investigate the effect of viewing time on user preferences towards web pages.
- To compare two separate evaluation techniques used to identify user preferences towards web pages.
- To investigate the order of criteria within card sorts, specifically whether patterns exist regarding the timing of the most and least important sorts.

1.1.3 Thesis overview

Chapter 2 presents the literature on a number of topics, including an exploration of the effect of visual attributes of artefacts on decision-making, and explanations for this. It also covers examples of proxies within computing and other domains, for example, recommender systems and search engine optimisation. The importance of further research into the effect of aesthetic aspects of a page is explored, and placed within the context of existing web design research.

Chapter 3 presents two web-based studies that consist of the collection and analysis of subjective ratings for a set of web pages. The information is gathered through a survey using Likert style visual analogue scales. It explores potential underlying reasons for the ratings through an exploration on the visual attributes of the pages.

Chapter 4 presents two studies that consist of the collection of objective values for potential predictors of subjective opinions, based on the literature presented in Chapter 1. Correlations are analysed to see if a connection between the objective measures found and the subjective values gathered in Chapter 3 can be supported. A comparison of web page domain is then conducted through an analysis of the similarities and differences between the two sets of findings.

Chapter 5 describes a study that investigates user categorisation of web pages. A card sorting activity was used with a set of thirty web page screen shots. Laddering was conducted to investigate the belief structure that influenced the categorisation process. The results were then thematically analysed and conclusions drawn. The findings from this study have been contrasted with the findings from Study Two as described in Chapter 3.

Chapter 6 summarises the conclusions of the research, the implications of the findings and highlights suggestions for further work in this research area.

Chapter 2 Literature review

2.1 Introduction

This section explores the background literature that forms the basis of this research topic, to understand what has been researched previously, what unanswered questions there are, and to inform the design of this research. A more study-specific literature review can be found at the beginning of each chapter. The focus of the research conducted in this thesis is within the field of computer science. However, in order to fully understand the context of this research, an overview of related areas of psychology has been provided. This contributes to the justification for the research undertaken in this thesis.

2.2 Decision making

Payne (1976) hypothesised that for decision-making concerning information processing, the information used to make the decision differs dependent on the complexity of the task. He found that when individuals were faced with a large number of options they employed a process to reduce the problem space, or number of options, using a simplistic method, before making a rational choice from the remaining options. It is conceivable that this is what happens when users are interacting with web pages. If an information location task is proving too difficult they may try to reduce the amount of information processing they have to do by locating an alternative web page that is not as complex. Web designers that have an awareness and understanding of these types of cognitive thinking biases have an advantage over those that do not. An alternative would be to determine where the boundaries of too much information lie, and create software to analyse a page to determine whether a site is too complex.

Within the field of psychology it is widely accepted that cognitive processing can be placed into two categories: conscious and subconscious processing. Throughout this thesis the terms of System 1 and System 2, originally proposed by Stanovich and West (2000), and later adopted by

Kahneman (2011), will be used. System 1 is concerned with swift, subconscious, immediate, low effort decision making (Evans, 2008). These decisions are often based on pattern matching, or through a learnt skill or knowledge, rather than a rational, evidence based, decision-making process (Kahneman, 2011). System 2 deals with more complex processing, often making use of additional information or evidence, in order to make a decision (Stanovich & West, 2000). When System 2 is being used, a higher level of effort is required, and the ability to continue performing previously effortless tasks, is impaired. The brain will always have a preference to singularly make use of System 1, if possible (Kahneman, 1973) to reduce draw on other systems. System 2 can also be thought of as a rule based system (Evans, 2008) and will make use of information from System 1 in order to make decisions. Due to these inherent preferences and mechanisms for decision-making, it is conceivable that, even when System 2 is used, individuals do not always accurately evaluate evidence when making decisions (Andrade & Ariely, 2009). These departures from rationality, or errors in judgement, are collectively known as thinking biases (Stanovich & West, 2008). Examples include the use of framing, primes, objective bias, and belief bias (*Ibid.*).

2.3 First impressions of websites

Lindgaard *et al*. (2006) conducted a study to determine how quickly first impressions of websites are formed. They showed participants websites for either 50ms or 500ms and asked them to rate the sites using a visual analogue scale. They compared the responses between the different time conditions, and compared the responses to previously gathered ratings from two other studies reported in the same paper. They found that participant's responses were consistent between the two time conditions, and also with the previously gathered ratings. They concluded that first impressions could be determined in less than 50ms. These findings were specific to a single set of web pages, however, they have subsequently been replicated and extended by Tractinsky *et al.* (2006), and Papachristos and Avouris (2011). All of the mentioned studies considered participant's ratings of web pages which they looked at, but could not interact with.

Studies by Lindgaard *et al.* (2006), Tractinsky *et al.* (2006), and Papachristos and Avouris (2011), all showed participants web pages for a very short period of time and gathered ratings for the web pages. The length of exposure to the visual stimuli was not sufficient to read all the content, and therefore, it was deemed that the ratings must be based on the visual, not verbal, content (Lindgaard *et al.*, 2006). This introduces questions about what visual elements of the page led to the ratings.

A study by Lavie and Tractinsky (2004) looked at different factors that were used to evaluate web page aesthetics, and they categorised these factors into two dimensions: Classical aesthetics and expressive aesthetics. In a later paper Tractinsky *et al.* (2006) used these two dimensions to evaluate the sets of web pages used for their studies into first impressions, and they found that participants' ratings were consistent with the aesthetic evaluations, showing participants could, and do, judge aesthetics very quickly. This finding replicated that of Schenkman and Jönsson (2000), who also found correlations between visually appealing pages and users' ratings of a site. Schenkman and Jönsson (*Ibid.*) also found links between pages with a high level of symmetry and those with high user ratings. However, Tuch *et al.* (2010) explored symmetry, and focussed on differences in preferences due to gender. They found that high levels of symmetry correlated with high aesthetic judgements only for males, not females. Tuch *et al.* (2009) looked for correlations between visual complexity of web pages and user's experiences, ability to complete tasks, and judgements. They found that web pages with a low level of complexity were more favourable to users. They also found that task completion was negatively correlated with visual complexity, with low complexity pages having higher task completion values.

In order to understand how important it is to create a positive first impression, it is worth investigating how long lasting first impressions are and how easily they can be changed. A user's initial impression of a site affects whether they stay on the site, or return to their search results (Bilal, 2000). Kim & Fesenmaier (2008) looked at website first impressions in the domain of tourism. In this domain, most customers find a destination web site through a search engine. If a user has a negative first impression and choses to return to their search results, there is no opportunity for the site to alter this initial impression. In the context of web design, how well a website can hold the attention of the user is often referred to as the "stickiness" of a site (Holland & Menzel-Baker, 2001). If visual cues can be identified that encourage users to remain on a site, this would provide useful insight into decision making within the field of web design and act as an aid when improving the "stickiness" of a particular site. It has also been suggested, that once a first impression is formed, users seek out information that conforms to this initial impression (Lindgaard *et al.*, 2006; Tractinsky *et al.*, 2006; Kim & Fesenmaier, 2008; Papachristos & Avouris, 2011). This phenomenon is known as the Halo effect (Nisbett & Wilson, 1977) or confirmation bias (Koriat *et al.*, 1980). It is clear this first impression is important, but it is not clear how much weight this initial first impression carries. There has been a greater focus of research into the usability of sites than the aesthetics of sites (Michailidou *et al.*, 2008). However, usability will have no impact on a user if the user has left the site before interacting with it. Therefore, optimising first impressions is just as important as usability.

2.4 Aesthetics

Web pages have been said to be complex visual stimulus (Tractinsky, 2006). Therefore, literature relating to aesthetic preferences in other visual stimulus, such as art, has been considered.

In the field of art much research has been undertaken concerning aesthetic preferences and the underlying reasons for some of these preferences. There has been research into beauty and symmetry (Chen *et al.*, 1997; Zaidel & Cohen, 2005, Zaidel *et al.*, 2005). Some of this has focused

on images, while others have focused on faces. Faces have been explored to investigate whether there are ways to measure the face and predict how people will view you in relation to health, beauty, and symmetry (Chen *et al.*, 1997). Some of these experiments focus on whether people can perceive how symmetrical faces are, which usually involves manipulating images of faces to make them more or less symmetrical. Zaidel and Cohen (2005) found that both male and female participants preferred a symmetrical image of the right side of the face when looking at female images.

The Golden Ratio is a more commonly known concept which is often used when considering images, whether they are artwork, photographs, or videos. The Golden Ratio is a proportion which, when found in art, music and nature, is considered aesthetically pleasing. Euclid's (ca. 300 BC) definition of the Golden Ratio is, "A straight line is said to have been cut in extreme and mean ratio when, as the whole line is to the greater segment, so is the greater to the lesser."(cited by Livio, 2002). Works of art that incorporate the Golden Ratio, such as Dali's Sacrament of the Last Supper, are perceived to be highly aesthetically pleasing. However, research has been inconclusive as to why this is. The use of these proportions is taught in photography, to the extent that some cameras have a superimposed grid to allow the photographer to line up their subject correctly in relation to the thirds of the entire frame (Brewster *et al.*, 2012). The Golden Ratio and Fibonacci numbers are often found together with examples being the number of petals on flowers, the spacing of seeds and leaves, and the spirals found on pine cones (Knott, 2003). Other examples of the Golden Ratio and Fibonacci are found in the construction of musical instruments such as violins (Laser, 2012), and many Mozart sonatas (Putz, 1995).

2.4.1 Colour psychology

Colour psychology is a well-researched topic, however many questions still remain unanswered (Whitfield and Wiltshire, 1990). The research that is most relevant to the topic of this thesis concerns colour preferences, and colour associations. Colour preference research focuses on

individuals' reactions to visual stimulus, and attempts to generalise these reactions to specific groups, such as gender, age or cultural background. Kreitler and Kreitler (1972) have reviewed much of the early work in this field. Individual studies have found trends in group preferences, however, when these studies have been replicated or extended, the results have been inconsistent. The reporting detail of many of the early studies makes it difficult to draw conclusions across them, as the information reported makes it difficult to ensure the colours used were actually the same in any two studies (Whitfield and Wiltshire, 1990). More recent research has improved this situation, and usually includes information about hue, value and chroma, making comparison across studies more reliable. There has also been research into individual response to colours, or colour associations. The early work on this topic was conducted by Bullough (1908). He found that participants' reaction to colour was due to four different factors: the objective aspect, the physiological aspect, the associative aspect, and the character aspect (Bullough, 1908). This early work shows that it can be difficult to link a reaction to a single specific factor. Therefore, any work considering reactions to colour is affected by many additional factors, in addition to the colour being presented. The use of colour tile studies limits these additional factors, however, by removing the colour form any context, the influence of that context is also removed. Sivik (1974) tried to disambiguate objects and their colour in their experiments, but participants found it difficult to separate objects and their colour when making judgements.

2.4.1.1 Colour psychology and marketing

The incorporation of colour within a store environment is thought to affect purchasing intentions and brand perception. In the context of e-commerce, the purchase location, the store, is replaced by a web site (Calitz & Scheepers, 2002). Therefore, looking at the influence of store design on purchasing habits is relevant to the research in this thesis. Crowley (1993) conducted a survey by manipulating the background colour of a furniture store and measuring the effect this had on perceptions of the store environment and the merchandise within the store. The survey found that blue stores were viewed more positively than red stores, while red stores were perceived as

more active environments. The colour of the store was not found to affect perceived merchandise quality but other attributes were affected by colour, with red being associated with "up-to-date" merchandise. Babin *et al*. (2003) undertook a similar study, with a focus on women's fashion retail. In this study they compared blue and orange store colours and found that blue led to a more positive evaluation of the store than orange, with a positive correlation between store evaluation and purchasing intent. These findings were specific to the domain, and caution should be used when generalizing, as the attributes and perceptions desired in women's fashion may not be valid in other domains. In studies relating to colour, the colours used can be categorised a number of ways. A common grouping is into cool (green, blue, and violet) or warm (red, orange and yellow) colours (Crowley, 1993). Khouw (2002) used these groupings to investigate the effect of interior room colours on emotional response. She found differences in response, based on gender, as well as differences due to hue, value and chroma rather than just whether the colour was warm or cool. This contrasts with Crowley (1993), who only considered whether the colour belonged to the grouping of cool or warm. When considering purchasing habits via web sites it is not possible for a designer to control the colour of the room the purchaser is in, but they can influence the background colour of the site and any images of products visible.

Focussing on the example of packaging design, the effect of packaging colour choice is a highly relevant question for businesses. Labrecque and Milne (2012) investigated the effect of particular hues on a purchaser's perception of a product. They reviewed colour psychology literature to determine which hues were frequently associated with particular dimensions of brand personality and took a small number of these associations and tested them in isolation. One of their experiments examined whether a particular product would be perceived as more sophisticated if it were in packaging in a low-saturation, high-value purple hue rather than a high-saturation, low-value red hue. They also investigated whether the same product would be perceived as more rugged if the packaging were in a high-saturation, low-value red hue rather than a low-saturation,

high-value purple hue. Their experiment found this was the case in both instances, and also found that if participants were asked to choose a product matching these attributes (sophisticated or rugged) they were likely to choose the product the researchers had intended to be picked. This suggests that product packaging can be manipulated to influence a purchaser's perception of a product in predictable ways and that, if the attributes which are valued in a product can be identified, businesses can tailor the packaging to match. Bottomley and Doyle (2006) conducted a study which considered use of colour and the type of product being sold. They had two hypotheses: first that functional brands should be presented in functional colours rather than sensory-social ones, and second, that users would prefer the colour red over the colour blue to be used for functional products and the colour blue over red for sensory social. They used logos consisting of a fictional brand name in a consistent typeface for a number of product types. The logo was then presented in different colours (a single colour at a time) and participants were asked to rate each variation of the logo for appropriateness on an 11-point scale. This experiment found that both their hypotheses were supported by the evidence gathered suggesting that blue and red are linked to sensory- social and functional products respectively. One limitation of this research was the use of a single colour in the logos that is not reflective of everyday logos used within industry. This leaves the question of whether the colours must be used in isolation to have the type of links to different products observed within this experiment. Labrecque and Milne (2012) considered the representativeness of single colour logos and found that in a review of the logos for the top 100 companies, the majority used a single colour (e.g McDonalds), or a dominant colour with an accent colour (e.g. Starbucks).

Colour preference and the effect of colour is thought to differ based on demographic factors such as gender, culture and age (Aslam, 2006; Madden *et al.*, 2000). Madden et al. (2000) conducted research into the meaning of colours within eight different cultures. Participants answered questions relating to the meaning of a set of colours and their personal rating of the colours. The participants were then presented with partially coloured in hypothetical logos and asked to shade

the rest of the logo in a colour of their own choosing from the set provided, which gave information about subconscious colour combinations. Colours which generally appeared to be well liked were red, black, green, blue and white. However, only the last three of this set were found to have consistent meanings cross culturally. Another aspect thought to affect colour perception, and its subsequent affect, is current trends in colour preference. It should be noted that this key research into cultural differences in colour meaning and preference is now thirteen years old, and this work should be revisited to see if the passage of time has led to any differences in the meaning and perception of colours cross culturally.

2.5 Web site evaluation

Web sites are commonly evaluated for usability. This measure concerns how easily a user can interact with a page. As the majority of interactions take more than a matter of seconds, it is unlikely that a user forms their first impression entirely based on usability (Lindgaard *et al.*, 2006), as the time taken to interact with the page is usually longer than the time taken to make a decision about the page. If a user rejects the page during this first pass stage, the usability of the site does not have an opportunity to influence the user. Once an initial decision to remain on a page has been made, the usability of the page will help form the user's future opinion. Robins and Holmes (2008) considered the effect of website design on the perceived credibility of a page, finding that users make their judgements on credibility swiftly and before they have examined the content. The judgement of a website varied based on the level of aesthetic treatment, leading them to suggest that decisions regarding credibility of websites were influenced by aesthetics. Tractinsky *et al*. (2000) considered the question of whether aesthetic website designs were also usable, and in their research this proved to be the case.

Usability can be measured through heuristics, and commonly used sets of heuristics have been proposed by Nielsen (1995), Norman (2002), Shneiderman (2010), and Resmini and Rosati (2011). Each set of heursitics covers similar themes, and the focus is on content rather than aesthetics.

Of Nielsen's (1995) heuristics, only one discusses the aesthetics side of design. The heuristic encourages designers to use an aesthetic and minimalist design, but does not actually explain what would make the design aesthetic. The concept of minimalist also links to work by Tuch *et al.*(2009), who found less visually complex web pages were considered more favourably. However, there is no clear guidance where the line is drawn for when a web page is minimalist or not. Therefore, there is scope to explore this topic, in order to create guidance regarding the aesthetic design of a web page.

Previous research into the effect of visual stimulus on web page users has addressed beauty and web page usability. Tractinsky *et al*. (2000) found a positive correlation between web site perceived beauty and perceived usability, and this link remained once users had interacted with the sites. Ngo and Ch'ng (2001) looked at web page screen design and symmetry, and modified a number of web pages to match their screen design to previously identified ratios from the existing research. They conducted a survey to discover whether participants preferred the original or modified pages, and found that users preferred the modified pages. A subsequent study by van Schaik and Ling (2003) extended this experimental research and asked participants to carry out an information retrieval task using a number of web sites adhering to five different screen ratios. One of the ratios, corresponded to the Golden Section discussed above. Aspects which were explored included efficiency of task completion, speed of task completion, and subjective preferences for the display. The study found that the screen ratio corresponding with the Golden Section performed poorly compared to the other ratios for the performance measures, and was less likely to be selected as the preferred ratio for any of the subjective preferences. One possible explanation for this contradiction to the findings of Ngo and Ch'ng (2001) was the addition of the information retrieval task. It is conceivable that what users believe they will prefer, just by observation is inconsistent with their preferences after interacting with a site. Another key difference between the two studies is that van Schaik and Ling (2003) used web sites with multiple pages, whereas Ngo and Ch'ng (2001) appear to have used a single page from each site.

Mitra *et al*. (2005, pp12) found that "user criteria that make a web site most attractive are the basic textual content of the site and how well the site responds to the specific interest of the user", suggesting that users are not as influenced by the more aesthetic qualities that some research suggests.

Website quality is measured using many different factors. Which factors are used, and how they are weighted can often depend on who is analysing the site and the reason for the analysis. Much work has been done on how accessible and usable websites are. Research has demonstrated the perceived attractiveness and perceived usefulness of a website has a positive relationship highlighting that creating an attractive website is of high priority. However according to Tombros *et al*. (2005, pp. 3) quality can be "based on the presentation and source of the document". An alternative viewpoint for website quality was found by Borges and Levene (2006) who discuss systems for ranking pages by popularity within a site. However their methods cannot easily be extended for different sites due to the techniques used.

Another relevant question is whether web design should be culturally specific. Many suggest that in order to appeal to the target audience, the product should be tailored appropriately. The internet has made this more difficult, as it can be difficult to know who will be viewing each site, which has led to researchers exploring cross cultural preferences for websites, along with sensitivity to culture specific websites (Callahan, 2005; Cyr *et al*., 2005). User preferences can encompass navigational styles, particular images, graphical design, or particular colours. It is important that by tailoring a site to one set of users' preferences, another set is not alienated. Cyr *et al*. (2005) addressed this issue and identified preferences for different factors within the web sites used. They then trialled sites optimised for particular sets of users, and asked questions concerning the level of trust the users put in the web sites. They did not find users distrusting sites optimised for other countries. These results were different from those previously gained by Yamagishi and Yamagishi (1994). Satisfaction was also found to be similar for sites optimised for a

participant's own culture and a foreign culture, suggesting satisfaction is not influenced by cultural designs. Previous studies have also had difficulty pinpointing what would optimise a web site for particular cultures with any level of reliability. Hermeking (2005) suggests that the criteria looked for in a web site are similar cross-culturally, with main factors being site quality, usability and trust, although the indicators of these factors do vary.

2.6 Research aims

The research presented in this thesis adds to the existing research concerning first impressions of web sites. It seeks to identify visual stimulus that contribute to these very early impressions, with a particular focus on the effect of colour. It also seeks to identify aesthetic trends within high- and low-rated web pages. This research also seeks to provide a better understanding of the aesthetic elements of a web page that lead to users staying on, or leaving, a page. The information from this research seeks to provide recommendations for website developers and researchers.

Chapter 3 Subjective ratings of Web pages

3.1 Introduction

This chapter describes two independent studies, undertaken to facilitate the correlational studies described in Chapter 4. In order to complete the correlational studies quantitative values were gathered for two sets of web pages using a specifically developed online tool. The quantitative values represent users' subjective opinions of the two sets of web pages. This chapter describes the studies conducted to gather these subjective values.

The main question addressed by this chapter is whether trends exist in visual attributes of high- and low-rated web pages. This links back to the literature concerning first impressions of websites, discussed in Section 2.3. Two sub-questions will also be investigated. The first is whether user opinions of web pages are influenced by viewing time. The second is whether, within the context of web page design, the subjective terms 'encourage' and 'discourage' are polar opposites on a single scale.

The two studies were conducted at different stages of the research and have been presented in a single chapter for clarity. After completion of Study One there were a number of outstanding questions and areas worthy of further exploration, and so Study Two was conducted. The use of two studies also allowed for a comparison of the findings; however, any findings from this comparison should be treated with caution, as the second study was intentionally not a replication of the first.

3.2 Background

This section presents existing literature relating to visual attributes and user opinions of web pages.

Research by Tractinsky *et a*l. (2000), Ngo and Ch'ng (2001), van Schaik and Ling (2003) and Robins and Holmes (2008) covered in more detail in Chapter 2, all indicated links between user

preferences and aesthetic aspects of the visual screen design. The research described in this chapter addresses additional visual attributes within the context of web design, and contributes to the research in this field.

The remainder of this section describes the design decisions made in relation to the experimental design used in both studies described in this chapter. For each experimental design decision explored below each option has been discussed in turn, and finally, the chosen option stated.

3.2.1 A single scale of encouragement

Traditionally *discourage* is thought to be the polar opposite of *encourage*. If this were the case, you would expect to find a perfect negative correlation between ratings for how much a page encouraged you to explore the site and how much the page discouraged you. Therefore, it was anticipated that a reduction in encouragement would translate into an equal increase in discouragement. In e-commerce literature, *encourage* and *discourage* are often used as opposites in relation to decisions regarding purchase as evaluated by Chen and Dhillon (2002), O'Cass and Fenech (2003), and Rosen and Purinton (2004). In the domain of website design little research has been done on whether *discourage* should be considered a gradable antonym (two terms measured on the same scale) or a complementary antonym (two terms which complement each other but are not measured on the same scale) of *encourage*. This distinction depends on whether they can be considered opposing points on the same relational scale or not. Research into this concept in relation to web pages was reported by Rugg (2013) and found that web pages were not consistently either attractive or unattractive. In a small number of cases web pages were rated highly for both encouragement and discouragement. Bem (1974) found a similar pattern when exploring individuals exhibiting masculine or feminine personality traits. There is also an absence of research into which verbs can be considered gradable antonyms when gathering opinions of images (Wagner, 2005; Whitaker, 2005).

3.2.2 The effect of viewing time

Study One incorporated an investigation of the effect of viewing time on subjective responses. There is evidence to suggest decisions regarding whether or not to stay on a web page are made very quickly (Robins & Holmes, 2008; Briggs, 2002; Weinschenk, 2011). It has also been said that, for web pages, content is one of the most important aspects of the page (Nielsen, 1999; Chase, 1996; McCarthy, 1995; Rodriguez *et al.*, 2006). However, the decisions regarding staying on a page are made too quickly for a user to read the content of the page. This raises questions about how this decision is reached. This question has also been posed by Robins and Holmes (2008) who found that users make judgements about the credibility of a site too quickly to have based this judgement on the content.

Based on the research discussed above a question was posed regarding the effect of viewing time on user decisions regarding web pages. By manipulating the exposure time (viewing condition) in an experiment it is possible to explore what effect, if any, manipulation of viewing time has. The research was restricted to respondents' immediate reaction to web pages, and for this reason two distinct viewing times were used: one second and five seconds. These short time frames ensured respondents were reacting to the initial visual affect of the page, not the textual content. In Study One participants were arbitrarily assigned to a viewing condition in turn, as part of a between subject design. There was no reason to believe any participants would perform better or worse in a particular group due to prior experience or knowledge. The allocation was done with no interaction from the researcher.

3.3 Method

The research described in this chapter explores trends in visual attributes of high- and low-rated web pages through two independent studies. The process was carried out through two studies, each using a different set of web pages. Study One is presented first, with Study Two being presented later, in Section 3.6. During this chapter the screenshots used in Study One are referred

to as images and the screenshots used in Study Two are referred to as cards, in order to clearly distinguish between the two sets of web pages.

3.3.1 Design

The following section explores suitable techniques to address the research themes posed in this thesis.

Correlational studies are often used for investigating links between two variables. Topics that have been explored this way include the effect of violence in computer games (Anderson & Dill, 2000), the effectiveness of computer-aided instruction (McNulty *et al*., 2009), and links between computer use, cognitive styles, and academic achievement (Altun & Cakan, 2006). There has been some criticism of the use of correlational studies as some researchers incorrectly interpret correlation to show causation (Coe & Fitz-Gibbon, 1998; Reinhart *et al*., 2013).

In addition to correlational studies it is possible to use statistics to seek correlations between more than one set of data. One technique for this is called Principal Component Analysis and it is a type of multivariate analysis (Rees, 2001). Principal Component Analysis uses observations for multiple factors pertaining to an individual item and calculates what weighting of each factor is required to explain a correlation. One reason for this is to determine whether the number of factors being measured can be reduced. It also indicates which factors are most important. In order to use this technique, a large number of samples must be obtained. Principal Component Analysis was not considered appropriate for the research described in this thesis, due to the large number of observations needed for it to be valid.

A computational application of Principal Component Analysis is neural networks. In the field of artificial intelligence, neural networks are used to identify patterns and trends within data. One common area of application is image processing (Lippman, 1987). Neural networks are a type of processing intended to emulate that of the brain. Neuron processing techniques are particularly useful for forecasting as they can determine causality (McCulloch & Pitts, 1943) including

situations where there is non-linear correlation (Zhang *et al.*, 1998). Two types of neural networks are supervised and unsupervised. In a supervised network the algorithm needed to get from the input to the desired output is already known. In an unsupervised network the algorithm is not known, and the network learns or discovers the rule by using training data. It can be inferred that neural networks may be appropriate to identify objective predictors of subjective opinions, if a suitable training data set was available.

Having reviewed three possible avenues for investigating the proposed research topic the most appropriate technique was to conduct a correlational study. The following section reviews a similar study, which showed potential for further investigation.

A previous correlational study into user opinions of web pages was conducted by Holland (2004). This research obtained user opinions of a set of 100 university computer science department homepages. It investigated whether correlations existed between five subjective perceptions of web pages and a set of objective measures. The subjective information was gathered using visual analogue Likert-style scales. This work found statistically significant positive correlations between the complexity of the HTML file, and the number of unique colours used. It found negative correlations between the size of the image file, and the word count. Limitations of this study included the small sample size for the subjective data collection and the use of p values, instead of confidence intervals for the correlation testing. Confidence intervals are thought to be a more robust measurement of correlation (Davies & Crombie, 2009). The use of p values alone makes it more difficult to generalise the findings.

The first step towards identifying predictors of a user's emotional response to a web site is to identify common visual attributes within high- and low-rated web pages.

In order to identify high- and low-rated pages a number of options were considered; a usage review, identifying an existing set of high- and low-rated pages, and conducting a survey to gather this information.

The first option was to conduct a review of usage. This carries an implicit assumption that a highly-used page is therefore a high-rated page. This assumption is fundamentally flawed as not all decisions to consume a web page are based solely on the content of the web page itself. For example, online banking web pages are used to access a bank account online. In many cases an individual chooses their bank based on physical location, services on offer, previous experience or recommendation. It is not common for banks to offer a preview of the interface design and functionality of their site in advance of a customer joining the bank. The factors mentioned make it unlikely that, in the case of online banking, the design of the bank's website will factor into the decision regarding which bank to join, although it may influence a decision to change bank. The market share taken by a particular site within a domain could also have been reviewed, but similar issues to those discussed previously remained. It would be non-trivial to map all the potential external factors contributing to usage of a set of web pages, therefore usage statistics were not information to use to identify high- or low-rated pages.

A second option was be to identify popular and unpopular web pages, using an existing list. The distinction between popularity and usage is difficult to define, and no such list was available. Therefore, an existing set of popular and unpopular web pages was not used to identify high- and low-rated pages.

The third option was to conduct a survey to gather user ratings of a set of web pages. Surveys can be used to gather explanatory information about a particular phenomenon, which matches the question posed in this chapter. Surveys can be used when qualitative information, quantitative information or a combination of both, are sought. The type of information to be gathered and the data collection methods, are described later in this section. The survey method was used by

Holland (2004), a preceding study of a similar nature, discussed in detail in Chapter 1. This study gathered information on 100 web pages, however, each page was only reviewed by 8 participants. This very small sample size made it difficult to generalize the findings to a larger population and poses a threat to the validity of results found.

After reviewing the three options, the use of a survey was chosen to identify high- and low-rated pages from two sets.

3.3.1.1 Data collection method

The following sections provide a discussion of the data collection options, the option selected, the type of questions and response options used, the data collection medium, and how the influence of the order effect and framing were minimised.

The think-aloud method (Ericsson & Simon, 1985) is commonly used for website usability testing. This method is used to gather information about a site which can include user opinions, often while a task is being completed. It provides qualitative information about a site from a user perspective, which can then be analysed for content using a variety of quantitative and qualitative methods, for example using cognitive causal maps. A drawback to this approach is that participants use verbal fillers such as “um” and “er” when they are still processing the page mentally (Cooke, 2010), and although these fillers can be recorded, the processing going on at this time is not. Cooke (Ibid) found the content of verbalisations to fall into a small number of categories, which she labelled as *reading, explanation, procedure, observation* and *other*. Explanation only made up 5% of the verbalisations. For the research described in this chapter, the most relevant of these categories is explanation as it provides the reasons behind the rating given. Therefore the think-aloud method was not appropriate for use in this instance.

The following section describes the type of information gathered in order to identify the high- and low-rated pages.

The method of data collection used by Holland (2004) was a Likert-style visual analogue scale. This involves using a 101cm line with anchor points at either end, and the respondent adds a line at the point that relates to their level of agreement with the question or statement. There has, however, been literature, which suggests visual analogue scales offer more categories than the brain can discriminate (Linacre, 1998). This response method provides a number for each criterion, which can also be grouped with other individuals' responses to get the average subjective value. Using consistent anchor points for every question is preferable, as having different anchor points for each question, even with similar meanings, can lead to confusion, with subjects answering the question they think is being asked instead of the actual question.

The correlational study, described in Chapter 4, required quantitative information about user opinions of web pages. For this reason quantitative information was required. The option was available to also gather qualitative information. By asking open questions many possible reasons for the opinions would be gathered, however some may have been extrinsic reasons, which would not be useful in answering the questions posed in this chapter. To gather, transcribe, synthesise and analyse a large set of qualitative responses regarding user opinions of web pages was determined to be disproportional to the insight that could be gained.

In person data collection provides an opportunity for the researcher to answer questions or clarify misconception that the participants may have. However, it can be a time consuming process. When using a questionnaire, careful design and piloting can help to ensure the questions are easy to understand and can reduce the requirement for any interaction with the participants. Holland (2004) conducted the data collection in person and this decision may explain the small sample size. Online data collection, when using a questionnaire, can yield large samples relatively quickly and simply. There are also a number of opportunities to disseminate an invitation to take part, in a relatively short amount of time, such as email, and social media websites. Another benefit if

online data collection is that the results can be collected in a way that makes any data manipulation before analysis a swift process.

For the data collection described in this chapter, images of the web pages were used rather than the live sites. This allowed control over the length of viewing time for the pages and prevented the user from being able to interact with the pages.

The order in which the web pages were presented to the participants was considered, and a random presentation order, which differed for each participant, was determined to be the most appropriate option. The random order was achieved using the RAND() function within PHP.

User opinions can be affected by framing, that is to say by their exposure to other web pages prior to viewing the survey. To reduce the effect of previous websites and the page used to gather background information a masking image was used, in a similar way to a distraction task. The masking image was a collage of all the sites, tiled to fill the majority of the screen.

Participants were asked to provide responses to 6 questions for each set of images using a visual analogue Likert-style scale. The mean of the values provided was calculated for each site, for each question. The mean values were then used to rank the sites. From this, a set of high-rated and low-rated pages was gathered for each study. These pages were then analysed for common visual attributes that could have led to the high or low rating. The discussion of the attributes identified can be found in Section 3.5.1.1.

In order to investigate whether *encourage* and *discourage* can be considered to be opposing points on a scale of encouragement, two questions were used. These questions were: *How much does this page encourage you to explore the rest of the site?*, and *How much does this page discourage you to explore the rest of the site?*. If *encourage* can be considered a gradable antonym, it would be expected for the results of one to be the inverse of the other, for each participant's rating.

To test this hypothesis the Pearson Product Moment Correlation Coefficient was used. A discussion of various correlation coefficient options and their suitability for different types of information is discussed in Section 4.3.5. In a situation where encourage and discourage are polar opposites on a single scale, a negative correlation would be expected. Therefore, a one-tailed test, looking for a negative correlation, was used. The two hypotheses were:

H_0: There is no correlation between responses for Question C and Question F

H_1: There is a negative correlation between responses for Question C and Question F

3.3.2 Participants

Participants were recruited via convenience sampling (Gelo *et al.*, 2008) and snowball sampling (Goodman, 1961). An invitation to take part was circulated via email and via a social network site, Facebook. 237 sets of data were collected. After reviewing the data and removing incomplete sets, 86 sets of complete data were analysed. There were practical issues in accessing the demographic information of participants and the focus of the analysis is non-demographic.

3.3.3 Materials

The text content of a web page is one element that is thought to influence user opinions of web pages. It was decided that control of this variable was a key requirement for the set of pages used in Study One. *CSS Zen Garden* is a website that sets out to demonstrate "what can be achieved through CSS-based design" (Shea, 2006). *CSS Zen Garden* has an HTML template and each design links the HTML file to a different cascading style sheet (CSS). This results in consistent text with different styling. The use of web pages from this site enabled the researcher to control the text content.

The materials on *CSS Zen Garden* were collated and published (The Zen of CSS Design: Visual Enlightenment for the Web, Shea & Holzschlag, 2005). The pages used in the research described in this chapter are a sample of the pages in the book. Designs featured in *The Zen of CSS Design*

are in six categories: design, layout, imagery, typography, special effects and reconstruction. For this research all the pages from design, layout, imagery and typography were used. The full set of images used in the research described in this chapter can be found in Appendix A.

The number of pages included in Study One was 24. The decision on the number of pages was determined, mindful of the length of time participants would be willing to spend on data collection. The number of pages also needed to be large enough to draw meaningful conclusions from the correlational study described in Chapter 4.

Screen shots of the web pages were taken, and these images were displayed through a website created by the researcher. The images were acquired by taking screen shots at 1024x800 resolution. Using Paintshop Pro the images were cropped to display in a web browser to approximately fill the screen when viewed in a web browser if using 1024x800 resolution. It is understood that not all users will have the same resolution, but this was an acceptable compromise, compared to the viable number of participants estimated if using a controlled environment, such as a consistent machine for all participants.

The questions used for data collection are shown below and the rationale behind the questions is discussed in Section 3.3.3.

Question A (QA) - How eye-catching do you find this page?

Question B (QB) - How visually attractive do you find this page?

Question C (QC) - How much does this page encourage you to explore the rest of the site?

Question D (QD) - How well designed is this page?

Question E (QE) - How good is this page as an advertisement for the website?

Question F (QF) - How much does this page discourage you to explore the rest of the site?

The response for each question was gathered on a unipolar visual analogue scale (VAS). These are a type of continuous response scale and information of this type can be treated as continuous data (Allen & Seaman, 2007). This allowed use of descriptive statistics such as mean and standard deviation in the discussion section. The questions are the same as those used by Holland (2004). This allowed for a comparison between the findings in that study, and those in this thesis. The phrasing of some of the questions is not as clear as it could be, particularly question E, however modifying the questions would have limited the ability to compare the results.

3.3.4 Procedure

A specifically developed online tool was used for the data collection. The researcher carried out the design and creation of the tool. It was created using HTML and PHP with a MySQL database to store the results. The tool allowed for the 24 images to be displayed in turn to participants for a specified length of time, according to the allocated viewing condition. The participant's responses to the six questions, found in Section 3.3.3, were gathered and recorded. The results were downloaded from the server in CSV format, subsequently allowing them to be analysed using Microsoft Excel.

The data collection was completed online by the participants with no direct interaction with the researcher. Participants completed the data collection on their own computer, using their usual browser and resolution settings. This ensured the experience was consistent with their usual browsing experience. The experience may not have been consistent between users.

An instruction sheet was shown which explained the task. Participants first completed a background questionnaire with 5 questions, which can be found in Appendix B. The answers from this were intended to be used for grouping results during analysis. A masking image was then displayed. The masking image was a mosaic of all 24 images, tiled 240 times, and can be seen in Appendix C. Participants were then shown an image of a web page for either one or five-seconds,

followed by a screen where the six questions were displayed on the screen in unison. The same set of questions was used for every page.

The images were shown in a random order for the reasons discussed in Section 3.3.1.1 and the viewing condition was set to either one or five seconds consistently for each participant. The condition was assigned when the background questionnaire was submitted, with the viewing condition being assigned in turn. The answers to all questions were stored in a database and exported to a spread sheet for analysis. Any data sets where the participant did not provide responses for all 24 images were removed before analysis.

In order to calculate the subjective rating of each page the mean was used. The use of this type of average does contribute to some smoothing out of the individual results, but also allows for all the individual results to have an effect on the calculation of the average. The type of data being collected was ordinal data ranging from 0 to 100 gathered using a Likert-style visual analogue scale. Data of this type can be considered continuous. In this situation the use of the mode would have been inappropriate, due to the large number of possible outcomes. The participants were not selecting individual values, they were selecting a position along a line. The median was also not appropriate, as it does not fully take into account the varying responses of the other participants.

3.4 Results

This section presents the results of the data collection regarding subjective opinions of web pages from the two studies described in this chapter. It also provides results from each viewing condition for Study One, and from the correlation test between the level of encouragement and discouragement perceived in response to each card.

3.4.1 Study One: CSS Zen Garden

3.4.1.1 Subjective ratings for images

The following two tables show the mean subjective rating for each question asked, for each web page. The means have been calculated using information from 86 participants.

Image	Mean Qa	St Dev Qa	Mean Qb	St Dev Qb	Mean Qc	St Dev Qc
1	68.44	22.64	64.60	23.89	56.27	24.04
2	54.40	23.89	58.24	24.68	48.36	23.64
3	45.09	23.78	58.60	23.49	49.87	22.77
4	45.67	24.98	40.52	24.86	36.12	22.22
5	48.10	25.60	48.81	25.20	47.24	24.40
6	55.27	24.11	54.76	25.64	49.51	24.64
7	41.80	20.83	46.33	22.00	43.49	23.24
8	60.03	23.48	57.53	25.60	52.88	24.79
9	53.04	25.01	56.63	23.60	51.55	25.23
10	69.69	23.79	58.81	26.39	51.19	28.13
11	57.59	25.67	60.05	23.96	52.44	23.63
12	39.14	23.55	40.18	25.15	42.60	27.37
13	54.81	23.12	57.64	24.40	50.80	23.60
14	67.35	21.23	55.94	25.92	50.06	26.85
15	57.74	25.45	56.75	25.38	49.33	22.57
16	44.25	23.68	47.64	23.31	45.54	24.42
17	44.64	24.32	49.02	22.66	41.16	23.68
18	52.94	25.04	41.84	25.75	39.18	25.27
19	51.35	25.42	59.07	22.54	49.51	23.20
20	37.44	22.28	31.06	20.20	28.12	17.52
21	59.84	24.70	41.24	24.49	40.20	24.29
22	31.10	20.66	36.20	22.71	36.41	22.59
23	39.55	24.52	27.54	23.38	23.72	22.52
24	52.57	21.10	53.65	21.99	46.24	21.93

Table 1 Mean and standard deviation for all 24 web pages (questions A-C)

Image	Mean Qd	St Dev Qd	Mean Qe	St Dev Qe	Mean Qf	St Dev Qf
1	55.78	22.48	55.47	25.31	24.88	19.51
2	51.86	23.72	49.36	22.42	26.21	23.12
3	54.92	22.13	47.66	25.34	28.51	23.25
4	37.13	20.91	34.64	22.77	35.81	29.24
5	49.44	24.23	42.72	22.50	33.18	23.88
6	48.42	24.28	45.37	25.10	33.87	27.29
7	49.69	22.36	40.92	21.75	31.96	26.10
8	50.70	25.19	48.13	23.75	29.28	23.37
9	52.51	26.16	50.42	24.82	27.29	24.76
10	52.02	25.28	52.24	26.75	30.90	27.34
11	53.14	23.46	51.37	26.20	28.39	25.32
12	46.89	27.51	39.21	24.70	34.84	25.92
13	51.13	22.28	49.91	22.07	25.79	21.58
14	49.51	24.47	48.86	25.43	33.13	26.97
15	51.87	22.77	49.24	24.32	29.06	25.24
16	48.64	24.50	42.49	23.90	33.00	23.28
17	44.54	22.15	40.06	23.00	32.22	25.72
18	41.76	23.55	36.93	24.16	40.70	30.61
19	48.91	22.94	47.32	24.45	28.06	24.23
20	29.73	18.31	28.68	18.02	50.37	29.75
21	40.71	23.69	38.22	24.24	43.15	30.25
22	41.38	21.28	31.23	19.36	40.10	29.67
23	23.28	20.87	21.71	18.54	61.79	28.86
24	49.29	21.84	47.35	20.93	30.99	25.10

Table 2 Mean and standard deviation for all 24 web pages (questions D-F)

The following graph shows the mean values for each of the 24 images. The graphical representation shows the spread of ratings for each question, using the mean rating value for each image.

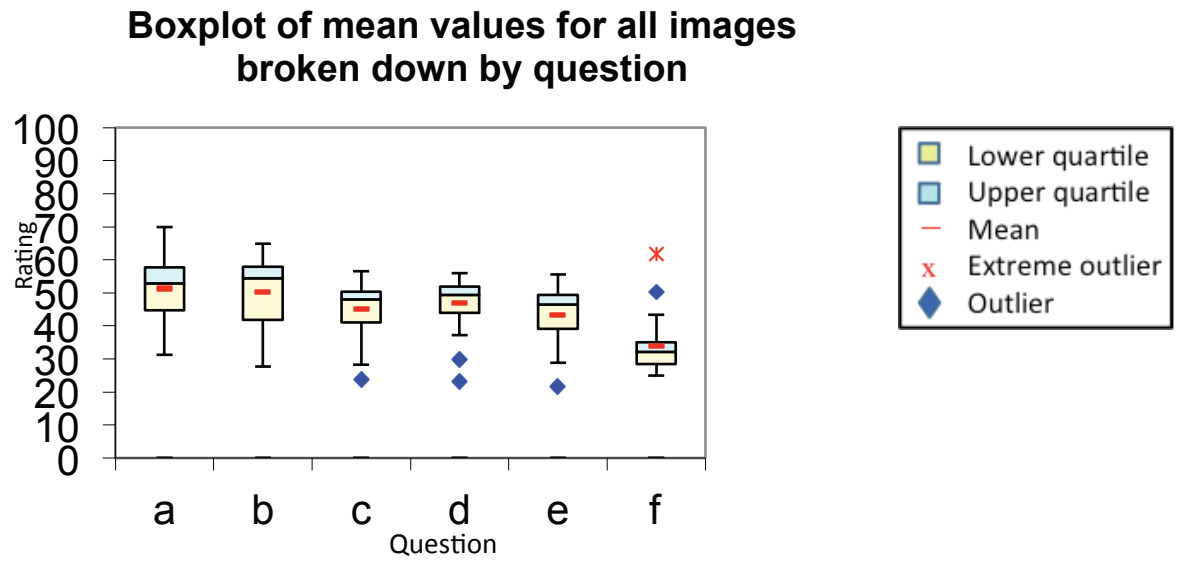


Figure 1 Boxplot of mean values for all images broken down by question

Addressing questions A-E first, the boxplot representation shows all the mean value for the set of images, for each question. The mean is indicated by the red line, and is between 40 and 55. The mean value is below the median for all five questions. The lower quartile is larger than the upper quartile for all five questions. There are outliers for Question C, Question D and Question E, which all lie below the rest of the values. When looking at the plot for Question F, the mean is above the median, and the upper and lower quartile ranges are similar. There are outliers for Question F, including one extreme outlier, indicated by the red asterisk. The outliers for Question F all lie above the rest of the values.

3.4.1.2 Subjective ratings split by condition

The following six graphs show the mean values from the subjective data collection on *CSS Zen Garden* web pages, broken down by viewing condition. Tables showing the mean and standard deviation for each question, broken down by image, for overall and each viewing condition, can be found in Appendix J. The first viewing condition was one second, and the second viewing condition was five seconds. The rationale for the timings can be found in Section 3.2.6.

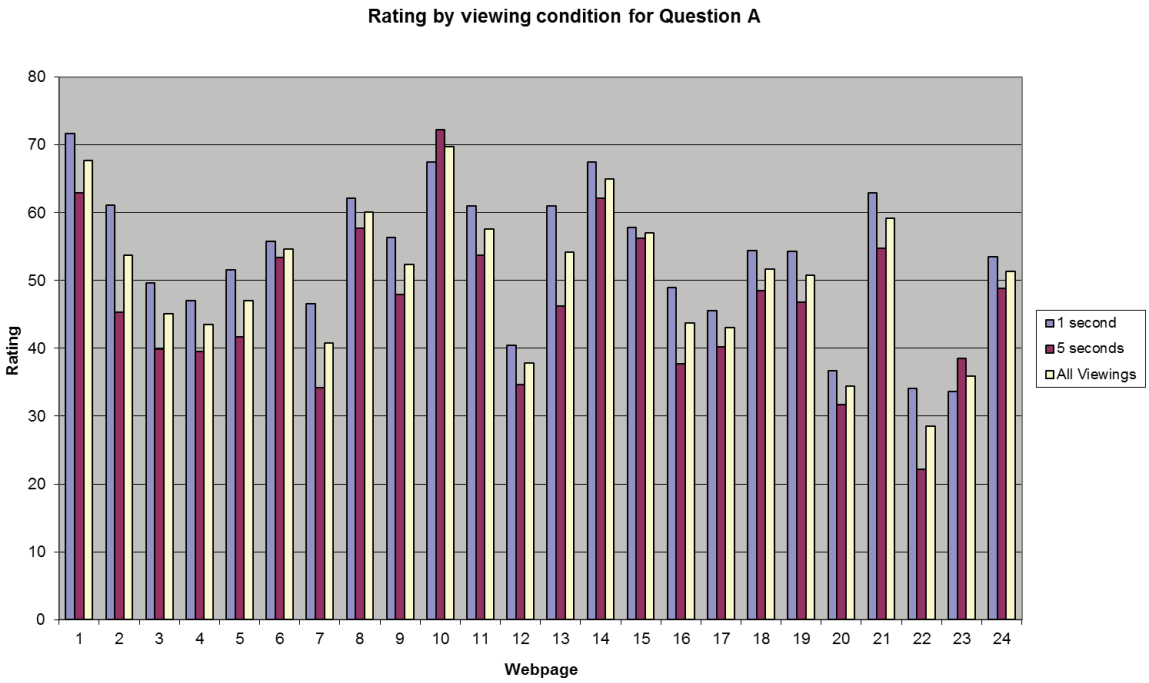


Figure 2 Mean rating for each image Question A - How eye-catching do you find this page?

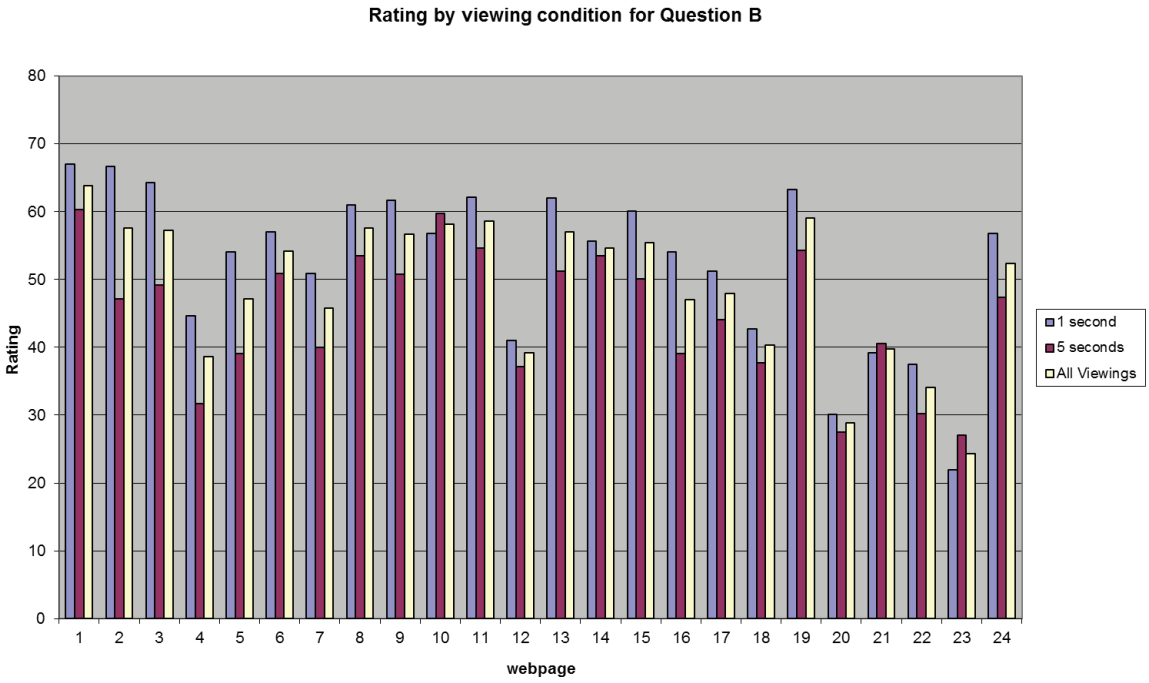


Figure 3 Mean rating for each image Question B - How visually attractive do you find this page?

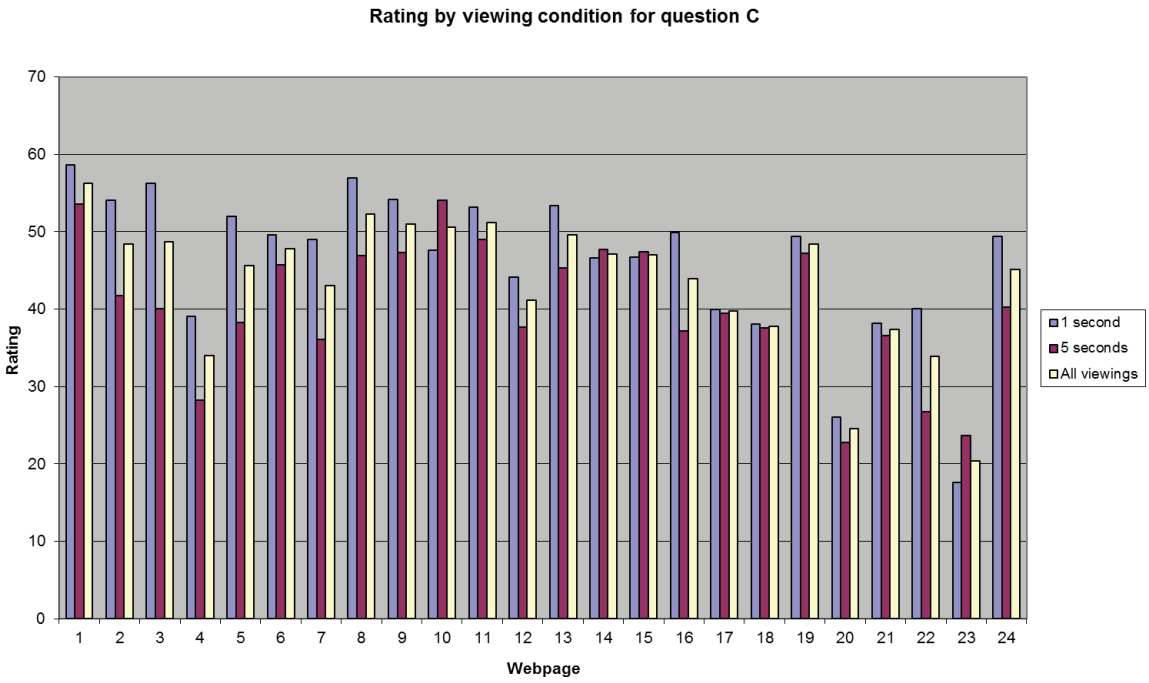


Figure 4 Mean rating for each image Question C - How much does this page encourage you to explore the rest of the site?

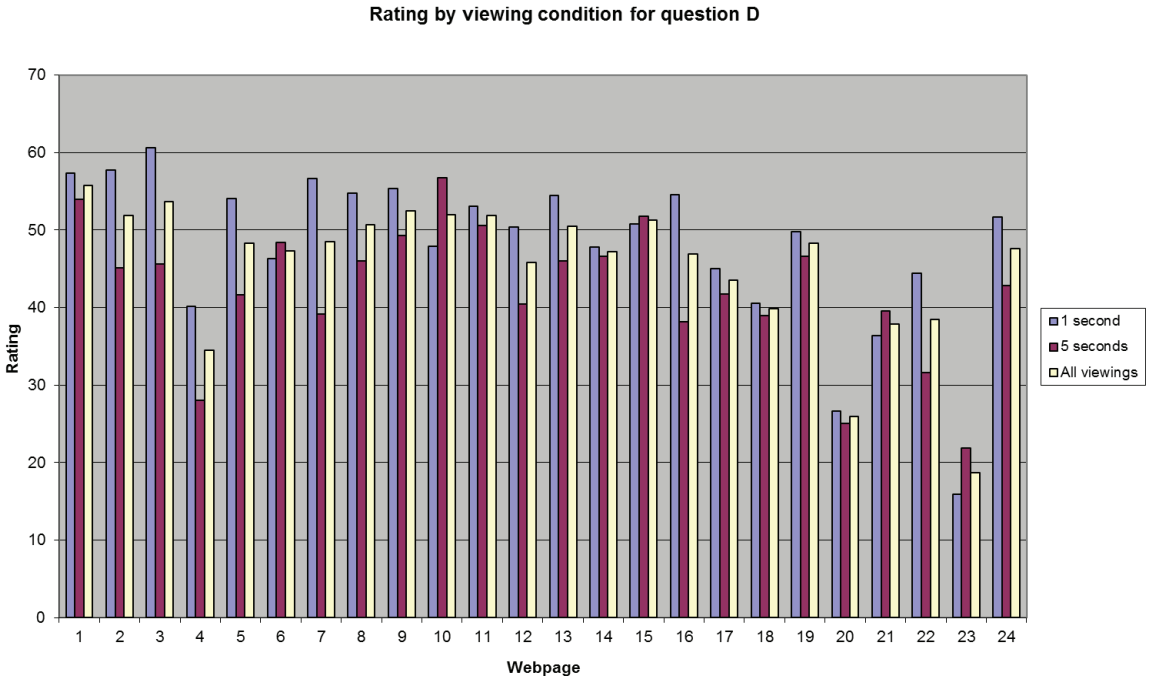


Figure 5 Mean rating for each image Question D - How well designed is this page?

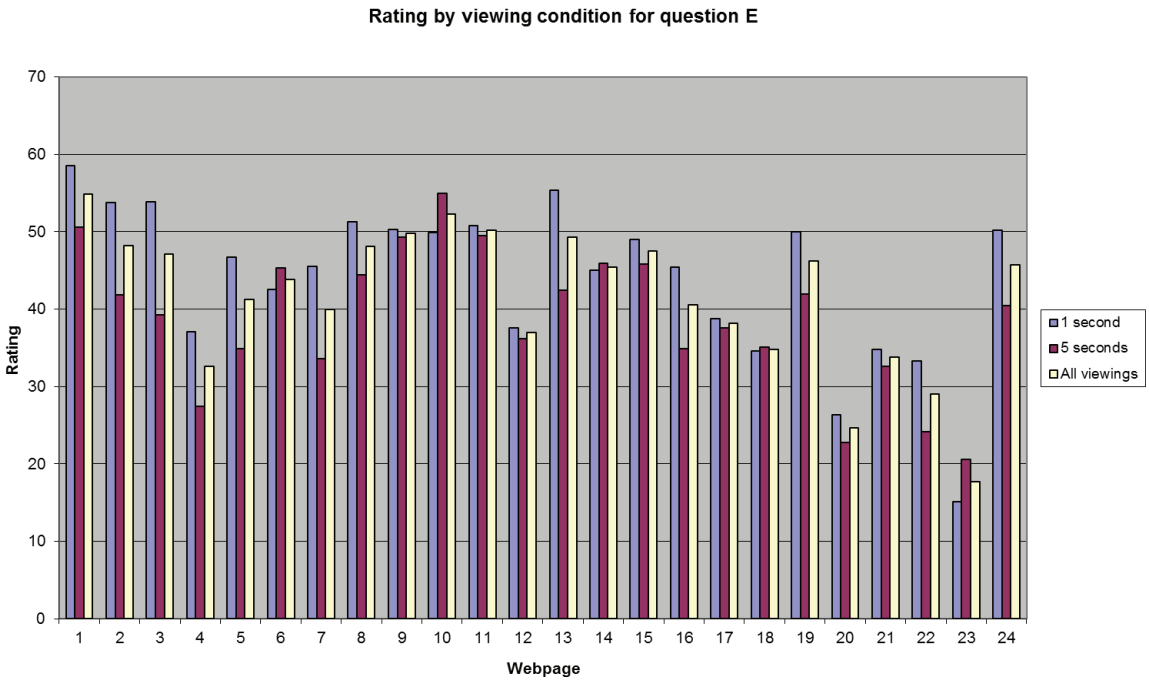


Figure 6 Mean rating for each image Question E - How good is this page as an advertisement for the website?

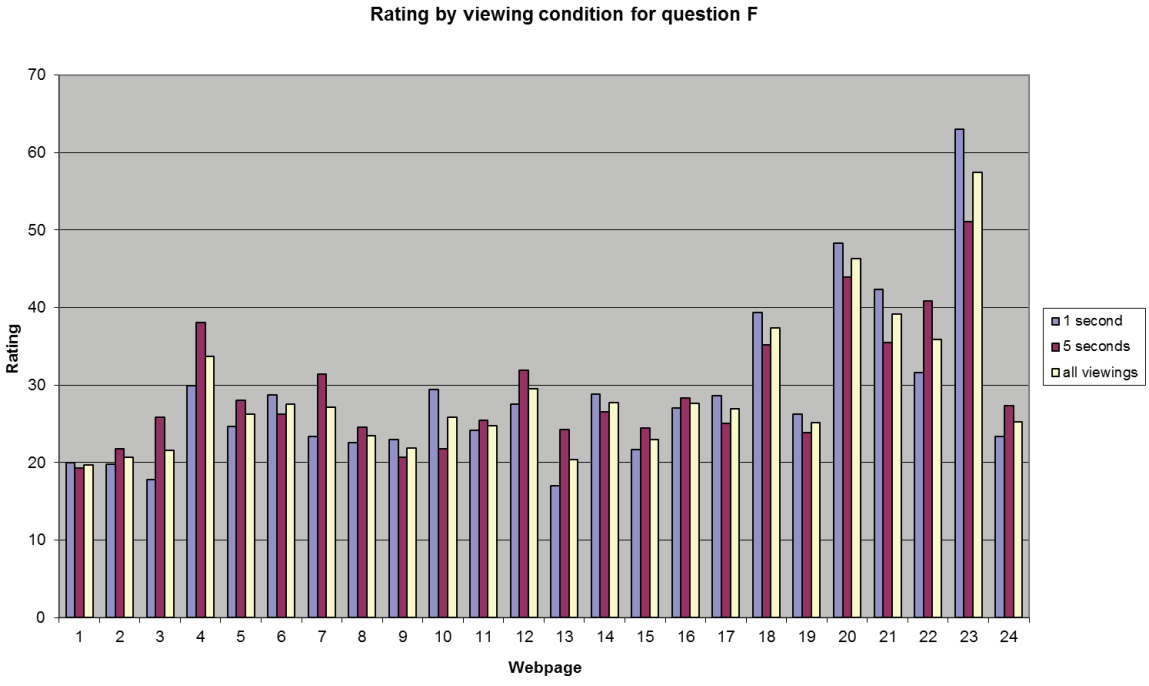


Figure 7 Mean rating for each image Question F - How much does this page discourage you to explore the rest of the site?

3.4.1.3 Correlation test for encourage and discourage

The following table shows the value for r for correlations between Question C and Question F, for each image. This was calculated using Pearson's correlation coefficient. It used the individual participant responses for every question.

It can be seen that the correlation is statistically significant for all images, for all responses, regardless of viewing time. When looking at viewing condition A in isolation the correlation is statistically significant for 20 out of 24 images. When looking at viewing condition B in isolation the correlation is statistically significant for 9 out of 24 images.

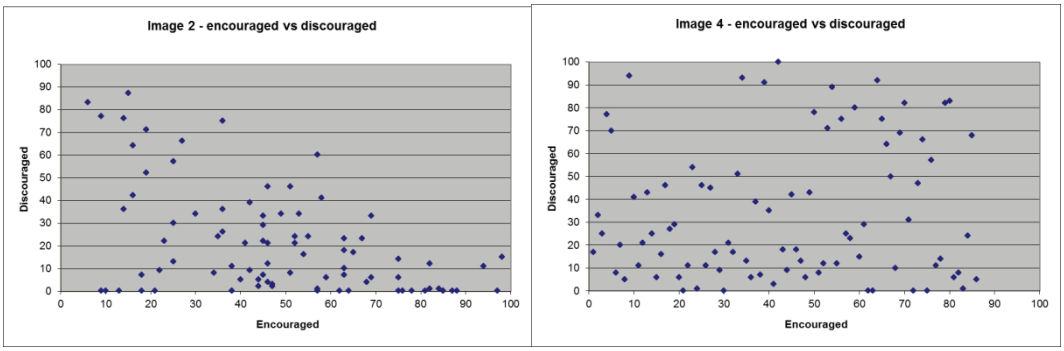


Figure 8 Example scatter graph

Scatter graphs for the correlations have been produced. Figure 8 above shows two examples, the full set can be found in Appendix K. The graphs have been presented with a large number visible on each page, to allow identification of patterns across the set.

	Correlation Coefficients		
Site	All Conditions	Condition A	Condition B
1	-0.471724235	-0.41	-0.546358822
2	-0.485363897	-0.59618273	-0.40095652
3	-0.49867475	-0.51834757	-0.440462399
4	-0.472183546	-0.54378309	-0.372012573
5	0.807236683	0.763639734	0.836465308
6	-0.480999011	-0.4661691	-0.513228367
7	-0.412341893	-0.35567236	-0.426394446
8	-0.525193646	-0.6621287	-0.330600853
9	-0.333907813	-0.49423475	-0.221504424
10	-0.553368357	-0.52848162	-0.565013351
11	-0.40706489	-0.51541258	-0.295667316
12	-0.528625029	-0.48846839	-0.559986623
13	-0.333922486	-0.28941057	-0.342263971
14	-0.494210712	-0.48643782	-0.502317272
15	-0.519210155	-0.52442838	-0.524955276
16	-0.339658344	-0.49512535	-0.199845625
17	-0.405003136	-0.45948324	-0.32449689
18	-0.536837436	-0.50341206	-0.589870677
19	-0.304733417	-0.39146921	-0.229257263
20	-0.447734476	-0.5861727	-0.310685638
21	-0.482227771	-0.50310518	-0.474369094
22	-0.433540423	-0.34876863	-0.492985341
23	-0.523250946	-0.61593056	-0.420763731
24	-0.31357846	-0.45473078	-0.142615816
	Significant at 0.0005		
	0.294	0.408	0.49
	n=86	n=46	n=40
	df=84	df-44	df-38

Table 3 Pearson's correlation coefficient for each image based on all users

3.5 Discussion

This section discusses the results presented for each study and draws attention to results which address the questions posed at the beginning of the chapter. These findings are placed within the context of the literature, and unexpected findings have been explored.

3.5.1 Study One: CSS Zen Garden

3.5.1.1 Subjective ratings for all pages

	Question a		Question b		Question c		Question d		Question e		Question f	
Rank	Website	Score	Website	Score	Website	Score	Website	Score	Website	Score	Website	Score
1	10	69.69	1	64.60	1	56.27	1	55.78	1	55.47	23	61.79
2	1	68.44	11	60.05	8	52.88	3	54.92	10	52.24	20	50.37
3	14	67.35	19	59.07	11	52.44	11	53.14	11	51.37	21	43.15
4	8	60.03	10	58.81	9	51.55	9	52.51	9	50.42	18	40.70
5	21	59.84	3	58.60	10	51.19	10	52.02	13	49.91	22	40.10
6	15	57.74	2	58.24	13	50.80	15	51.87	2	49.36	4	35.81
7	11	57.59	13	57.64	14	50.06	2	51.86	15	49.24	12	34.84
8	6	55.27	8	57.53	3	49.87	13	51.13	14	48.86	6	33.87
9	13	54.81	15	56.75	19	49.51	8	50.70	8	48.13	5	33.18
10	2	54.40	9	56.63	6	49.51	7	49.69	3	47.66	14	33.13
11	9	53.04	14	55.94	15	49.33	14	49.51	24	47.35	16	33.00
12	18	52.94	6	54.76	2	48.36	5	49.44	19	47.32	17	32.22
13	24	52.57	24	53.65	5	47.24	24	49.29	6	45.37	7	31.96
14	19	51.35	17	49.02	24	46.24	19	48.91	5	42.72	24	30.99
15	5	48.10	5	48.81	16	45.54	16	48.64	16	42.49	10	30.90
16	4	45.67	16	47.64	7	43.49	6	48.42	7	40.92	8	29.28
17	3	45.09	7	46.33	12	42.60	12	46.89	17	40.06	15	29.06
18	17	44.64	18	41.84	17	41.16	17	44.54	12	39.21	3	28.51
19	16	44.25	21	41.24	21	40.20	18	41.76	21	38.22	11	28.39
20	7	41.80	4	40.52	18	39.18	22	41.38	18	36.93	19	28.06
21	23	39.55	12	40.18	22	36.41	21	40.71	4	34.64	9	27.29
22	12	39.14	22	36.20	4	36.12	4	37.13	22	31.23	2	26.21
23	20	37.44	20	31.06	20	28.12	20	29.73	20	28.68	13	25.79
24	22	31.10	23	27.54	23	23.72	23	23.28	23	21.71	1	24.88

Table 4 Images in rank order

The table above presents the mean rating values for each image in rank order. This allows a comparison across all questions. The following table summarises the highest and lowest rated five images for each question, to allow an at-a-glance comparison of the images in each of these sets.

	Highest rated five images	Lowest rated five images
How eye-catching do you find this page?	10, 1, 14, 8 and 21	22, 20, 12, 23 and 7
How visually attractive do you find this page?	1, 11, 19, 10 and 3	23, 20, 22, 12 and 4
How much does this page encourage you to explore the rest of the site?	1, 8, 11, 9 and 10	23, 20, 4, 22 and 18
How well designed is this page?	1, 3, 11, 9 and 10	23, 20, 4, 21 and 22
How good is this page as an advertisement for the site?	1, 10, 11, 9 and 13	23, 20, 22, 4 and 18
How much does this page discourage you to explore the rest of the site?	23, 20, 21, 18, and 22	1, 13, 2, 9 and 19

Table 5 Highest and lowest rated images when in rank order based on mean rating

In the following section the images that have appeared in the highest or lowest rated five images have been reviewed for common visual features, in order to address the main question posed in this chapter: What visual aspects of a web page affect user ratings?

The following figure shows screenshots of all pages present in the highest five images, for one or more question.

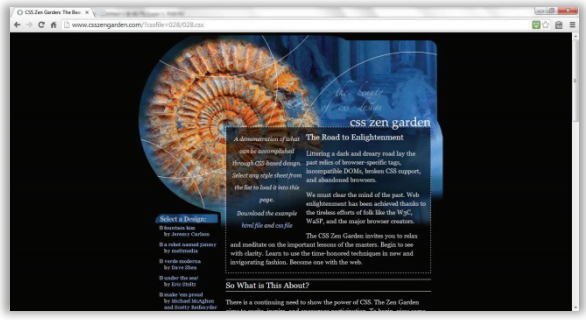


Image 1

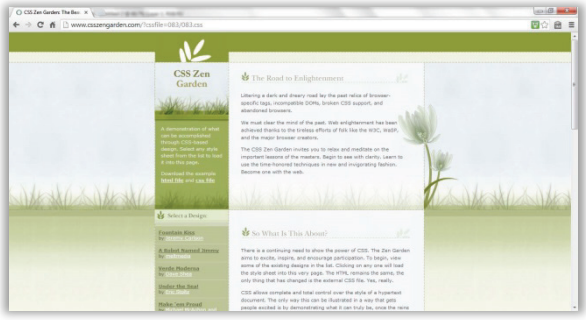


Image 3

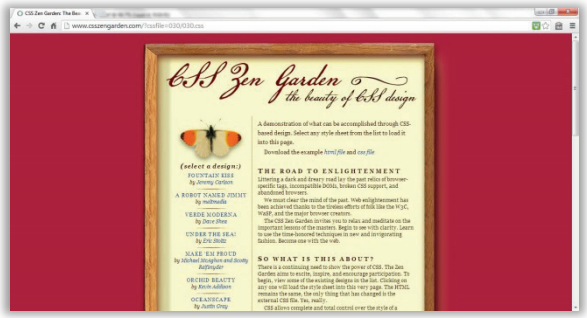


Image 8

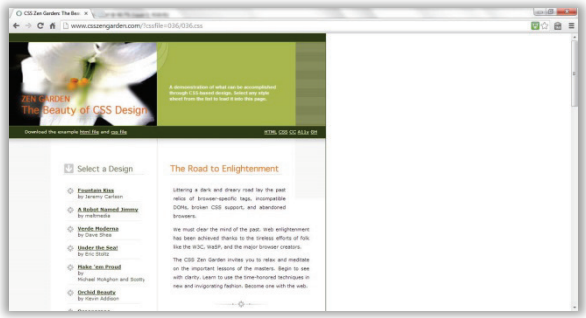


Image 9

Image 10

Image 11

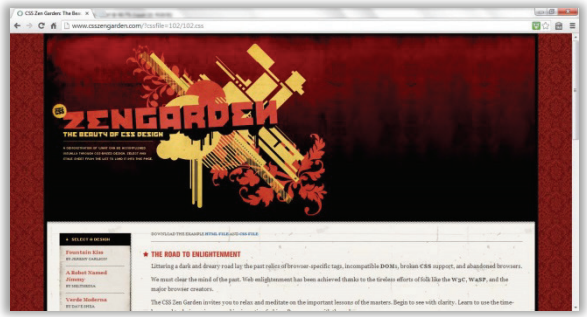


Image 14

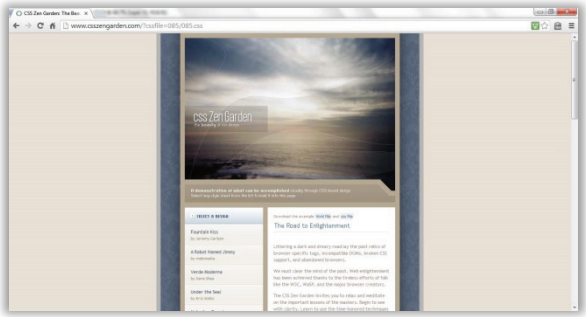


Image 19

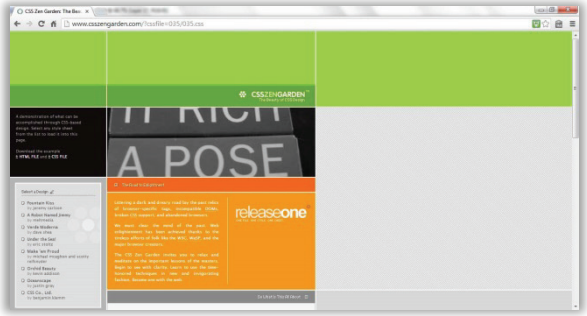


Image 21

Figure 9 Images rated in the highest five for Question A-E

By reviewing the cards in rank order it can be seen that some images were favoured over others. The highest rated five pages for every question contain at least one example of both a dark background with light coloured text, and a light background with dark coloured text. This shows that one is not consistently preferred over the other.

The highest rated five pages for QA (How eye-catching do you find this page?) included two pages with deep red backgrounds (Image 8 and Image 14 in Figure 9 above), which suggests either the colour red or the intensity of colour may be a factor related to how eye catching a page is. These were the only pages within the sample that had deep red backgrounds. These two pages did not appear in the top five for any other question.

Another colour which is present in multiple pages in the top five for most of the questions is the colour green. The shade of green and where it occurs, as a background or feature colour, is less consistent. A more detailed discussion of colour psychology and the research linked to these theories in a number of domains including web design can be found in Chapter 2.

Image 1 is an anomaly as most other pages which have a less obvious contrast between text and background have scored badly. The inclusion of an image in the background also detracts from the main text. It is unclear what this page does that makes it score well when others that have similar features do not.

The following figure shows screenshots of all pages present in the lowest five images, for one or more question.

Image 4

Image 7

Image 12

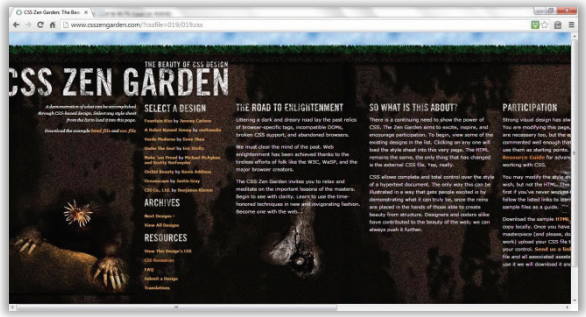


Image 18

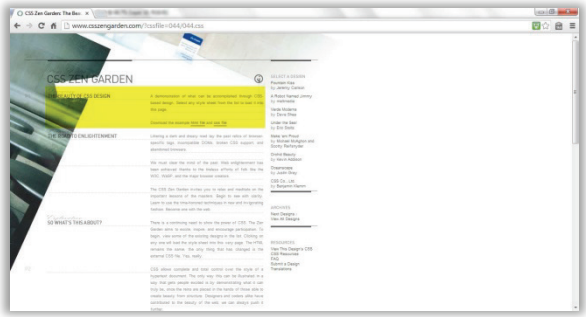


Image 20

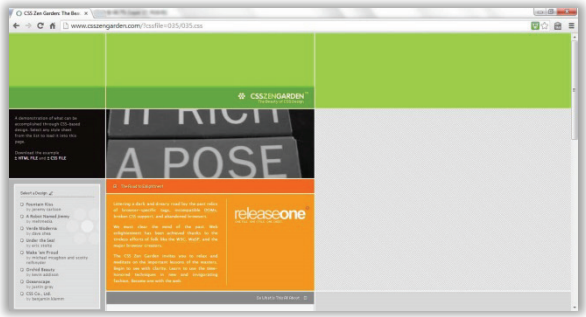


Image 21

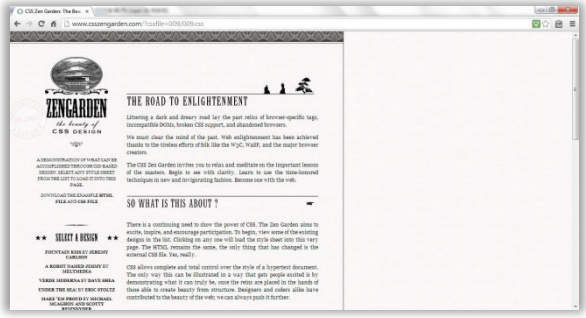


Image 22

Image 23

Figure 10 Images rated in the lowest five for Question A-E

Looking at the sites that appear in the lowest five for Question A to Question E it can be seen that four of the pages have a light coloured background. This also featured in the highest rated set of pages making it unclear whether this is a contributing factor to high or low ratings.

Two of the pages in the lowest rated set (shown in Figure 10 above) have the main text on background images. In contrast, one page in the highest rated five (Image 1 shown in Figure 9 above) also has a background image, but in this instance the text is contained within a text box which appears to overlay the image. This makes the separation of the body text and image clearer than in the case of Image 18 and Image 23 (shown in Figure 10). It may be that the level of contrast between the text and the image causes issues with readability. High contrast between text and background is a commonly cited web accessibility principle (WCAG, 2008).

When comparing the web pages that appear in the lowest rated five it can be seen that three pages occur in this category for every question (the highest rated five in the case of Question F). These were Image 23, Image 20 and Image 22 (shown in Figure 10 above). These pages all came from the same category in *The Zen of CSS Design*: Typography.

It is interesting to note that Image 21 occurs in the highest rated five for Question A (How eye-catching do you find this page?) and in the lowest rated five for Question D (How well designed is this page?). This suggests that an eye-catching page is not necessarily considered to be a well-designed page. This calls into question the hypothesis, discussed in Section 3.5.1.4 below, that each of the pages was rated based on the same criteria, without taking the different questions into account.

After review of the high- and low-rated pages the key findings were that high-rated images often contained the colour green. Images containing deep red were rated highly for Question A (How eye-catching do you find this page?). Low-rated images often displayed a low level of contrast

between the background and the text. The findings regarding the intensity of the background colour were mixed.

3.5.1.2 Subjective ratings split by condition

In order to analyse the effect of differing viewing conditions the mean ratings for the images were calculated for each condition. A summary table of the differences can be found below and the tables showing the values can be found in Appendix L.

	Number of instances where rating for condition a is higher than for condition b (a>b)	Exceptions to a>b
QA	22	10,23
QB	21	10, 21, 23
QC	20	10,14,15,23
QD	19	6,10,15,21,23
QR	19	6,10,14,18,23
QF	13	1,6,9,10,14,17,18,19,20,21,23

Table 6 Instances where rating for condition A is higher than for condition B

For Question A through to Question E it can be seen that ratings of the sites are higher for viewing condition A (one second) than viewing conditions B (five seconds) in the majority of cases. A possible explanation for this is that if a user does not have long enough to make an informed judgement there is a positive tendency present. Image 10 and Image 23 have a higher rating when viewed for a longer period of time (five seconds) for all six questions. The images were reviewed for common attributes such as similar background colours, use of colours, use of images, or similarities of font, but no common visual factors were apparent. When considering whether the sites were ranked in the highest or lowest five (shown in Table 4) Image 10 appears in the highest rated five, and Image 23 appears in the lowest rated five for Question A through to Question E.

Image 10 Image 23

Figure 11 Site 10 and Site 23

When considering all the images when ordered by rank, as shown in Table 4, it can be seen that image 10 appears within the highest rated five for all questions under viewing condition B (five seconds), however it only occurs once under viewing condition A (one second). When considering all viewing times Image 10 appears five times. This suggests immediate opinions for the image are not as high as if the user has time to take in the entire page. In contrast, Image 13 (shown in Figure 12 below) is in the highest five for all questions under viewing condition A, and does not appear in the highest five for any question under viewing condition B. This shows the image is rated higher when only seen for a short time, and the rating is lower if the user has a longer time to make the decision. Image 13 has a lot of white space on one side, which could be a contributing factor to the lower rating when users have longer to make their decision.

Image 13

Figure 12 Image 13

3.5.1.3 Correlation test for encourage and discourage

Using Pearson's correlation coefficient and looking at all the pairs of data (regardless of which image it is for) the value for r is -0.44515. This show a statistically significant correlation and provides a p value of p>0.0005 and allows us to reject the null hypothesis. The value allows us to accept the hypothesis that there is a negative correlation between QC (How much does this page encourage you to explore the rest of the site?) and QF (How much does this page discourage you to explore the rest of the site?). It is interesting to note that although this is a statistically significant correlation it is not a perfect correlation. A value for r of approximately -0.8 or beyond (allowing for noise) would show a perfect correlation. The implication of this finding is that although they are correlated there must be something else contributing to the values, rather than being able to treat them as opposite ends of a scale of encouragement.

The relative merits of p values and confidence intervals have been discussed in Section 4.3.5. The correlation between Question C (How much does this page encourage you to explore the rest of the site?) and Question F (How much does this page discourage you to explore the rest of the site?) was -0.45; the 95% confidence interval was -0.48 to -0.41.

The scatter graphs showing the correlation between Question C (How much does this page encourage you to explore the rest of the site?) and Question F (How much does this page discourage you to explore the rest of the site?) are shown in Appendix K. The graphs have been presented with a large number visible on each page, to allow identification of patterns across the set. If a line was drawn on each of the graphs to represent a direct negative correlation, interestingly all the graphs show the results either a clear cluster in the bottom left triangle under the line, or appear random. Five images, Image 1, Image 2, Image 8, Image 9, Image 13, show a clear cluster in the bottom left triangle. Four images, Image 20, Image 21, Image 22, Image 23, appear random. The remaining fifteen images have over half their results in the bottom left

triangle, but display a less obvious clustering. There are no instances where the results cluster in the top right triangle above the line.

A graphical representations of this, including the line of direct correlation, can be seen in Figure 13 below. The left image shows clustering in the bottom left triangle, and the right image shows an example with less clear clustering. With the addition of the direct correlation line the right image does show over half the results are in the bottom left triangle.

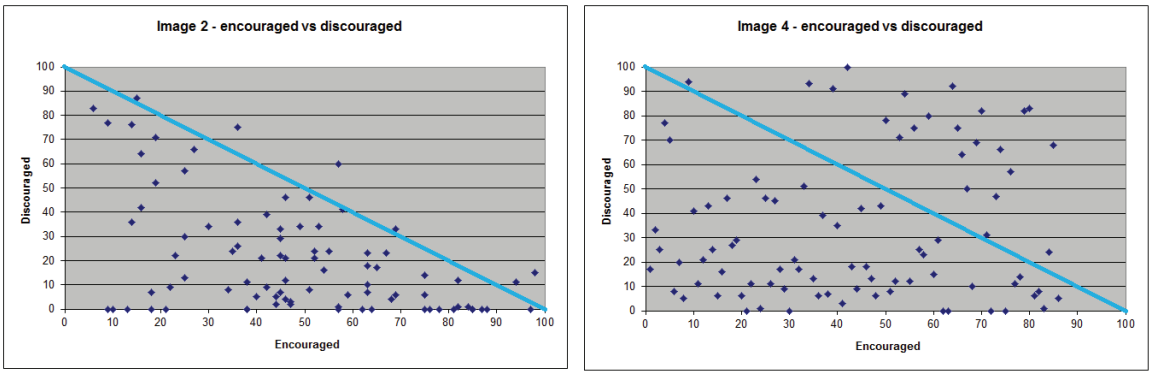


Figure 13 Scatter graphs of encourage and discourage for Image 2 and Image 10

This clustering matching the left hand image shows few pages rated highly for both encouragement and discouragement, as you might expect. However, contrary to the expected result of a negative correlation, there are a number of pages that rated low for encouragement and low for discouragement. It is possible that there are multiple factors contributing to web page user encouragement and discouragement, and when the question is framed differently an individual's emotional response to it may be different. It also shows that a page that does not encourage may also not discourage. However, it is also possible that the question was not entirely clear. Kahneman (1973) proposes a similar concept in relation to decision-making and rationality. When the same question is framed positively and negatively individual's emotional responses are often not directly proportional. An example of this is when explain the risk level of an operation "The statement that "the odds of survival one month after surgery are 90%" is more

reassuring than the equivalent statement that “mortality within one month of surgery is 10%” “(Kahneman, 2011, pp.88).

This also links to the concept of voting systems and the differing results if asked to vote for preferred items or least preferred items. This can be seen in many popular reality shows with voting, such as Big Brother. If the results of a vote for who should stay were gathered, and the inverse, a vote for who should leave were also gathered, the resulting lists are often thought not to be the inverse of each other. In reality, it is rare for both sets of information to actually be gathered, making a systematic comparison impractical. There are often characters in the show that are not favoured, but are also not offensive, and they end up with few votes whichever question is asked. It is conceivable that some web pages are not favoured, but are also not offensive.

When looking at the pairs of data broken down by image, as shown in Table 3, all images within this study had statistically significant correlations. Interestingly, for one image, Image 5, the correlation is positive, when for all other images it is negative. The result has been reviewed and no evidence is present to suggest this anomaly has been introduced by the researcher during the analysis process.

3.5.1.4 General Discussion of Study One

The following section considers whether respondents were basing their answers for the first five questions on the same attributes. This question arose from a review of the pages appearing in the highest- or lowest-rated five. The level of variation of pages appearing within the highest or lowest five, once the pages have been ranked, is small.

Each set of answers was defined as one respondent’s answers to the five questions relating to one image. This meant there were 2064 sets of data in total. The minimum and maximum values were gathered, and the range of these values was calculated by subtracting the minimum from the maximum.

Range	Quantity
0	44
1-10	333
11-20	564
21-30	433
31-40	291
41-50	170
51-60	118
61-70	55
71-80	35
81-90	14
91-99	3
100	4

Table 7 Range of values given for questions a-f per participant per image

As shown in Table 7, the number of responses with no variation was 44. These sets of data were reviewed to see if the response pattern was consistent for any individual participants. The largest number of no variation responses from any one participant was 5. Due to the numbers shown, it is clear that users were not rating all five questions with the same value. In order to take into account variation due to the measurement tool, the percentage with a range of 5 or less was also calculated and this number was 9%.

Another check was performed regarding the variation of response over all six questions, in case any participant entered consistent ratings for all six questions. There were six responses with no variation, and 2% of the sets of data had a range of 0-5. This number is not high enough to infer there was a problem with the measurement tool, or an issue with participants just clicking on the same or similar value to get to the end of the data collection quicker.

It is also conceivable that if the participants had been asked to provide a score, with no guidance as to what they should base the score on, similar pages would have scored well and badly. The possibility that the various aspects considered in this research described in this chapter (aesthetic appeal and "fit for purpose") are all judged using the same criteria is something which warrants further investigation.

In the study described in this chapter there was no opportunity to ask the participants what they were basing their answers on, and whether different criteria were being considered when answering each of the six questions. This is something worthy of investigation in the future, as the results from this study do not investigate the cause of any similarity or difference in ratings for each question.

3.5.1.5 Informal evaluation of the data collection method and tool

Having gathered some informal feedback from people that participated in the Study One there are some aspects to consider if conducting a similar survey in the future. More information concerning the task, and exactly what it entails, should be given in the invitation to take part. Some participants had issues remaining focussed on the task until the end; this could be countered by using fewer images in each show. Some people reported that they had felt they did not have long enough to make up their mind about an image. The timing was deliberate and an integral part of the study. Phrasing could be incorporated into the instructions to let the participants know the research is concerned with their initial reaction. Some participants reported having used the back button to get another look at the image. A valid option is to disable the browser back button being used with a script on the web page. The study had a low completion rate, despite large numbers of participants starting the task. It is not clear why this was. Some participants encountered issues when progressing to the next image, and had to end the study as there was no way to recover from the issue. One participant reported confusion about the meaning of the Question F (How much does this page discourage you to explore the

rest of the site?), they were unsure exactly what they were looking for due to the use of negatives and the absence of anchor points to aid their understanding.

3.6 Study 2

3.6.1 Method

The design of Study Two was the same as Study One, with some key differences: the set of web pages used, the language in the pages, and the viewing time. The study specific information, design differences and the rationale behind the decisions is covered in the relevant sections below.

3.6.1.1 Participants

61 sets of data were collected. After reviewing the information obtained, removing participants that could understand Finnish (an exclusion criterion), and removing incomplete sets, 43 complete sets of data were analysed.

The gender profile of the participants was 20 male, 22 female, and 1 other / prefer not to say. This is in line with the gender profile of the UK population (Office for National Statistics, 2014). 18 of the participants had no web design experience, 25 had designed a few websites, and no participants had designed lots of websites. The level of hot air balloon familiarity across all participants was low with no participants being either moderately or extremely familiar. Five participants were somewhat familiar with hot air balloons, 20 were slightly familiar and 18 were not at all familiar.

3.6.1.2 Materials

When selecting the images for Study Two, it was important that the purpose of the web pages was clear. Participants from Study One had informally said they struggled to answer some of the questions due to the ambiguous purpose of the pages being reviewed. A decision was made to

use an e-commerce environment of some description, as this is a familiar web environment. A site selling an activity or experience was identified as a suitable candidate.

When considering which activities would be appropriate for the study, factors such as gender participation, how familiar the activity is, equipment required, location required and status of specialist equipment were taken into consideration. An example of this would be a track day in a particular type of car, which may lead to participants rating the car not the web page. The topic of holidays would require consistency in the type of holiday (beach / cultural / activity) to minimise the effect of personal holiday preferences. The aim was to minimise the effect of these external factors. A list was created through a brain-storming session, which consisted of the following activities: skiing holidays, mountain biking experiences (holidays or day activities), hot air balloon rides and fishing holidays.

A search for suitable web pages was performed for all of the activities on the shortlist. Using the search term "Skiing holidays" resulted in a large number of sites advertising the accommodation for skiing holidays. Some of the pages had no images, others included solely images of the accommodation, and a small number included images of either people skiing or the mountains you could go skiing on. Images of accommodation were chalets in a number of cases, and this may have led to similar status related responses due to the appearance of the chalets, not the web pages so this domain was rejected. Using the terms mountain biking or fishing did not produce enough results to be suitable. The search for hot air balloon rides provided enough results to appear worth pursuing and had the perceived advantage that you did not have to be an enthusiast to enjoy the activity. Therefore, the domain of hot air balloon rides was selected.

The decision on number of pages to include in Study Two had two constraints. First, the number needed to be large enough for the correlational study described in Chapter 4. For correlation tests it is preferable to have a set of 25 or more pairs of results. This was a limitation of Study One, where only 24 pairs were collected. Second, the number of pages was chosen before a final

decision had been made about whether the online data collection would use live pages or images of the pages. The number had been chosen to allow for removal of a small number of pages, if needed, where the content had been modified dramatically during the data collection window. The decision to use images instead of live sites was made after the card sorting activity, discussed in Chapter 5, had already been completed. Therefore, in order to allow comparisons between the two sets of findings to be drawn, the same set of pages needed to be used. Study Two did not manipulate the viewing time, as this was addressed by Study One. The length of viewing time used was a maximum of 30 seconds, to enable the study to be completed within a reasonable amount of time.

The following section explores the choice of language, and the process used to translate the set of web pages.

In order to minimise the effect of text content on user opinions of web pages Study Two made use of a language that could not be understood by the participants at a glance. The language chosen needed to be one not derived from any of the more commonly spoken languages, as this would potentially exclude a high number of participants. A language using the western character set was preferred, as the use of one in a different character set was considered to introduce too high a novelty factor. It was unrealistic to attempt to locate 30 hot air balloon experience websites in one language, however, the use of multiple foreign languages was not desirable as it may have caused confusion, and would exclude more people than using a single language. Google Translate provides the ability to translate any web page into your chosen language. Therefore, sites from multiple countries were used and translated into a consistent language.

There were some limitations to the approach used for translation; some sites include words in an image, and cannot be translated. The effect of this limitation was minimised as the pages were not originally in English, so this did not lead to a readable word in most cases. A native speaker was not used to verify translation accuracy because participants were chosen who would not be

able to read the text. The participants were not being asked to understand any of the text, or prevented from knowing the purpose of the site, meaning that loanwords (Haugen, 1950) or un-translated words have no affect on the results. Before use, the pages were translated into English to verify they were for companies selling hot air balloon experiences to the public.

The following section explores the process used to gather the set of 30 pages.

In order to replicate a user's search procedure as closely as possible, pages were selected using a consistent search engine (Google), and were not restricted to homepages for the companies. The decision to replicate the user experience as closely as possible was based on the assumption that most users locate new service providers for activities of this type on the internet using a search engine (Kim & Fesenmaier, 2008). Therefore, the page a search engine links to is the first impression they have of a particular website. Many users will not explore a site past this initial page, before making decisions on their opinion of a website (Bilal, 2000). The search term "Hot Air Balloon" was translated into the following languages: Finnish, German, Italian, Romanian, Swedish and Catalan. The search was conducted using the native Google site for the language and then translated into Finnish using Google Translate. The Google search results were previewed in English and filtered to ensure the content was selling hot air balloon trips, and were not fan sites, blogs, newspaper articles or other search results irrelevant to the study.

The following section describes the process of acquiring the screenshots.

The images included the border of the browser window to indicate to the participants that the pages were static, and no interaction was possible. The use of an image increased the likelihood that the pages were viewed at a consistent screen size and had not been modified due to the participant's individual browser settings. The screenshots used were taken under consistent settings using the most popular browser at the time, Chrome (StatCounter, 2013). The window size was set to A4 size, as this was the size of the cards used for the study described in Chapter 5,

and consistent images were used for both studies. The use of images instead of live sites also allowed review for any situations where the translation method had not altered content, such as images displaying words. Pages were reviewed to ensure the majority of the page had been translated. There were a small number of aspects that were not translated and the decision was taken not to identify every instance and manually modify it as to accurately identify all such instances and modify them convincingly would have taken a disproportionate amount of time.

For the data collection described in Section 3.6, a background questionnaire was used, which can be found in Appendix Q. It consisted of four questions. One question was used to include or exclude respondents from the study and three questions were used to allow for grouping of results by demographic in later sections of this chapter.

A screening question relating to the level of understanding of the Finnish language was included in order to ensure no participants could understand the text content of the cards at a glance. The responses to this question were used to exclude participants if they could understand Finnish at a glance.

The demographic question relating to gender had three answers: male, female and other / prefer not to say. The inclusion of a so called “bucket” category covering an innocuous category and a potentially more revealing, personal category follows best practice. Several countries, including Australia and New Zealand, have introduced a Gender X option on passports, which also suggests a move towards providing categories in addition to the traditional two gender categories. A previous diversion from the traditional two categories is the use of trans-gender.

The level of detail provided by experts and novices in a particular topic can be vastly different and for an easily identifiable reason. For example, experts in archery would be able to identify the probable cost of equipment being used in a competition and infer how good the archer is by merely looking at the fletches on the arrows used. In order to control for differences between

experts and novices a background question was included to address the level of familiarity the participant had with the topic. The following phrasing used was:

> How familiar do you consider yourself to be with hot air balloons?
>
> Not at all familiar, Slightly familiar, Somewhat familiar, Moderately familiar, or Extremely familiar.

This allowed the participant to choose the point of view and to self-identify with a particular statement. The phrasing prevented the researcher's perception of an expert or novice level of knowledge of the domain affecting the responses. The phrasing does not distinguish between hot air balloons and hot air ballooning, and this ambiguity may have affected the responses given. The question was not used to exclude participants, but allowed for analysis to be undertaken on differences between these different groups of respondents.

Another area where expert or novice knowledge might have caused differences between respondents was in the area of web design. There are well known guidelines for web design, and it was possible that different levels of knowledge of these would influence the perception of each of the web pages. For this reason a question was included to find out how much experience of website design the participant had. The question used was:

> What is your level of web design experience?
>
> Never designed a website, Designed a few websites (1-10), or Designed many websites.

This represents a fairly small number of options as a higher level of granularity between the actual number of sites designed, as used in Study One, was not considered necessary. The question was not used to exclude participants, but allowed for analysis to be undertaken on differences between these different groups of respondents.

A specifically-developed online tool was used for the data collection. Adam Stanton, a colleague with professional expertise in this area, created the tool, based on the researcher's specification. It was created using HTML and PHP with a MySQL database to store the results. The tool used MD5 encryption to create a unique identifier from participant's email addresses. The email address entered was not stored anywhere within the tool, making the stored information anonymous. The tool allowed for the 30 cards to be displayed in turn to participants. The participant's responses to the six questions, found in Section 3.3.3, were gathered and recorded. The responses were downloaded from the server in CSV format, allowing them to subsequently be analysed using Microsoft Excel. A web page was created to display each screenshot at a consistent size and the screen shot was centred in the web page, with white space around the top, sides and bottom. This was to imply to the participant it was a screenshot and not an interactive page which could be navigated.

3.6.1.3 Procedure

Participants first completed a background questionnaire with four questions, which have been covered in detail in Section 3.6.1.2. An instruction sheet was shown which explained the task. A consent form was then displayed, as required by the University Ethics Panel. A copy of these materials can be found in Appendices G to I respectively, along with a copy of the ethical approval letter. In a departure from the procedure for the study described in Section 3.3.4, a masking image was not displayed due to time constraints, and practical constraints identified within the masking image creation process. Participants were then directed to a web page which displayed on the screen for 30 seconds. After five seconds the first question appeared with the web page still visible below. Once an answer had been provided for the first question the next question immediately appeared, and this continued until all six questions had been answered. This procedure is different from the one used in Study One, where all six questions were visible at the same time. Once all six questions had been answered the data collection tool displayed the next web page. It is possible that the web page may no longer have been displayed during the

answering of some questions, if the answers had not been provided within the 30 second viewing time.

3.6.2 Results

3.6.2.1 Subjective ratings for cards

Table 8 shows the mean rating for each card and question, ordered by CardID, the standard deviation can be seen in Table 9 below. The mean was calculated using the results from 44 participants who completed the study.

Card ID	Mean QA	Mean QB	Mean QC	Mean QD	Mean QE	Mean QF
1	53.81	48.86	44.02	42.21	41.84	31.86
2	29.40	31.44	32.28	39.40	32.84	40.16
3	64.56	52.84	48.02	48.12	54.67	34.88
4	41.88	40.00	39.28	44.44	45.79	32.47
5	48.79	44.09	38.33	36.98	40.93	43.26
6	39.67	38.40	45.51	48.21	46.93	33.35
7	35.23	35.91	35.49	37.84	34.47	35.56
8	26.28	29.60	34.65	41.47	33.56	38.60
9	36.56	35.56	42.53	44.12	41.91	26.63
10	50.79	46.70	45.44	49.37	49.53	28.81
11	17.23	16.09	19.72	19.12	20.77	52.60
12	38.23	39.93	41.33	43.74	41.67	34.86
13	39.79	39.40	39.74	35.88	38.95	34.98
14	69.88	67.23	61.14	62.26	63.88	16.33
15	35.72	36.81	33.72	34.07	39.84	44.42
16	57.70	54.86	54.84	54.53	57.05	22.49
17	62.37	48.09	49.00	49.84	56.09	27.84
18	40.09	40.35	39.33	46.49	48.23	32.95
19	45.49	43.49	40.44	41.67	39.12	32.77
20	56.56	54.12	52.77	55.00	56.79	25.09
21	48.77	30.70	27.93	25.05	32.26	48.77
22	43.12	46.00	39.95	41.26	43.56	33.84
23	48.26	44.00	45.12	45.12	48.51	28.51
24	36.44	33.14	29.93	27.00	27.79	41.53
25	46.93	45.16	36.81	33.84	39.14	39.74
26	22.88	19.63	20.84	27.60	27.65	53.26
27	57.42	57.21	51.37	53.23	56.70	24.02
28	58.98	59.30	51.72	52.98	53.37	29.37
29	47.26	44.44	39.26	36.77	39.37	31.51
30	59.53	48.14	48.93	48.72	47.37	32.74

Table 8 Mean subjective ratings for hot air balloon cards

Table 9 shows the mean rating and standard deviation for each card and question, ordered by CardID. The mean was calculated using the results from 44 participants who completed the study.

	Question A		Question B		Question C		Question D		Question E		Question F	
Card ID	Mean	St Dev	Mean	St Dev	Mean	St Dev	Mean	St Dev	Mean	St Dev	Mean	St Dev
1	53.81	33.10	48.86	27.68	44.02	24.55	42.21	26.86	41.84	27.69	31.86	26.35
2	29.40	23.54	31.44	24.43	32.28	24.05	39.40	24.93	32.84	23.23	40.16	32.55
3	64.56	24.90	52.84	28.08	48.02	25.21	48.12	25.38	54.67	22.31	34.88	27.46
4	41.88	21.76	40.00	24.40	39.28	22.57	44.44	22.13	45.79	24.38	32.47	27.47
5	48.79	28.48	44.09	29.03	38.33	25.53	36.98	25.44	40.93	26.79	43.26	28.93
6	39.67	24.01	38.40	23.37	45.51	22.98	48.21	23.84	46.93	24.40	33.35	28.16
7	35.23	23.06	35.91	20.77	35.49	20.52	37.84	21.53	34.47	19.64	35.56	26.74
8	26.28	23.57	29.60	26.09	34.65	28.58	41.47	28.18	33.56	24.47	38.60	30.23
9	36.56	25.29	35.56	25.36	42.53	28.73	44.12	27.43	41.91	24.49	26.63	24.02
10	50.79	25.50	46.70	25.21	45.44	24.48	49.37	24.34	49.53	23.47	28.81	23.44
11	17.23	17.25	16.09	17.71	19.72	20.45	19.12	19.09	20.77	21.98	52.60	35.40
12	38.23	27.93	39.93	30.99	41.33	27.14	43.74	31.08	41.67	28.89	34.86	28.98
13	39.79	20.11	39.40	21.34	39.74	19.65	35.88	21.34	38.95	21.93	34.98	26.25
14	69.88	22.48	67.23	19.86	61.14	21.09	62.26	19.03	63.88	22.87	16.33	17.35
15	35.72	19.88	36.81	20.66	33.72	19.28	34.07	19.91	39.84	19.56	44.42	28.28
16	57.70	24.85	54.86	27.30	54.84	25.22	54.53	28.14	57.05	27.73	22.49	22.58
17	62.37	23.12	48.09	27.36	49.00	26.60	49.84	22.80	56.09	25.00	27.84	27.14
18	40.09	26.05	40.35	25.87	39.33	23.26	46.49	25.50	48.23	26.12	32.95	24.89
19	45.49	28.19	43.49	29.49	40.44	25.86	41.67	29.81	39.12	27.05	32.77	27.44
20	56.56	24.89	54.12	23.28	52.77	24.86	55.00	22.28	56.79	22.58	25.09	22.13
21	48.77	32.67	30.70	27.75	27.93	25.21	25.05	22.02	32.26	29.41	48.77	35.44
22	43.12	26.67	46.00	26.50	39.95	23.85	41.26	25.81	43.56	25.33	33.84	28.28
23	48.26	25.64	44.00	23.79	45.12	21.20	45.12	23.33	48.51	25.04	28.51	20.29
24	36.44	26.29	33.14	26.35	29.93	24.40	27.00	24.82	27.79	22.79	41.53	33.72
25	46.93	30.25	45.16	33.19	36.81	28.48	33.84	25.70	39.14	29.40	39.74	31.78
26	22.88	22.86	19.63	20.66	20.84	16.24	27.60	25.01	27.65	23.27	53.26	31.83
27	57.42	24.56	57.21	24.57	51.37	24.56	53.23	26.74	56.70	24.30	24.02	23.39
28	58.98	32.48	59.30	30.43	51.72	31.85	52.98	31.40	53.37	29.11	29.37	27.46
29	47.26	26.81	44.44	24.77	39.26	22.25	36.77	20.87	39.37	21.69	31.51	25.17
30	59.53	26.72	48.14	28.50	48.93	26.88	48.72	29.96	47.37	27.82	32.74	28.40

Table 9 Mean and Standard deviation for subjective ratings of hot air balloon cards

One observation when reviewing the information in Table 9 above, is that it is clear the largest level of variance in the results for Question F, which shows the level of agreement between

respondents is smallest for this question. The level of agreement was not one of the stated aims, however, it is worth reviewing the information for this as it shows how well the mean reflects the full set of responses given for each card.

3.6.2.2 Correlation test for encourage and discourage

The following table presents the correlation coefficients found between the answers to Question C (How much does this page encourage you to explore the rest of the site?) and Question F (How much does this page discourage you to explore the rest of the site?), for each image.

Card ID	Correlation Coefficient
1	-0.30
2	-0.51
3	-0.60
4	-0.39
5	-0.51
6	-0.60
7	-0.50
8	-0.48
9	-0.27
10	-0.38
11	-0.42
12	-0.71
13	-0.39
14	-0.47
15	-0.52
16	-0.20
17	-0.62
18	-0.57
19	-0.45
20	-0.54
21	-0.66
22	-0.29
23	-0.48
24	-0.54
25	-0.72
26	-0.53
27	-0.46
28	-0.55
29	-0.26
30	-0.75
Overall	-0.53

Table 10 Pearson's correlation coefficient for each card

The following scatter graph plots the values for encourage and discourage across all cards and all respondents.

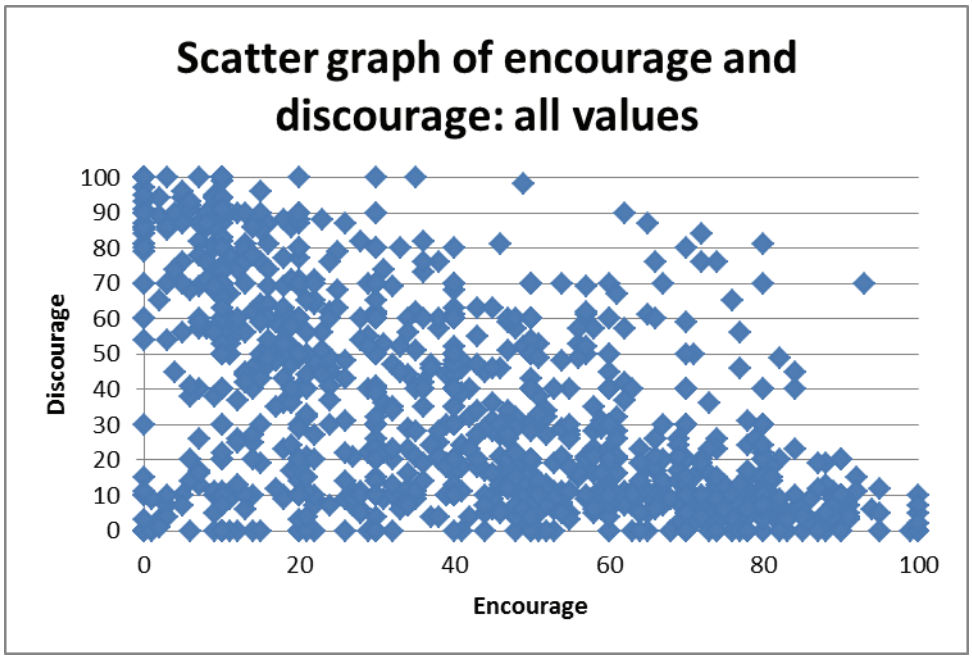


Figure 14 Scatter graph to show encourage and discourage: all values

3.6.3 Discussion

3.6.3.1 Subjective ratings by rank

	Question A		Question B		Question C		Question D		Question E		Question F	
Rank	Card	Mean	Card	Mean	Card	Mean	Card	Mean	Card	Mean	Card	Mean
1	14	69.88	14	67.23	14	61.14	14	62.26	14	63.88	26	53.26
2	3	64.56	28	59.30	16	54.84	20	55.00	16	57.05	11	52.60
3	17	62.37	27	57.21	20	52.77	16	54.53	20	56.79	21	48.77
4	30	59.53	16	54.86	28	51.72	27	53.23	27	56.70	15	44.42
5	28	58.98	20	54.12	27	51.37	28	52.98	17	56.09	5	43.26
6	16	57.70	3	52.84	17	49.00	17	49.84	3	54.67	24	41.53
7	27	57.42	1	48.86	30	48.93	10	49.37	28	53.37	2	40.16
8	20	56.56	30	48.14	3	48.02	30	48.72	10	49.53	25	39.74
9	1	53.81	17	48.09	6	45.51	6	48.21	23	48.51	8	38.60
10	10	50.79	10	46.70	10	45.44	3	48.12	18	48.23	7	35.56
11	5	48.79	22	46.00	23	45.12	18	46.49	30	47.37	13	34.98
12	21	48.77	25	45.16	1	44.02	23	45.12	6	46.93	3	34.88
13	23	48.26	29	44.44	9	42.53	4	44.44	4	45.79	12	34.86
14	29	47.26	5	44.09	12	41.33	9	44.12	22	43.56	22	33.84
15	25	46.93	23	44.00	19	40.44	12	43.74	9	41.91	6	33.35
16	19	45.49	19	43.49	22	39.95	1	42.21	1	41.84	18	32.95
17	22	43.12	18	40.35	13	39.74	19	41.67	12	41.67	19	32.77
18	4	41.88	4	40.00	18	39.33	8	41.47	5	40.93	30	32.74
19	18	40.09	12	39.93	4	39.28	22	41.26	15	39.84	4	32.47
20	13	39.79	13	39.40	29	39.26	2	39.40	29	39.37	1	31.86
21	6	39.67	6	38.40	5	38.33	7	37.84	25	39.14	29	31.51
22	12	38.23	15	36.81	25	36.81	5	36.98	19	39.12	28	29.37
23	9	36.56	7	35.91	7	35.49	29	36.77	13	38.95	10	28.81
24	24	36.44	9	35.56	8	34.65	13	35.88	7	34.47	23	28.51
25	15	35.72	24	33.14	15	33.72	15	34.07	8	33.56	17	27.84
26	7	35.23	2	31.44	2	32.28	25	33.84	2	32.84	9	26.63
27	2	29.40	21	30.70	24	29.93	26	27.60	21	32.26	20	25.09
28	8	26.28	8	29.60	21	27.93	24	27.00	24	27.79	27	24.02
29	26	22.88	26	19.63	26	20.84	21	25.05	26	27.65	16	22.49
30	11	17.23	11	16.09	11	19.72	11	19.12	11	20.77	14	16.33

Table 11 Mean subjective ratings for hot air balloon web pages in rank order

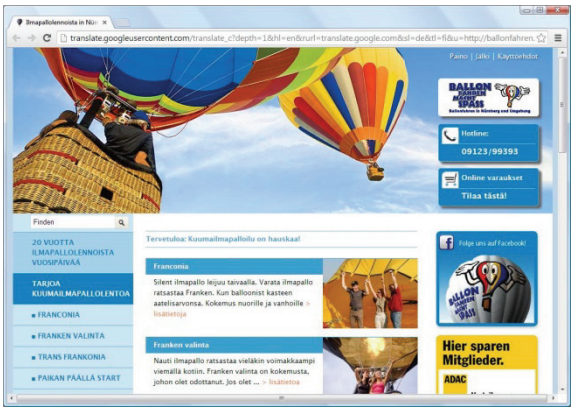
Table 11 shows the cards in rank order, ranked by mean rating for each question. The mean was calculated using the results from 44 participants who completed the study.

	Highest rated six cards	Lowest rated six cards
How eye-catching do you find this page?	3, 14, 17, 27, 28, 30	2, 7, 8, 11, 15, 26
How visually attractive do you find this page?	3, 14, 16, 20, 27, 28	2, 8, 11, 21, 24, 26
How much does this page encourage you to explore the rest of the site?	14, 16, 17, 10, 27, 28	2, 8, 11, 21, 24, 26
How well designed is this page?	14, 16, 17, 20, 27, 28	11, 15, 21, 24, 25, 26
How good is this page as an advertisement for the site?	3, 14, 16, 17, 20, 27	2, 8, 11, 21, 24, 26
How much does this page discourage you to explore the rest of the site?	5, 11, 15, 21, 24, 26	9, 14, 16, 17, 20, 27

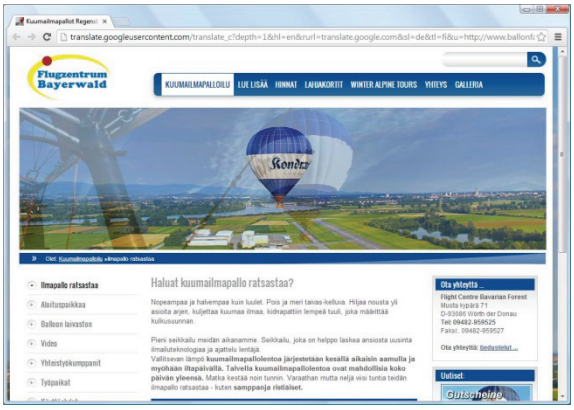
Table 12 Highest and lowest rated web pages when in rank order based on mean rating

In the following section the cards that have appeared in the highest or lowest rated six cards have been reviewed for common visual features, in order to address the main question posed in this chapter: What visual aspects of a web page affect user ratings?

Card 3



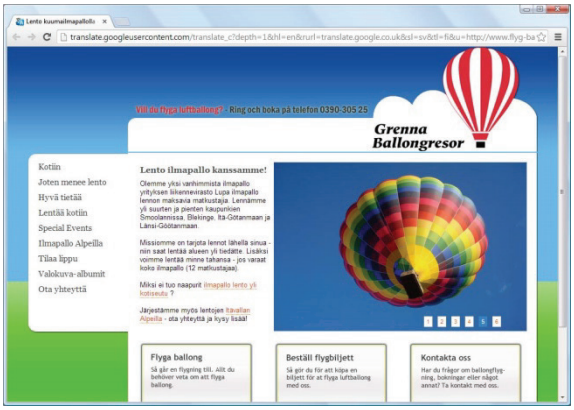
Card 14



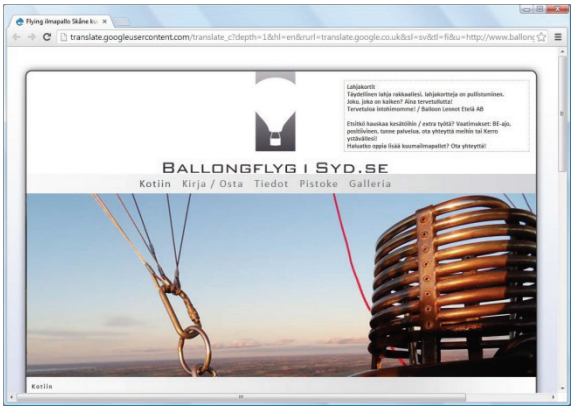
Card 16

Card 17

Card 20



Card 27



Card 28

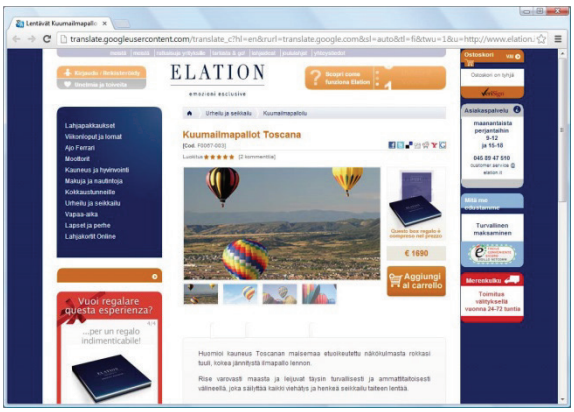
Card 30

Figure 15 Cards rated in the highest six for questions A-E

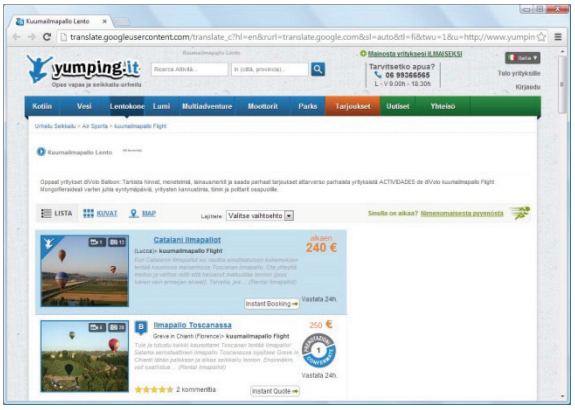
By reviewing the cards in rank order it can be seen that some cards were favoured over others. Card 14 was rated highest for every question and Card 27 appeared within the top six for every question. Both cards use bright colours and have blue as the background in the top portion of the page. A number of the images appearing in the top six for any of the questions make use of blue

for their colour scheme. The prevalence of the use of blue in the top rated images contrasts with the findings in Section 3.5.1.1, where the colour green was found in those rated highly across all questions, and the colour red was commonly found in those rated highly for how eye-catching they were. It may be that there are domain specific preferences, and for the domain of hot air balloons, blue has positive links. One possible explanation for the positive association between the colour blue and hot air balloon web pages is that hot air balloons fly in the sky, which is blue. Therefore, the use of blue primes the user to think of the hot air balloon in a commonly associated location. One participant for the card sorting activity reported in Chapter 5 gave this as their reasoning for placing cards with blue backgrounds in a preferred category. This concept is discussed in more detail in Section 5.6.8. Cards 16, 17 and 20 each appeared in the top six a total of five times, and all contained blue, to varying degrees. When looking at the standard deviation of responses for the two highest rated cards, Card 14 and Card 27, which can be found in Table 9, it can be seen that Card 14 has a small standard deviation, relative to the other standard deviations calculated, whereas Card 27 has a standard deviation which closer follows the mean of all the deviations calculated. A small standard deviation corresponds to a high level of agreement between participants. All the cards rated in the highest six cards for one or more questions, shown in Figure 15, include a large picture. Three of the cards include images of one of more balloon, with a design incorporating a rainbow of bright colours.

Card 2

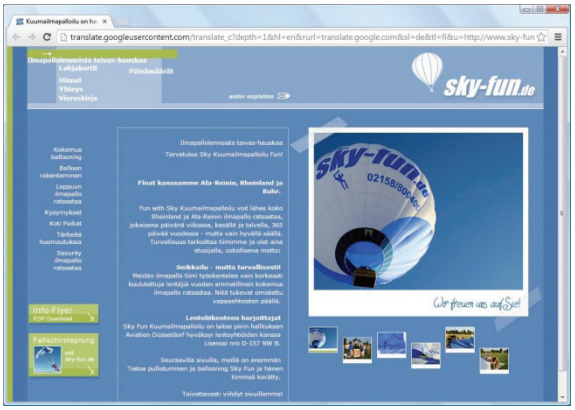


Card 7

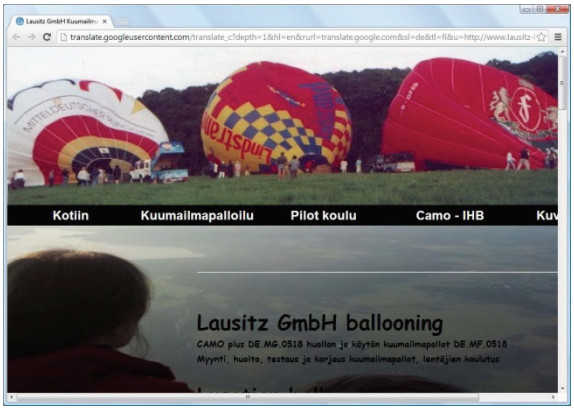


Card 8

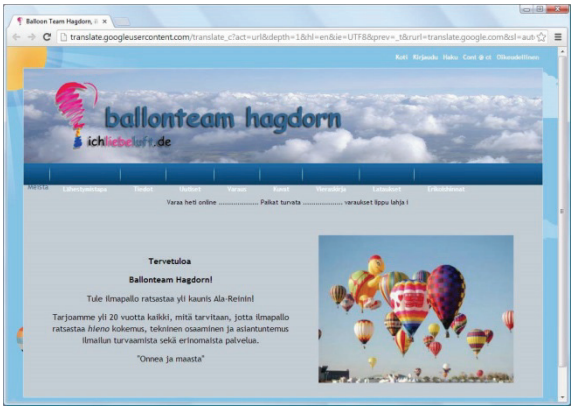
Card 11



Card 15

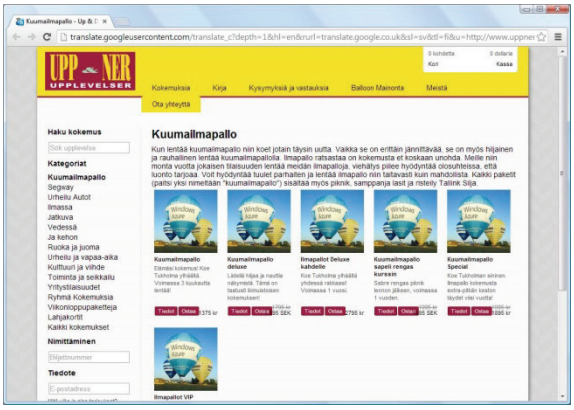


Card 21



Card 24

Card 25



Card 26

Figure 16 Cards rated in the lowest six cards for questions A-E

Card 11 and card 26 both appeared in the lowest six cards for every question with Card 11 being rated lowest five times. A number of the cards in the lowest rated six cards for any of the questions used a blue colour scheme, with varying quantities of blue present in the page. In three of the images blue was the predominant colour. In three other cards blue was the background colour used for the page, although the text had a white background. The use of a blue colour scheme was also identified in a number of the cards in the highest rated six cards for one or more question, meaning the use of a blue colour scheme alone does not predict placement in the highest or lowest rated six cards. When considering the use of images within the cards, the majority included in lowest six cards for one or more question, shown in Figure 16, have a number of small images. A trend identified in the cards included in the top six for one or more question, shown in Figure 15 was the inclusion of at least one large image. It is conceivable that for a website of this type, that being one trying to sell an experience, users would like to see an image that gives them idea what the activity will be like, and this is potentially communicated more successfully through large images, than smaller images.

When looking at the standard deviation of responses, shown in Table 9, for the two lowest rated cards, Card 11 and Card 26, it can be seen that Card 11 has a small standard deviation, relative to the other standard deviations calculated, for four questions, and Card 26 has a small deviation, relative to the other standard deviations calculated, for three questions. For Question F (How

much does the page discourage you to explore the rest of the site?), Card 11 and Card 26 both have a large standard deviation, relative to the other standard deviations calculated.

An informal analysis of the mean value results broken down by gender, shown in Appendix M, found females ranked pages with higher values for four of the questions (How eye-catching do you find this page?; How visually attractive do you find this page?; How much does this page encourage you to explore the rest of the site?; How much does this page discourage you to explore the rest of the site?). The standard deviation of the mean ratings from females was higher than males for all questions except Question D (How well designed is this page?). This shows the level of variation within responses was higher for females than for males. As there was only one participant in the gender category "Other / prefer not to say", this category was not reviewed for comparison.

Having reviewed the correlation calculations, shown in Table 10, and the scatter graphs, shown in Appendix N, the results are consistent with those found in Study One. The main purpose of including this for both studies was to investigate for consistency. A further systematic analysis of the level of difference between the two sets of information would be worthy of further investigation.

3.6.3.2 General discussion of Study Two

One observation when reviewing the information in Table 9, that shows the mean and standard deviation for subjective ratings of hot air balloon web pages, is that the largest level of variance in the results is for Question F (How much does this page discourage you to explore the rest of the site?), which shows the level of agreement between respondents is smallest for this question. This may be due to the unusual nature of the question, or alternatively, respondents trying to work out how they answered Question C (How much does this page encourage you to explore the rest of the site?), in order to provide the inverse. Anecdotally, a small number of participants contacted the researcher after they had taken part to apologise that they did not always answer

Question F accurately. They stated that they sometimes got confused, as Question F was the inverse of the other five questions.

3.7 General discussion

3.7.1 Similarity of visual attributes

The high- and low-rated paged were reviewed to see if there were any trends in the used of colour, and no trends were found. There was, however, commonality in the intensity of colours within high-rated pages. In Study One, high-rated pages for Question A (How eye-catching do you find this page?) contained deep red, and high-rated pages from Study Two used bright colours. There were also inconsistencies regarding the use of background images, and the colour of the background. Due to the differing domains and, in particular, purpose of the web pages used for each study any generalisations of trends, or absences thereof, are tentative.

3.7.2 Similarity of encourage and discourage correlations

The findings from both Study One and Study Two showed negative correlations between Question C (How much does this page encourage you to explore the rest of the site?) and Question F (How much does this page discourage you to explore the rest of the site?). In each case the correlations were not directly proportional. The trend identified in Study One regarding the clustering of results in the bottom left hand triangle and the absence of results in the top right hand triangle was also present in Study Two, as shown in Appendix K and Appendix N respectively.

3.8 Conclusions

This chapter has described two related studies that gathered information about subjective opinions of web pages. The results were analysed in order to identify visual trends within high- and low-rated pages.

The key conclusions from the research described in this chapter were:

- Encouragement and discouragement of further exploration of a web site are two distinct concepts
- User reactions to colour are based on the intensity of colours, not just the hue
- Immediate reactions are frequently more positive than considered reactions
- A large, relevant image of an activity creates a positive reaction (domain specific finding from Study 2)

The implications of each of these are discussed in turn below.

When reviewing the results pertaining to encouragement and discouragement, and whether they are on the same scale, it can be seen that although a negative correlation exists they do not appear perfectly correlated. This finding means that they should be considered as two separate concepts, which replicates the findings of Rugg (2013). The implication of this finding is that evaluations of websites should address encouragement and discouragement separately. The terms should not be used as opposing anchor points for Likert or Likert-style scales. When designers are trying to optimise their sites they should consciously consider whether they are trying to increase encouragement levels, or minimise discouragement levels. To use more conventional web development terminology, this is whether they are trying to maximise positive impressions, or minimise negative impressions. For pages with high levels of discouragement designers have not addressed the hygiene factors.

The research presented in this chapter identified trends regarding the impact of the intensity of the colours within high-rated pages. As a general finding, bright, high intensity colours gave positive responses, however, for some domains high intensity colours may not be appropriate. Most current guidance concentrates on choice of hue, but this research shows that colour intensity is also a significant factor.

When considering the difference between initial reactions to viewing a web page (first impressions) and those formed after further viewing (considered impressions) the research should that the immediate reactions tended to be more positive. Confirmation bias (Koriat *et al.*, 1980) would imply that web designers would be better served to strengthen immediate impressions in order to achieve the best overall impression.

Within the Study 2 domain the results showed that a large, dominant image of the domain subject created a positive reaction. This is potentially an important factor in the design of activity or experience orientated web pages, but further investigation would be required to confirm this.

3.8.1 Limitations and further work

Two studies following similar experimental design have been conducted for the research described in this chapter. The following sections will reflect firstly on limitations with the overall experimental design, and then each study in turn.

The research presented in this chapter has an implied assumption that subjective opinions of web pages are nomothetic. The sampling methods used for both studies were convenience and snowball sampling. The absence of a defined population means the results from this study cannot be generalised to the full population. Due to the nature of the data collection the participants all had access to computers, been of an age to understand the survey to be able to take part, and must have had a working knowledge of English. This means there are limitations to the generalisation of the findings. This piece of research was intended as a proof of concept, and has produced some very interesting findings. These findings suggest it is a suitable technique to use for gathering subjective opinions of web pages. Future studies with defined populations would enable a further question to be addressed; are subjective opinions of web pages an example of nomothetic or idiographic preferences? This question was outside of the scope of the research described in this chapter.

The research shows that the intensity of colours used is a factor worthy of further exploration and this could be achieved through further studies, which include an objective measurement of the colour intensity values.

The research undertaken in Study One used a sample of web pages from *CSS Zen Garden* (Shea, 2005). Each of the pages had been selected by the editors to be included in the book, from a larger set of websites. The standard of web design within the sample is, therefore, likely to include less variation than a more realistic sample of pages within a particular domain. Study Two addressed this issue by targeting a specific domain and collecting the sample from the internet through a search engine, a process which closely emulates user web page identification practices. This limitation has implications for the results of Study One and also on the results of any comparison of the two sets of results.

Due to the anonymous nature of the data collection for Study Two, and the timing of the card sorting activity described in Chapter 5 in relation to the online data collection for Study Two, it is not possible to know whether any individuals participated in both studies. It is conceivable that any individuals participating in both studies would have a more favourable response to the cards due to the Mere Exposure Effect (Zajonc, 2001). That is, the existing familiarity with the cards led to an increased positive response towards the cards, compared to if they had been an unfamiliar set of cards. This poses a threat to the validity of the findings from Study Two, however, the impact of this effect would be consistent across the full set of cards used, therefore limiting any impact on the results. In cases where participants asked the researcher if they were able to take part in the study described in this chapter, they were discouraged from doing so. For logistical reasons it was not possible to add a screening question, or an exclusion criteria, to account for this situation once this potential overlap of participants became apparent. The number of participants in the study described in Chapter 5 was 18, and the number of participants in Study

Two was 43. The benefit of using the same set of cards for two studies was great enough to justify this decision.

In Study One the viewing time was manipulated in order to address question about the impact this would have. This variable was not manipulated in Study Two. The findings from a comparison of the two sets of information is worthwhile, and produced some interesting results, however as the time manipulation was not consistent no tangible comparisons can be inferred.

This chapter was concerned with gathering information about subjective opinions of web pages, and using it to identify common visual features within high- and low-rated pages.

Chapter 4 Correlations between subjective ratings and objective measures of Web pages

4.1 Introduction

The purpose of the research described in this chapter was to investigate whether there are objective measurements that could be used to predict subjective opinions of web pages.

In order to address this question two independent studies were conducted. For each study a set of objective measures was determined, through both a review of relevant literature and using the findings from Chapter 3. Subsequently, a test was devised to assess whether any of the measures could be used as predictors of subjective opinions.

The following section provides an overview of some additional areas of interest: information complexity and text complexity. Based on this literature two sets of objective measures, one for each study, are proposed. It also explores potential options for identifying objective measures to be used as proxies for subjective opinions, each of which is presented below.

4.2 Background

The level of visual information present within a page, and the effect this has on individuals, has not yet been covered. The following section presents an overview of the research into complexity of documents and how this applies to web design.

4.2.1 Complexity

Images that are stored on a computer can be stored in various different file formats. The format that you chose will depend on the purpose of the image and how much space is available to store it. Compression techniques have been created in order to reduce the space a file will take up. In order to store an image, rather than storing every pixel's colour value, the file stores patterns of information, such as ten white pixels in a row, instead of ten individual pixels all being white. For

any given compression algorithm, the more complex an image the less the file can be compressed. Therefore, by seeing how much a file has been compressed, one can gain a measure of the complexity of a page. An application of Shannon's information theory (1949), which was concerned with the level of noise and complexity of sound to be submitted through a channel (Sayood, 2012), is file compression. Tuch *et al.* (2009) also proposed the size of file once compressed, as a possible proxy for visual complexity.

There has been some research into web page quality using card sorts to obtain attributes and categories (Hurd, 2001). One common category is how cluttered the page looks. This could be due to a bad layout, or by trying to put too much information on one page; this possible "cluttered" feel obviously lending itself to information theory (Shannon & Weaver, 1949). The compression ratio of the page may, therefore, be able to indicate how cluttered a page is.

Information theory also gives values to documents concerning other criteria, for example how novel they are. Research into facial beauty and fractal geometry found that individuals preferred less complex faces which required less processing, so long as certain features were in proportion with each other (Rhodes *et al.*, 1998). However previous research into web pages (Holland, 2004) found that individuals preferred the more complex web pages that took more processing, leaving it unclear whether there is one overriding preference, or whether it varies by context or individually.

Usability heuristics are rules of thumb used by designers which provide guidance on a number of different aspects of website design. Nielsen (1995) suggests web pages should have a minimalist design. One web page which takes minimalist design to the extreme is Google (Rugg, 2013). The functionality of this website is key, and the Google web site does not allow design to detract from this. A search engine has a single purpose and therefore, this minimalist approach links well to the purpose of the site. However, may sites have more than one purpose, meaning this approach is suitable for a search engine but is not suitable for many other types of site (Nielsen & Tahir,

2001). It is unclear whether there is a cut-off point before which a page is considered uncluttered, after which it is cluttered, or whether there is a grey area where a page is not considered minimalist, but is not considered cluttered either. It is also unclear whether the line for minimalist or cluttered is at a similar point for users in general, or whether it is a personal thing. Therefore, an investigation of whether the information complexity level of a website can be used to predict opinions, is a justified one.

4.3 Method

A number of potential objective measures were identified through the literature, and through the findings of the research described in Chapter 3.

4.3.1 Design

The two studies described in this chapter were correlation studies, that is, two sets of information were obtained, hypotheses regarding correlations were formed and correlations between these two sets of information were tested in order to accept or reject the proposed hypotheses.

Objective measures were gathered for two sets of web pages. The web pages were the same two sets as used in the research described in Chapter 3 and were screen shots of web pages from *CSS Zen Garden* and web pages for selling a hot air balloon experience. The rationale behind these domains has been given in Section 3.3.3 and Section 3.6.1.2. The same set of measures has been used for each study. The values were gathered either through software or by eye. The subjective ratings were those gathered in Chapter 3 through an online survey. A null and alternative hypothesis was formed for each measure, based on the literature, regarding the type of correlation anticipated. The objective values and subjective ratings were analysed for correlations using Spearman's Rank Correlation Coefficient, as the values did not follow a normal distribution.

4.3.2 Objective measures

The objective measures have been based on background literature into preferences and the findings from Chapter 3.

It should be noted that, for each study, the hypothesis formation and testing took place after the analysis of the subjective results had been undertaken. This meant that the domain specific findings revealed in Chapter 3 were taken into account.

4.3.2.1 Colour

To measure the colours used within web pages a number of values were gathered using image manipulation software: GIMP, ImageJ, and Adobe Photoshop CS3. The software was chosen due to availability and the features required. In order to determine the amount of red, green and blue within each page, RGB values (the proportions of red, green, and blue light required to generate each pixel of the image) were recorded from GIMP for Study One, and ImageJ for Study Two. The colour histogram for each web page was reviewed and the values for the mode and standard deviation were recorded for red, blue and green, the overall value mode (showing the most commonly occurring colour value within the page), and the standard deviation. The value for mode given by GIMP or ImageJ takes all colours used within the page into account. The standard deviation shows the level of contrast within the page: a page with a high level of contrast has a large standard deviation and a page with a low level of contrast has a small standard deviation. The values given for red, green and blue refer to the amount of each of these hues.

The number of different colours used within the page was also of interest and this was found by looking at the properties of the image within Adobe Photoshop CS3. The name of this property was “Number of unique colours”.

4.3.2.2 Complexity

File compression from a lossless file format into a lossy file format involves taking an existing file, and reducing the file size by storing only some of the information. This is done using compression

algorithms: the compression results in a loss of information, hence the name. This concept is discussed in more detail in Section 4.2.1 above. To put a value on how much a file could be compressed a compression ratio was calculated. In order to obtain a value for the amount of compression able to be performed on a file each screen shot was firstly saved in TIF format, a lossless file format, and the file size was recorded. Using Adobe Photoshop CS3 the same file was then saved in JPEG format, a lossy file format, at a consistent level of detail for each image, and the file size recorded.

It is worth noting that had alternative lossless or lossy file types been used the compression ratio may differ as the algorithm used may differ: for the purpose of this study a consistent lossless file format and a consistent lossy file format were used.

4.3.2.3 Text

To gather information about the number of words visible in the portion of the page used for subjective data collection the image was looked at by eye and a count of words performed manually. An automated process for this would be more appropriate if this measure were to be used for a larger sample size or in subsequent work.

4.3.3 Objective measures not tested

There were various measures that were not tested. Reading scores for the web page content were not used as the content had been translated into a language which would not be understood by the participants at a glance. As participants were aware of this, they were unable to spend any time comprehending the text.

As part of the image histograms used to gather the mode and standard deviation for each web page, the mean and median were also displayed. Links between user perception of images and the mean or median colour values were not present in the literature. Therefore, values were not obtained for the mean or median for each web page.

4.3.4 Subjective measures

The subjective measures used are the same set as in Study Two in Chapter 3. A discussion of what they were, how they were gathered, and the findings can be found in Chapter 3.

4.3.5 Correlation tests

A number of hypotheses were formed based on the findings from Chapter 3 and the discussion of the literature in Chapter 2 and Section 4.2. The hypotheses were tested for statistical significance using Spearman's Rank Correlation Coefficient, a decision which will has been further explored below. The 95% confidence interval for rho was also calculated. The correlation tests were performed using R.

- *Quantity of red will have a positive correlation to how eye catching a page is (based on findings in Chapter 3),*
- *Quantity of green will have a positive correlation for all questions (based on findings in Chapter 3),*
- *Complexity of the page will be negatively correlated to Questions A – E,*
- *The font used for body text will have no correlation with subjective user ratings (will be consistent regardless of font used),*
- *Number of unique colours will be positively correlated with Questions A – E.*

To test for correlations an appropriate statistical test was required. Pearson's correlation coefficient is used in situations where the data follows a normal distribution, while Spearman's rank correlation coefficient is used when the data does not follow a normal distribution (Rees, 2001). In order to use data that does not follow a normal distribution the data is placed in rank order, and the correlation calculated using this order, instead of the original values. In cases where this transformation results in uniformly distributed data, Spearman's rank correlation coefficient can then be used (Rees, 2001). After consultation with an independent statistician, Spearman's rank correlation coefficient was used for the analysis of the subjective and objective

measures presented in this chapter, as the results did not follow a normal distribution. The results have been presented with the rho value and the 95% confidence interval. The 95% confidence interval provides the range of values the true value is likely to lie between, and is used to infer whether any statistically significant correlations found within the sample used can be generalised to the population as a whole (Hopkins, 2000). Confidence intervals are more useful than p values alone as they show the strength of the effect in the population (Davies & Crombie, 2009). The research described in this chapter involves calculating the correlation coefficient for a large number of tests.

4.4 Results

4.4.1 Study One: CSS Zen Garden

4.4.1.1 Objective measures

Image	Comp-ression ratio (%)	GIMP value st dev	GIMP value mode	GIMP red st dev	GIMP red mode	GIMP green st dev	GIMP green mode	GIMP blue st dev	GIMP blue mode	GIMP unique colours	Words	AQUINE
1	74	78.97	0	83.43	0	79.17	0	81.31	0	152841	162	91.3
2	92	32.47	255	32.47	255	40.68	255	54.32	255	64977	413	42.8
3	93	35.70	247	35.70	247	33.01	249	63.41	251	38834	221	12.4
4	97	72.85	102	72.85	102	53.66	153	53.67	153	55627	227	42.4
5	91	109.55	0	109.55	0	106.01	0	13.95	0	47371	175	58.9
6	92	70.85	43	70.85	43	68.72	20	69.37	0	72864	256	68.3
7	94	32.17	255	32.17	255	37.16	255	41.23	255	55755	327	39.1
8	91	44.33	181	44.33	181	104.58	30	75.35	61	76796	209	25.7
9	94	58.76	255	58.76	255	56.70	255	76.84	255	54190	163	28.5
10	88	42.72	255	36.86	255	43.61	255	53.12	255	84531	369	15.9
11	91	46.07	255	44.91	255	49.94	255	49.81	255	50866	230	32.9
12	93	57.16	255	70.98	255	57.88	255	47.30	255	59892	316	16.9
13	94	44.85	255	48.20	255	36.84	255	52.39	255	76453	345	33.1
14	77	100.18	0	96.31	0	108.75	0	105.01	0	84373	78	45
15	91	47.97	181	69.08	170	50.79	187	31.32	187	43723	191	47.3
16	95	32.41	255	32.41	255	31.98	255	57.64	255	44246	333	22.2
17	91	36.59	255	36.59	255	36.52	255	39.77	255	38067	244	42
18	64	83.38	255	83.38	255	86.54	255	90.55	0	81155	275	30.2
19	89	57.33	232	57.33	232	54.68	224	51.33	213	34159	111	65.8
20	91	45.34	255	45.34	255	40.49	255	66.86	255	43322	316	43.3
21	96	64.23	255	64.30	255	61.79	219	96.67	219	44457	186	38.2
22	95	48.77	247	48.77	247	46.94	244	47.78	243	36905	240	46.8
23	95	61.31	255	61.31	255	69.82	255	70.85	255	76806	263	37.9
24	89	38.35	235	38.35	235	38.64	232	57.82	208	98774	236	13.3

Table 13 Objective measures for *CSS Zen Garden* web pages

4.4.1.2 Correlations

In order to test for correlations between the objective measures presented in Section 4.3.2 and the subjective ratings presented in Section 3.4.1.1, Spearman's rank Correlation Coefficient was used. The 95% confidence interval was also calculated. All tests were performed using R. The following six tables show the results of the correlation tests performed.

Table 14 below shows a negative correlation between Question A (How eye-catching do you find this page?) and compression ratio, along with a negative correlation with the green mode. It goes on to show a positive correlation between Question A (How eye-catching do you find this page?) and the number of unique colours.

Question A				
How eye-catching do you find this page?				
Objective measure	Spearman's Rank Correlation Coefficient (Rho)	95% Confidence Interval		
Compression ratio (%)	-0.44	-0.72	to	-0.05
GIMP value st dev	0.23	-0.19	to	0.58
GIMP value mode	-0.31	-0.64	to	0.10
GIMP red st dev	0.22	-0.20	to	0.57
GIMP red mode	-0.31	-0.64	to	0.10
GIMP green st dev	0.39	-0.02	to	0.68
GIMP green mode	-0.41	-0.70	to	-0.01
GIMP blue st dev	0.38	-0.03	to	0.68
GIMP blue mode	-0.40	-0.69	to	0.01
GIMP unique colours	0.51	0.13	to	0.76
Words	-0.32	-0.64	to	0.09
AQUINE	0.08	-0.33	to	0.47

Table 14 Correlation between Question A and objective measures

Table 15 below shows a negative correlation between Question B (How visually attractive do you find this page?) and compression ratio.

Question B				
How visually attractive do you find this page?				
Objective measure	Spearman's Rank Correlation Coefficient (Rho)	95% Confidence Interval		
Compression ratio (%)	-0.45	-0.73	to	-0.06
GIMP value st dev	-0.17	-0.54	to	0.25
GIMP value mode	-0.23	-0.58	to	0.19
GIMP red st dev	-0.17	-0.54	to	0.25
GIMP red mode	-0.23	-0.58	to	0.19
GIMP green st dev	-0.06	-0.45	to	0.35
GIMP green mode	-0.17	-0.54	to	0.25
GIMP blue st dev	0.00	-0.40	to	0.40
GIMP blue mode	-0.09	-0.48	to	0.32
GIMP unique colours	0.16	-0.26	to	0.53
Words	-0.24	-0.59	to	0.18
AQUINE	0.02	-0.39	to	0.42

Table 15 Correlation between Question B and objective measures

Table 16 below shows no correlation between Question C (How much does this page encourage you to explore the rest of the site?) and any of the objective measures.

Question C				
How much does this page encourage you to explore the rest of the site?				
Objective measure	Spearman's Rank Correlation Coefficient (Rho)	95% Confidence Interval		
Compression ratio (%)	-0.37	-0.67	to	0.04
GIMP value st dev	-0.05	-0.45	to	0.36
GIMP value mode	-0.29	-0.62	to	0.13
GIMP red st dev	-0.05	-0.44	to	0.36
GIMP red mode	-0.29	-0.62	to	0.13
GIMP green st dev	0.13	-0.29	to	0.50
GIMP green mode	-0.25	-0.59	to	0.17
GIMP blue st dev	0.12	-0.30	to	0.50
GIMP blue mode	-0.14	-0.52	to	0.28
GIMP unique colours	0.29	-0.13	to	0.62
Words	-0.32	-0.64	to	0.10
AQUINE	-0.06	-0.45	to	0.35

Table 16 Correlation between Question C and objective measures

Table 17 below shows no correlation between Question D (How well designed is this page?) and any of the objective measures.

Question D				
How well designed is this page?				
Objective measure	Spearman's Rank Correlation Coefficient (Rho)	95% Confidence Interval		
Compression ratio (%)	-0.31	-0.63	to	0.11
GIMP value st dev	-0.25	-0.59	to	0.17
GIMP value mode	-0.15	-0.52	to	0.27
GIMP red st dev	-0.21	-0.56	to	0.22
GIMP red mode	-0.15	-0.53	to	0.27
GIMP green st dev	-0.11	-0.49	to	0.31
GIMP green mode	-0.07	-0.46	to	0.34
GIMP blue st dev	-0.06	-0.45	to	0.35
GIMP blue mode	0.02	-0.39	to	0.42
GIMP unique colours	0.21	-0.22	to	0.56
Words	-0.18	-0.54	to	0.24
AQUINE	-0.14	-0.51	to	0.28

Table 17 Correlation between Question D and objective measures

Table 18 below shows no correlation between Question E (How good is this page as an advertisement for the website?) and any of the objective measures.

Question E				
How good is this page as an advertisement for the website?				
Objective measure	Spearman's Rank Correlation Coefficient (Rho)	95% Confidence Interval		
Compression ratio (%)	-0.39	-0.69	to	0.01
GIMP value st dev	-0.14	-0.52	to	0.28
GIMP value mode	-0.17	-0.54	to	0.25
GIMP red st dev	-0.11	-0.49	to	0.31
GIMP red mode	-0.18	-0.54	to	0.25
GIMP green st dev	-0.02	-0.42	to	0.39
GIMP green mode	-0.12	-0.50	to	0.30
GIMP blue st dev	0.03	-0.37	to	0.43
GIMP blue mode	-0.02	-0.42	to	0.38
GIMP unique colours	0.34	-0.08	to	0.65
Words	-0.18	-0.54	to	0.24
AQUINE	-0.07	-0.46	to	0.35

Table 18 Correlation between Question E and objective measures

Table 19 below shows no correlation between Question F (How much does this page discourage you to explore the rest of the site?) and any of the objective measures.

Question F				
How much does this page discourage you to explore the rest of the site?				
Objective measure	Spearman's Rank Correlation Coefficient (Rho)	95% Confidence Interval		
Compression ratio (%)	0.26	-0.16	to	0.60
GIMP value st dev	0.33	-0.08	to	0.65
GIMP value mode	0.10	-0.32	to	0.48
GIMP red st dev	0.29	-0.13	to	0.62
GIMP red mode	0.10	-0.32	to	0.48
GIMP green st dev	0.24	-0.18	to	0.59
GIMP green mode	0.01	-0.40	to	0.41
GIMP blue st dev	0.13	-0.28	to	0.51
GIMP blue mode	-0.11	-0.49	to	0.31
GIMP unique colours	-0.10	-0.48	to	0.32
Words	0.14	-0.28	to	0.51
AQUINE	0.03	-0.38	to	0.43

Table 19 Correlation between Question F and objective measures

4.4.2 Study Two: Hot air balloon experiences

4.4.2.1 Objective measures for Study Two

The following two tables show the values for the objective measures gathered for the hot air balloon cards.

Card	Compression ratio (%)	Number of words	Number of Numbers	ImageJ value st dev	ImageJ value mode	GIMP unique colours
1	83	42	1	44.09	134	44126
2	77	211	9	54.63	255	38162
3	60	36	7	84.95	255	131203
4	81	87	2	59.20	255	30145
5	61	108	3	47.44	255	60277
6	72	139	14	64.35	255	68454
7	77	114	11	86.42	255	73012
8	82	182	9	52.11	255	28984
9	88	133	6	77.64	255	21143
10	74	98	0	61.09	255	69941
11	88	150	21	50.77	255	23515
12	74	79	9	73.49	255	75822
13	74	187	10	71.86	255	74981
14	56	74	3	67.31	255	137556
15	75	142	1	46.36	142	38895
16	65	120	6	65.12	255	143607
17	64	84	2	76.30	255	150277
18	78	135	7	55.14	188	46946
19	71	77	18	70.13	255	99483
20	61	75	0	59.40	239	135673
21	59	29	2	78.11	0	174726
22	72	92	4	62.42	230	51277
23	67	146	16	59.72	255	80186
24	69	63	0	43.96	206	95553
25	65	34	0	54.69	90	47663
26	90	217	13	55.26	252	12646
27	73	147	4	71.85	255	63621
28	67	53	0	68.77	255	76675
29	69	38	2	60.72	255	75436
30	55	154	2	71.51	255	165265
min	55	29	0	44	0	12646
max	90	217	21	86	255	174726

Table 20 Objective measures for hot air balloon web pages

Card	ImageJ red mean	ImageJ red st dev	ImageJ red mode	ImageJ green mean	ImageJ green st dev	ImageJ green mode	ImageJ blue mean	ImageJ blue st dev	ImageJ blue mode
1	135.44	55.77	83	160.91	50.27	129	192.91	67.5	191
2	189.68	87.93	255	221.42	51.87	255	236.27	47.5	255
3	151.93	90.86	255	159.04	93.10	255	153.67	83.0	255
4	210.51	64.25	255	213.88	62.23	255	222.53	58.1	255
5	171.71	63.38	255	198.62	46.99	255	209.99	54.2	221
6	189.65	75.79	255	198.24	66.18	255	203.72	69.6	255
7	172.69	98.94	255	172.67	92.54	255	184.52	79.9	255
8	218.09	60.34	255	225.7	49.51	255	228.85	53.9	255
9	168.73	99.21	255	186.05	77.53	255	201.25	62.9	255
10	185.24	72.47	255	193.86	62.93	255	206.43	72.5	255
11	217.21	60.22	255	226.37	56.27	255	238.94	49.3	255
12	208.04	78.44	255	208.97	74.84	255	209.71	76.4	255
13	157.81	99.04	255	183.71	74.72	255	208	64.0	255
14	173.26	83.69	255	188.66	68.10	255	193.34	86.0	255
15	129.53	62.71	95	162.62	48.00	140	197.02	39.6	192
16	180.08	79.74	255	194.7	64.21	255	205.31	60.8	255
17	191.63	67.84	255	169.84	82.98	255	153.13	96.3	255
18	197.79	50.79	180	191.4	66.29	191	195.64	64.8	193
19	186.37	80.71	255	197.56	70.74	255	208.62	65.9	255
20	150.21	68.34	229	164.46	61.86	237	187.17	64.9	250
21	124.81	76.99	0	109.64	84.25	0	114.57	82.0	0
22	195.62	59.73	230	187.4	67.25	230	186.89	67.5	230
23	202.52	62.96	255	206.52	58.36	255	202.44	72.0	255
24	170.23	55.48	198	188.55	45.76	207	204.63	44.3	214
25	59.904	81.85	0	131.94	51.55	104	175.73	43.4	165
26	220.49	58.59	253	220.81	57.00	252	203.87	83.9	252
27	163.53	92.97	255	184.74	74.32	255	197.11	72.9	255
28	198.48	64.98	255	198.86	71.00	255	201.56	75.7	255
29	207.04	67.77	255	208.6	66.69	255	214.37	62.5	255
30	178.69	81.05	255	175.41	79.30	255	189.18	81.4	255
min	60	51	0	110	46	0	115	40	0
max	220	99	255	226	93	255	239	96	255

Table 21 Continuation of Objective measures in Table 20

4.4.2.2 Correlations for Study Two

In order to test for correlations between the hot air balloon cards, and the objective measures presented in Section 4.3.2 and the subjective ratings presented in Section 3.6.2.1, Spearman's

rank Correlation Coefficient was used. The 95% confidence interval was also calculated. All tests were performed using Excel and the confidence intervals were calculated using a spread sheet from *A new view of statistics* (Hopkins, 2013). The following six tables show the results of the correlation tests performed. Each table is for a different question from the subjective values.

Table 22 below shows a negative correlation between Question A (How eye-catching do you find this page?) and compression ratio, number of words, and number of numbers. It shows a positive correlation between Question A (How eye-catching do you find this page?) and the blue standard deviation, the green standard deviation, and the number of unique colours.

Question A				
How eye-catching do you find this page?				
Objective measure	**Spearman's Rank Correlation Coefficient (Rho)**	**95% Confidence Interval**		
Compression ratio (%)	-0.78	-0.89	to	-0.59
Number of words	-0.56	-0.76	to	-0.25
Number of numbers	-0.54	-0.75	to	-0.22
ImageJ value st dev	0.37	0.02	to	0.65
ImageJ value mode	-0.02	-0.37	to	0.35
GIMP unique colours	0.72	0.48	to	0.86
ImageJ red st dev	0.20	-0.17	to	0.52
ImageJ red mode	-0.02	-0.38	to	0.34
ImageJ green st dev	0.39	0.03	to	0.66
ImageJ green mode	-0.02	-0.38	to	0.34
ImageJ blue st dev	0.51	0.18	to	0.73
ImageJ blue mode	-0.02	-0.38	to	0.34

Table 22 Correlation between Question A and objective measures

Table 23 below, shows a negative correlation between Question B (How visually attractive do you find this page?) and compression ratio, number of words, and number of numbers. It shows a positive correlation between Question B (How visually attractive do you find this page?) and the number of unique colours.

Question B				
How visually attractive do you find this page?				
Objective measure	Spearman's Rank Correlation Coefficient (Rho)	95% Confidence Interval		
Compression ratio (%)	-0.65	-0.82	to	-0.38
Number of words	-0.47	-0.71	to	-0.13
Number of numbers	-0.51	-0.73	to	-0.18
ImageJ value st dev	0.28	-0.09	to	0.58
ImageJ value mode	0.14	-0.23	to	0.48
GIMP unique colours	0.50	0.17	to	0.73
ImageJ red st dev	0.21	-0.16	to	0.53
ImageJ red mode	0.10	-0.27	to	0.44
ImageJ green st dev	0.25	-0.12	to	0.56
ImageJ green mode	0.14	-0.23	to	0.48
ImageJ blue st dev	0.35	-0.01	to	0.63
ImageJ blue mode	0.19	-0.19	to	0.51

Table 23 Correlation between Question B and objective measures

Table 24 below shows a negative correlation between Question C (How much does this page encourage you to explore the rest of the site?) and compression ratio. Table 24 also shows a positive correlation between Question C (How much does this page encourage you to explore the rest of the site?) and the blue standard deviation, and the number of unique colours.

Question C				
How much does this page encourage you to explore the rest of the site?				
Objective measure	Spearman's Rank Correlation Coefficient (Rho)	95% Confidence Interval		
Compression ratio (%)	-0.56	-0.77	to	-0.25
Number of words	-0.29	-0.59	to	0.08
Number of numbers	-0.36	-0.64	to	0.00
ImageJ value st dev	0.36	0.00	to	0.64
ImageJ value mode	0.30	-0.06	to	0.60
GIMP unique colours	0.51	0.18	to	0.73
ImageJ red st dev	0.28	-0.08	to	0.58
ImageJ red mode	0.27	-0.10	to	0.58
ImageJ green st dev	0.32	-0.04	to	0.61
ImageJ green mode	0.30	-0.07	to	0.60
ImageJ blue st dev	0.41	0.06	to	0.67
ImageJ blue mode	0.33	-0.03	to	0.62

Table 24 Correlation between Question C and objective measures

Table 25 below shows a negative correlation between Question D (How well designed is this page?) and the compression ratio. Table 25 below shows a positive correlation between Question D (How well designed is this page?) and the value mode, the red mode, the green mode, the blue standard deviation, the blue mode, and the number of unique colours.

Question D				
How well designed is this page?				
Objective measure	Spearman's Rank Correlation Coefficient (Rho)	95% Confidence Interval		
Compression ratio (%)	-0.40	-0.66	to	-0.05
Number of words	-0.12	-0.46	to	0.25
Number of numbers	-0.28	-0.58	to	0.09
ImageJ value st dev	0.33	-0.03	to	0.62
ImageJ value mode	0.41	0.06	to	0.67
GIMP unique colours	0.37	0.01	to	0.64
ImageJ red st dev	0.23	-0.14	to	0.54
ImageJ red mode	0.38	0.03	to	0.65
ImageJ green st dev	0.29	-0.08	to	0.59
ImageJ green mode	0.41	0.06	to	0.67
ImageJ blue st dev	0.43	0.08	to	0.68
ImageJ blue mode	0.43	0.08	to	0.69

Table 25 Correlation between QuestionD and objective measures

Table 26 below shows a negative correlation between Question E (How good is this page as an advertisement for the website?) and compression ratio, and the number of numbers. Table 26 also shows a positive correlation between Question E (How good is this page as an advertisement for the website?) and the blue standard deviation, and the number of unique colours.

Question E				
How good is this page as an advertisement for the website?				
Objective measure	Spearman's Rank Correlation Coefficient (Rho)	95% Confidence Interval		
Compression ratio (%)	-0.55	-0.76	to	-0.23
Number of words	-0.27	-0.57	to	0.10
Number of numbers	-0.39	-0.66	to	-0.03
ImageJ value st dev	0.35	-0.01	to	0.63
ImageJ value mode	0.25	-0.13	to	0.56
GIMP unique colours	0.48	0.15	to	0.72
ImageJ red st dev	0.19	-0.18	to	0.51
ImageJ red mode	0.23	-0.15	to	0.54
ImageJ green st dev	0.33	-0.03	to	0.62
ImageJ green mode	0.25	-0.13	to	0.56
ImageJ blue st dev	0.46	0.12	to	0.70
ImageJ blue mode	0.26	-0.11	to	0.57

Table 26 Correlation between Question E and objective measures

Table 27 below, shows a negative correlation between Question F (How much does this page discourage you to explore the rest of the site?) and the value mode, the green mode, the blue mode.

Question F				
How much does this page discourage you to explore the rest of the site?				
Objective measure	Spearman's Rank Correlation Coefficient (Rho)	95% Confidence Interval		
Compression ratio (%)	0.35	-0.01	to	0.63
Number of words	0.23	-0.14	to	0.54
Number of numbers	0.28	-0.08	to	0.58
ImageJ value st dev	-0.35	-0.63	to	0.01
ImageJ value mode	-0.37	-0.64	to	-0.01
GIMP unique colours	-0.36	-0.64	to	0.00
ImageJ red st dev	-0.30	-0.60	to	0.07
ImageJ red mode	-0.34	-0.62	to	0.02
ImageJ green st dev	-0.30	-0.60	to	0.07
ImageJ green mode	-0.37	-0.64	to	-0.01
ImageJ blue st dev	-0.34	-0.63	to	0.02
ImageJ blue mode	-0.41	-0.67	to	-0.05

Table 27 Correlation between Question F and objective measures

4.5 Discussion

4.5.1 Study One: CSS Zen Garden

4.5.1.1 Objective measures for Study One

Rank	Comp-ression ratio (%)	GIMP value st dev	GIMP value mode	GIMP red st dev	GIMP red mode	GIMP green st dev	GIMP green mode	GIMP blue st dev	GIMP blue mode	GIMP unique colours	Words	AQUINE
1	4	5	18	5	18	14	18	14	9	1	2	1
2	21	14	21	14	12	5	23	21	23	24	10	6
3	16	18	23	1	21	8	12	18	20	10	13	19
4	22	1	9	18	23	18	9	1	16	14	16	5
5	23	4	12	4	9	1	11	9	2	18	7	15
6	7	6	11	12	13	23	10	8	10	23	12	22
7	9	21	20	6	20	6	2	23	13	8	20	14
8	13	23	13	15	11	21	20	6	11	13	18	20
9	3	9	10	21	10	12	7	20	12	6	23	2
10	12	19	17	23	17	9	13	3	7	2	6	4
11	2	12	2	9	2	19	17	24	17	12	17	17
12	6	22	16	19	16	4	16	16	3	7	22	7
13	5	15	7	22	7	15	3	2	22	4	24	21
14	8	11	22	13	22	11	22	4	21	9	11	23
15	11	20	3	20	3	22	24	10	19	11	4	13
16	15	13	24	11	24	10	19	13	24	5	3	11
17	17	8	19	8	19	2	21	19	15	21	8	18
18	20	10	15	24	8	20	15	11	4	16	15	9
19	19	24	8	10	15	24	4	22	8	15	21	8
20	24	17	4	17	4	7	8	12	14	20	5	16
21	10	3	6	3	6	13	6	7	18	3	9	12
22	14	2	5	2	5	17	14	17	1	17	1	10
23	1	16	14	16	14	3	5	15	6	22	19	24
24	18	7	1	7	1	16	1	5	5	19	14	3

Table 28 Pages ranked based on objective values (largest to smallest)

The following eight figures show the images ranked in the highest and lowest five for four of the objective measures. The measures shown are compression ratio, unique colours, number of words, and value standard deviation.

Image 4

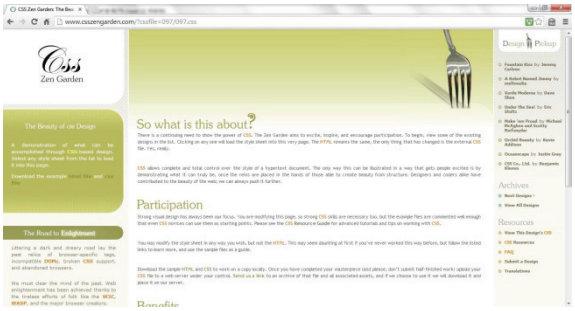


Image 16

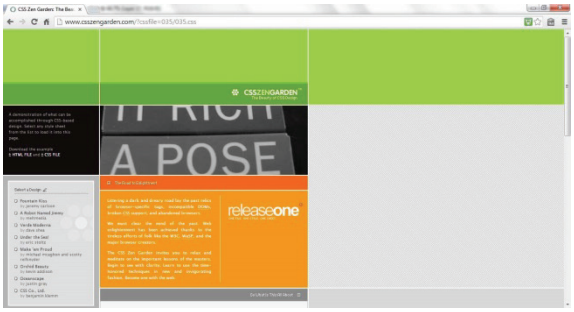


Image 21

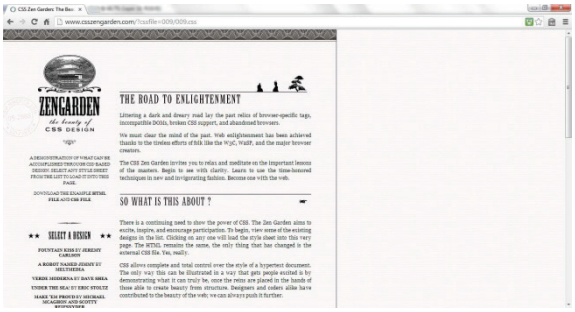


Image 22

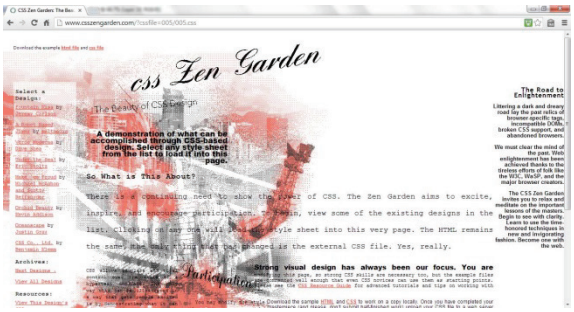


Image 23

Figure 17 Highest five images for compression ratio

The images displayed in Figure 17 above were ranked in the highest five for compression ratio. This means that a higher level of compression was possible when converting the images from a lossless file format into a lossy file format. By reviewing the images in this set it can be seen that they all contain a large amount of space in a single colour. This has probably contributed to the ability to compress the file, as there are large groups of neighbouring pixels with the same colour. The colour is a pale background in four cases.

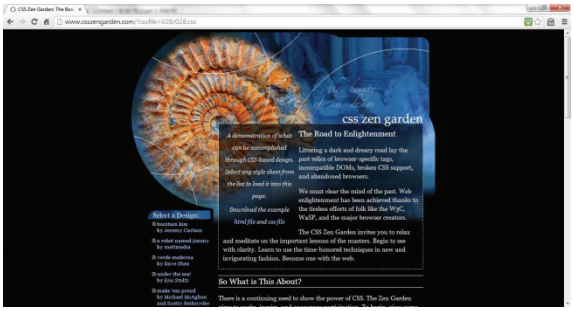


Image 1

Image 10

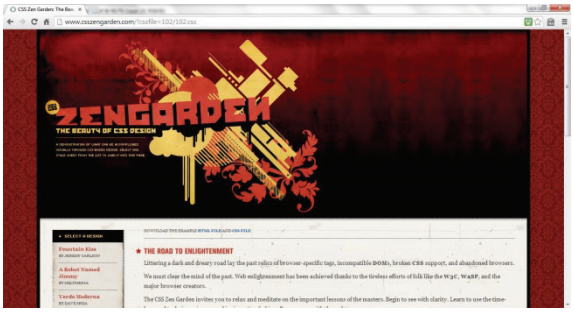


Image 14

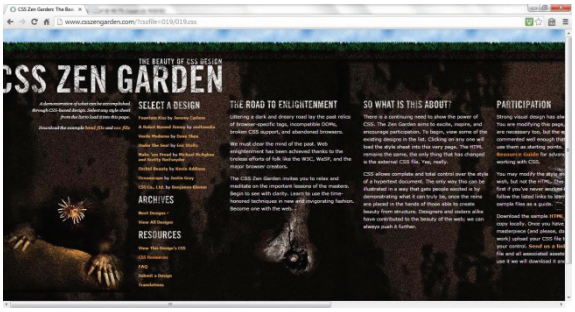


Image 18

Image 24

Figure 18 Lowest five images for compression ratio

The images displayed in Figure 18 above were ranked in the lowest five for compression ratio. This means that a lower level of compression was possible when converting the images from a lossless file format into a lossy file format. By reviewing the cards in this set is can be seen that four contain coloured backgrounds. Three of the backgrounds include an image as a large proportion of the background. This makes compression more challenging. Image 10 has a white background, however, it also contains a number of sections with multiple vibrant colours.

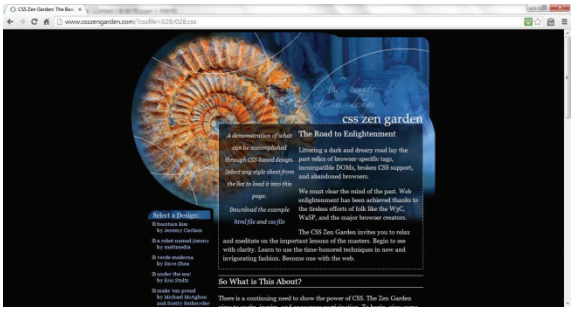


Image 1

Image 10

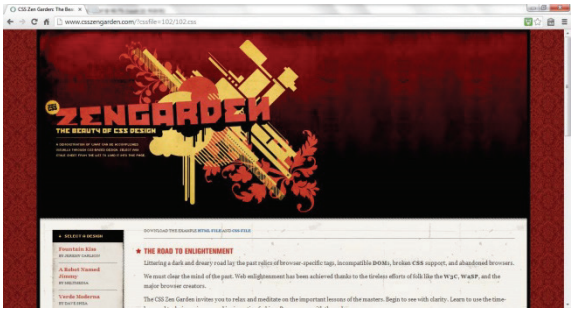


Image 14

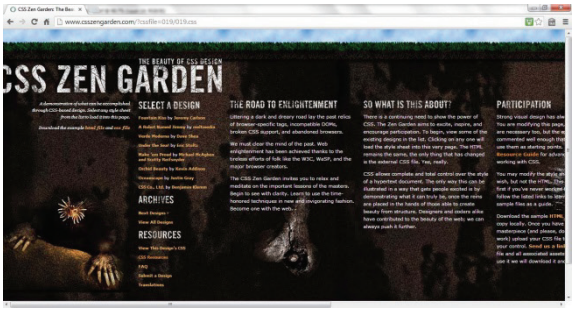


Image 18

Image 24

Figure 19 Highest five images for unique colours

The images in Figure 19 above were ranked in the highest five for unique colours. It is interesting to note that this is exactly the same set of five images as in Figure 18, showing the images with the least compression.

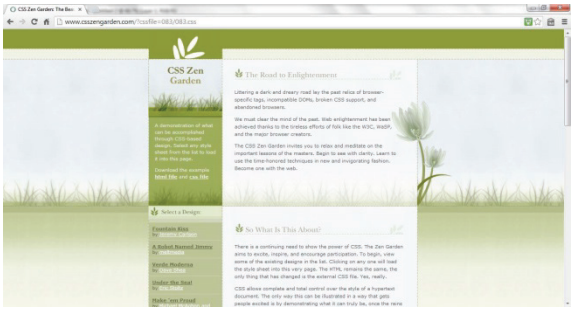


Image 3

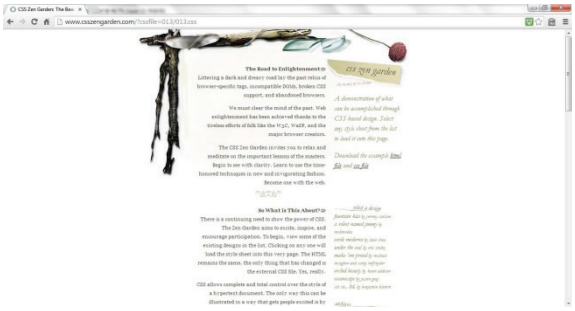


Image 17

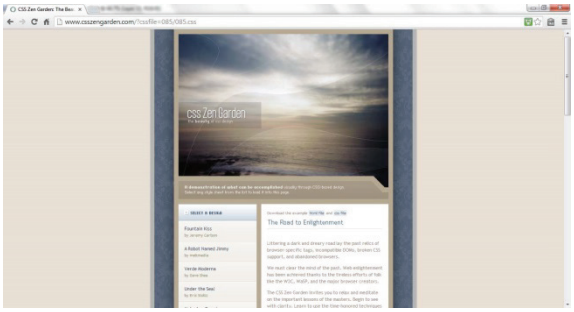


Image 19

Image20

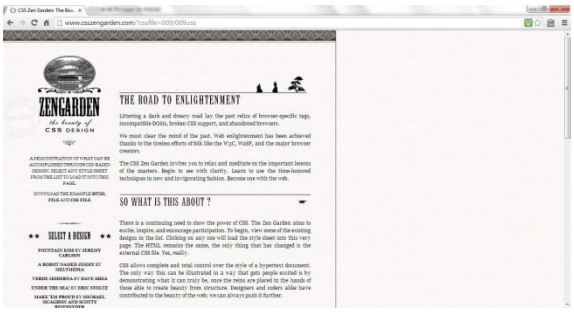


Image 22

Figure 20 Lowest five images for unique colours

The images in Figure 20 above were ranked in the lowest five for unique colours.

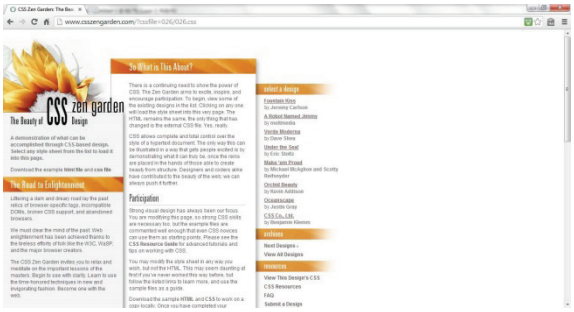


Image 2

Image 7

Image 10

Image 13

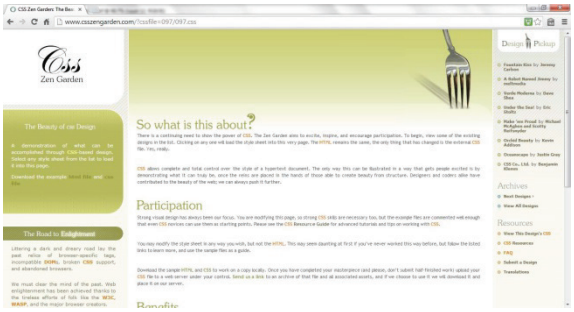


Image 16

Figure 21 Highest five images for number of words

The images in Figure 21 above were ranked in the highest five for number of words. Interestingly, in the case of the five images shown here, the pages also have a large amount of white space. This white space, coupled with a large amount of text, suggests the information that is present is not distributed evenly across the page. In the correlational study the number of words was not found to correlate with any of the subjective ratings.

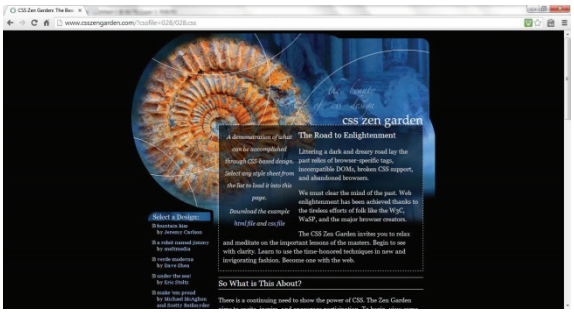


Image 1

Image 5

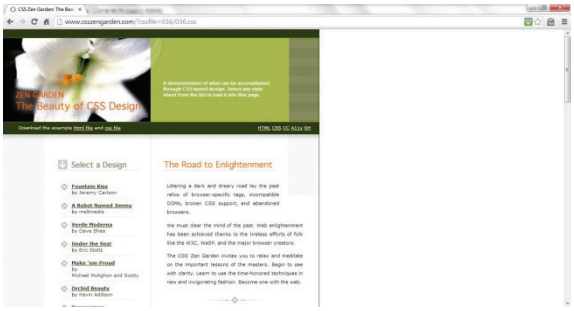


Image 9

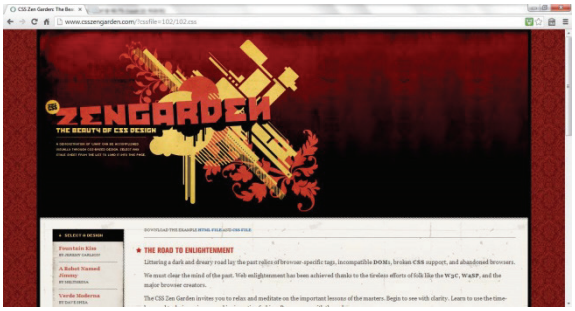


Image 14

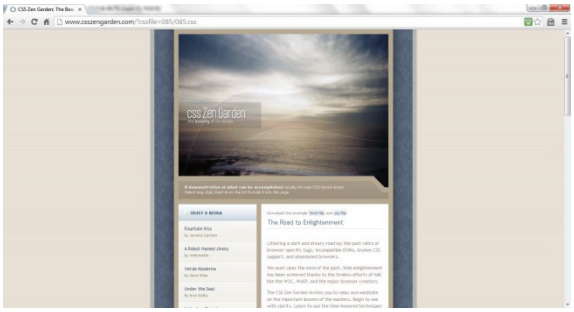


Image 19

Figure 22 Lowest five images for number of words

The images in Figure 22 above were ranked in the lowest five for number of words. By reviewing the set of images it can be seen they all incorporate a large amount of space around the content. In four cases this space is not white, which may add visual interest to the page. Four of the images place the text content in a central panel, which contributes to a feeling of symmetry about the pages.

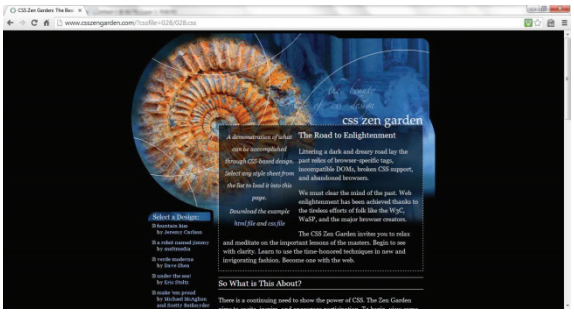


Image 1

Image 4

Image 5

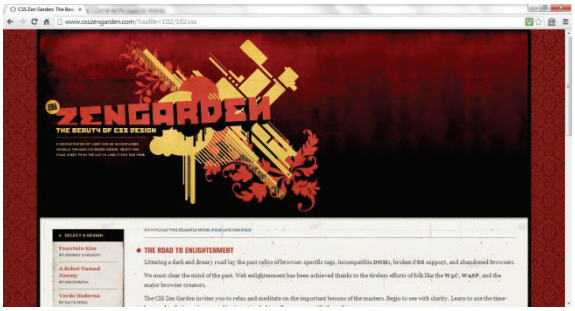


Image 14

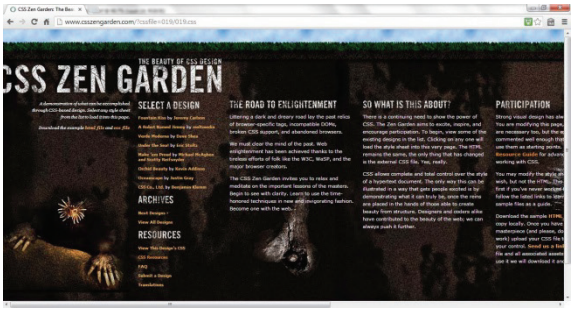


Image 18

Figure 23 Highest five images for value standard deviation

The images in Figure 23 above were ranked in the highest five for value standard deviation. This means there is a high level of colour contrast within these pages. By reviewing the five images in this set it can be seen that they all use dark colours. In two cases, Image 1 and Image 18, the text has an image behind it.

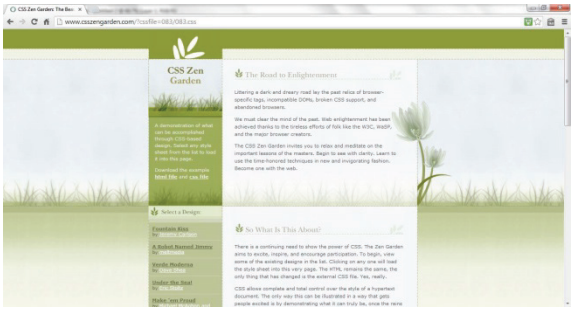


Image 3

Image 10

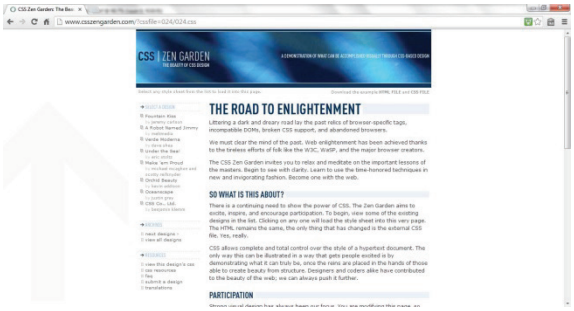


Image 12

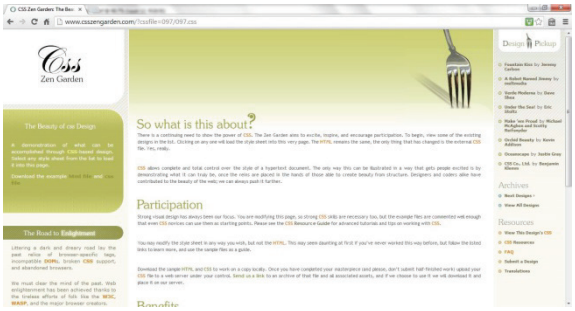


Image 16

Image 24

Figure 24 Lowest five images for value standard deviation

The images in Figure 24 above were ranked in the lowest five for value standard deviation. By reviewing the set of image sit can be seen that two images, Image 10 and Image 12, have white backgrounds. An unexpected pattern was that the remaining three images all make use of the colour green.

4.5.1.2 Correlations for Study One

As predicted, there was a negative correlation found between the compression ratio and Question A (How eye-catching do you find this page?). Eye catching pages often make use of a number of colours and contrast in order to stand out. This in turn makes a page more difficult to compress, due to the amount of visual contrast present.

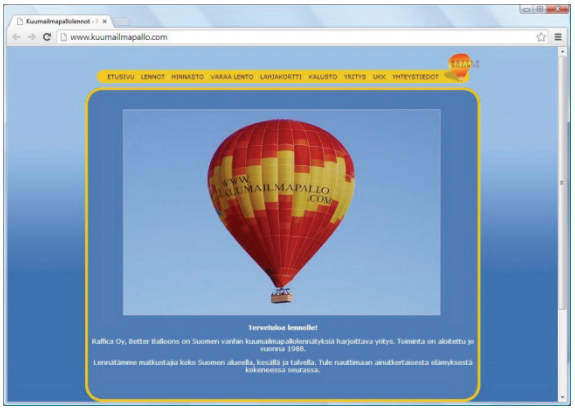
There were no correlations found between any of the objective measures and questions C-F (How much does this page encourage you to explore the rest of the site?; How well designed is this page?; How good is this page as an advertisement for the website?; How much does this page discourage you to explore the rest of the site?). One limitation from the images used for Study

One occurred from the sample of pages used. They were drawn from a website showcasing the power of cascading style sheets. Therefore, the perceived level of web design within these pages was high. This may have led to smaller variation than from an alternative set of pages. This limitation was addressed when selecting the pages for Study Two.

4.5.2 Study Two: Hot air balloon experience

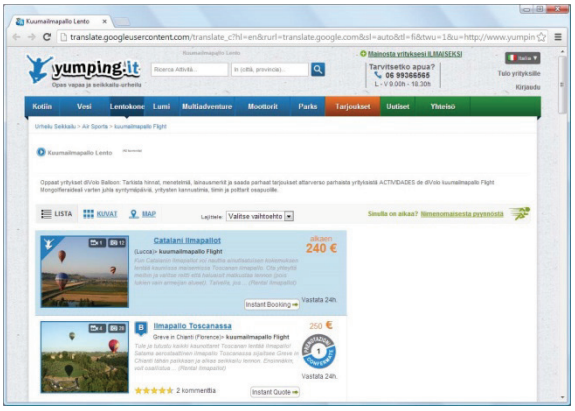
4.5.2.1 Objective measures for Study Two

The following six figures show the cards ranked in the highest and lowest six for three of the measures. The measures shown are compression ratio, number of words, and value standard deviation.

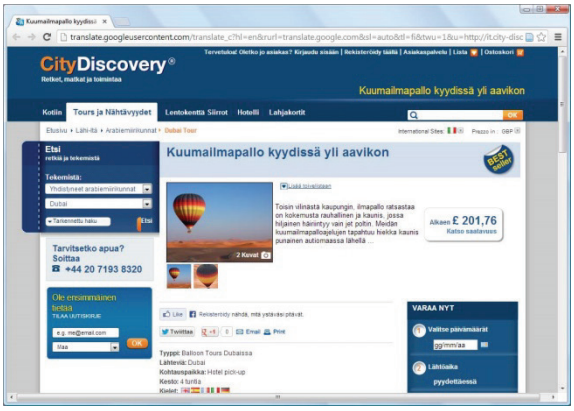


Card 1

Card 4

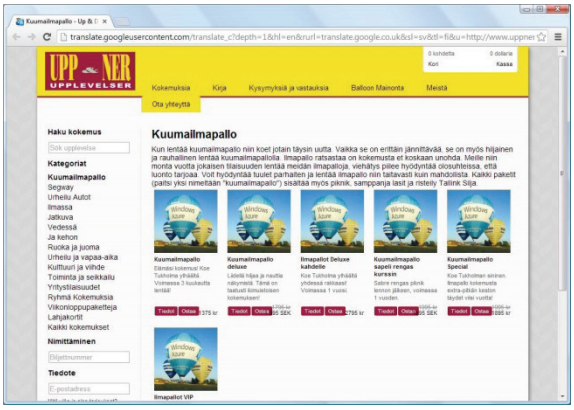


Card 8



Card 9

Card 11



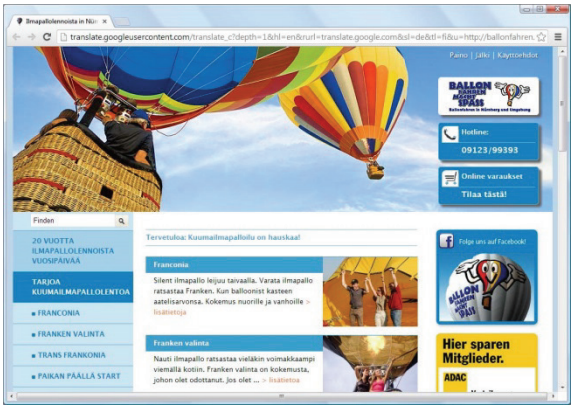
Card 26

Figure 25 Highest six cards for compression ratio

The cards displayed in Figure 25 above were ranked in the highest six for compression ratio. This means that a higher level of compression was possible when converting the cards from a lossless file format into a lossy file format. By reviewing the cards in this set it can be seen that the majority have a white background. The only exception is Card 1, which has a blue background. There is also a large amount of white space, proportional to the card size. This has probably contributed to the ability to compress the card, as neighbouring pixels in a uniform colour can often be stored efficiently. The size of the images present in the set of cards displayed in Figure 25 is small for all cards except for Card 1. A comparison of the set of cards in Figure 25 above, and the set of cards with a small number of words, shown in Figure 22, has been reported in the text after Figure 22.

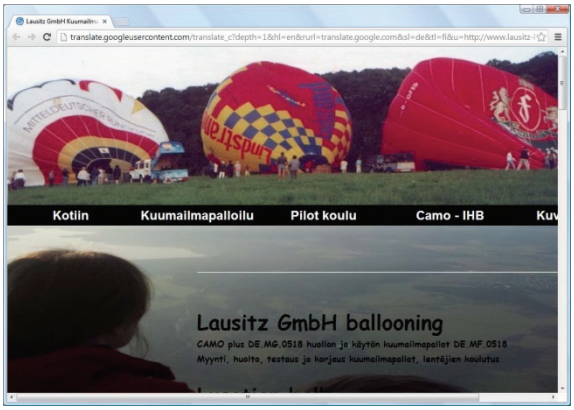
Card 3

Card 5



Card 14

Card 20



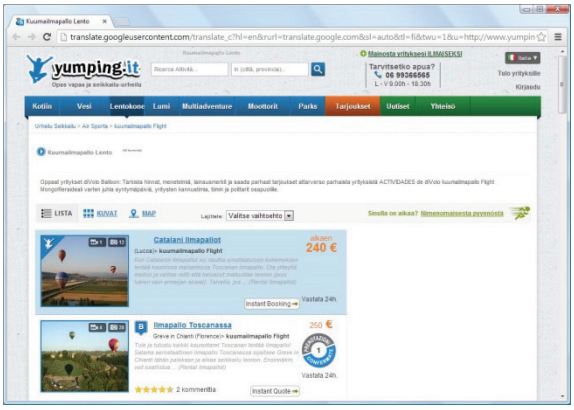
Card 21

Card 30

Figure 26 Lowest six cards for compression ratio

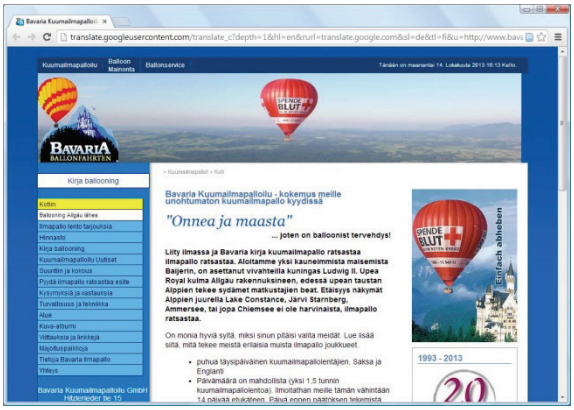
The cards displayed in Figure 26 above were ranked in the lowest six for compression ratio. This means that a lower level of compression was possible when converting the cards from a lossless file format into a lossy file format. By reviewing the cards in this set it can be seen that they all make use of a large number of colours and large images. It is logical that with the presences of lots of colours and large images the level of compression available was smaller than cards with a smaller number of colours or images. It is also logical to predict the set of cards in Figure 26 above should have some overlap with those exhibiting a large standard deviation. This point is addressed after Figure 29.

Card2

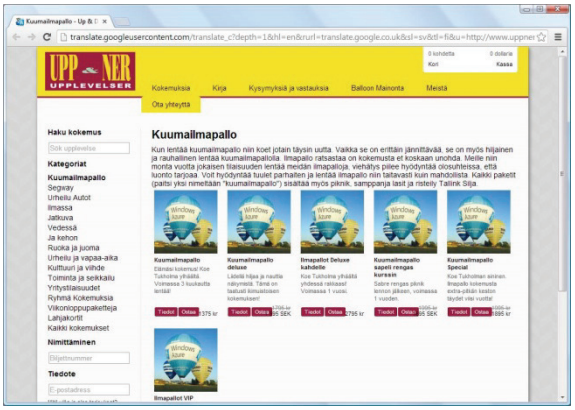


Card 8

Card 11



Card 13

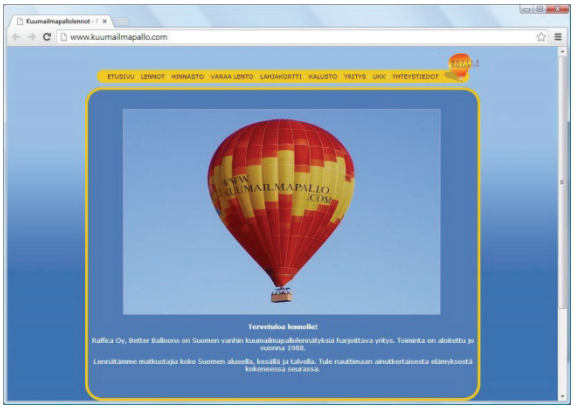


Card 26

Card 30

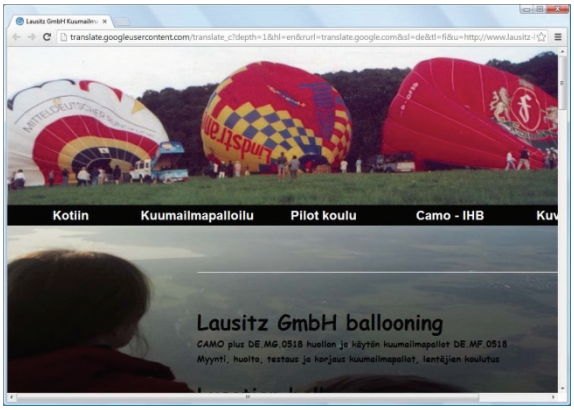
Figure 27 Highest six cards for number of words

The cards displayed in Figure 27 above were ranked in the highest six for number of words. It is conceivable that the purpose of a page is less clear on a cluttered page. Usability heuristics (Nielsen,1995) suggest pages should have a minimalist design, and a large number of words visible goes against this heuristic.



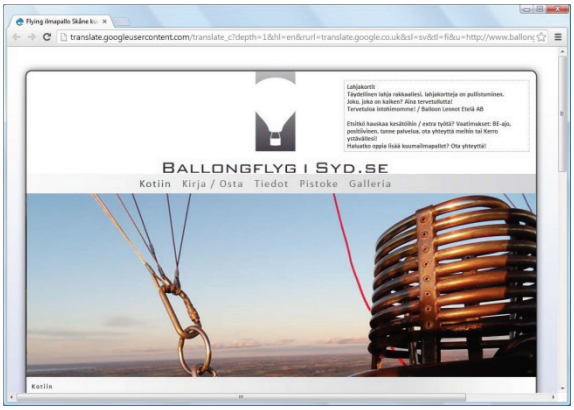
Card 1

Card 3

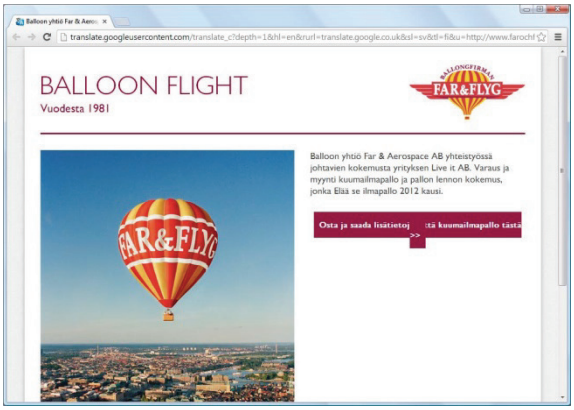


Card 21

Card 25



Card 28



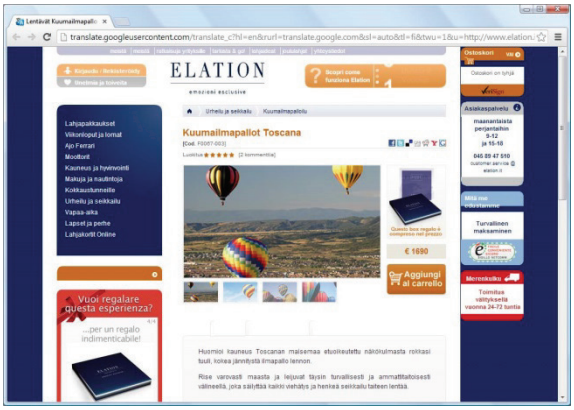
Card 29

Figure 28 Lowest six cards for number of words

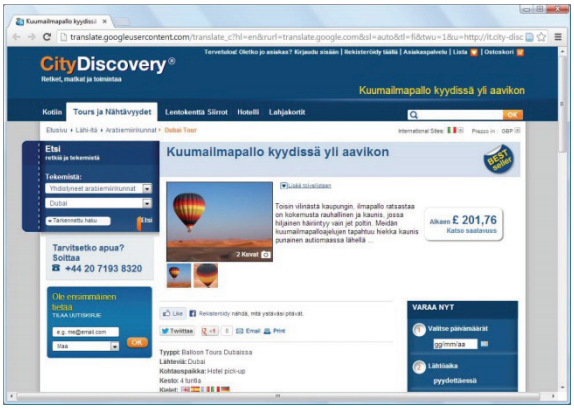
The cards displayed in Figure 28 above were ranked in the lowest six for number of words. From a text perspective these pages follow the usability heuristic regarding a minimalist design, however, the other aspects of visual interest do not necessarily also follow this heuristic. The compression ration better takes into account all visual aspects of the page. A comparison of the

set of cards in Figure 25 (the highest six cards for compression ratio), and the set of cards with a small number of words, shown in Figure 28 above, shows only one card, Card 1, is present in both sets. This finding adds strength to the argument for recording both values, and for looking for correlations between the subjective values and both measures.

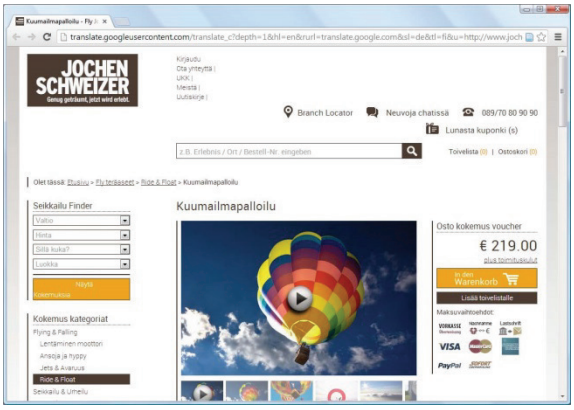
Card 3



Card 7

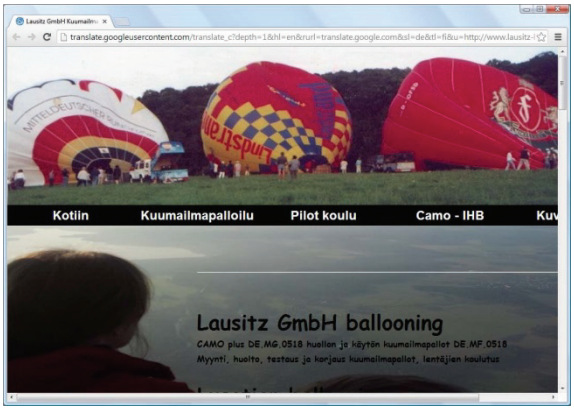


Card 9



Card 12

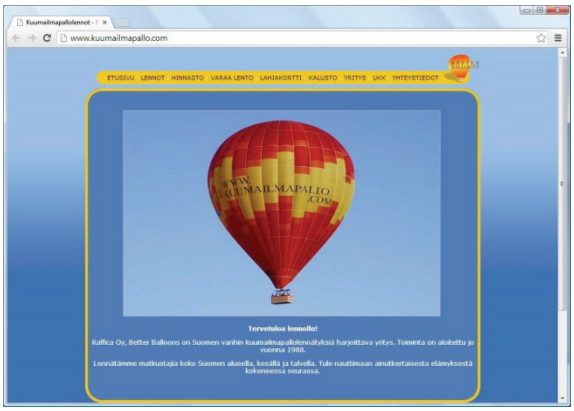
Card 17



Card 21

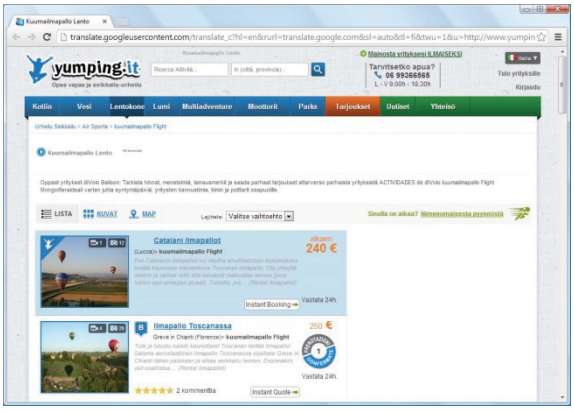
Figure 29 Highest six cards for value standard deviation

The cards displayed in Figure 29 above were ranked in the highest six for value standard deviation. This means that the level of colour contrast within these pages was high. This measure does not take into account the intensity of the colour. The majority of the cards in Figure 29 use quite dark colours. Interestingly, the only card with predominantly bright colours is Card 17. When reviewing the set of cards in Figure 26 (lowest six for compression ratio) and the cards in Figure 29 above (highest six for value standard deviation), it can be seen that there are two cards in both sets, Card 3 and Card 21. This finding adds strength to the argument for recording both values, and looking for correlations between the subjective values and both measures.



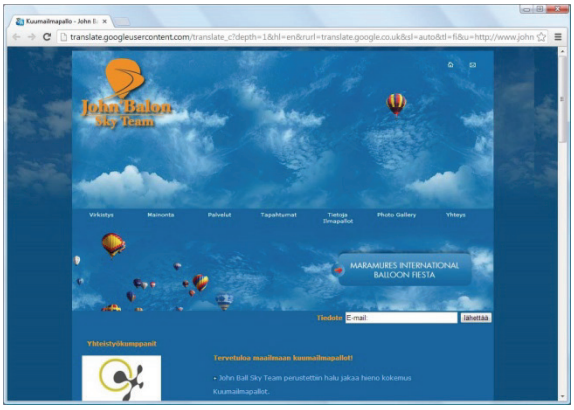
Card 1

Card 5



Card 8

Card 11



Card 15 Card 24

Figure 30 Lowest six cards for value standard deviation

The cards displayed in Figure 30 above were ranked in the lowest six for value standard deviation. This means that the level of colour contrast within these pages was low. The colour most prevalent within these pages was blue, with four cards using it as their main colour.

4.5.2.2 Correlations for Study Two

Objective measure	Number of correlations	Direction of correlations
Compression ratio (%)	5	Negative
Number of words	2	Negative
Number of numbers	3	Negative
ImageJ value st dev	1	Positive
ImageJ value mode	2	1 Positive and 1 negaitve
GIMP unique colours	5	Positive
ImageJ red st dev		
ImageJ red mode	1	Positive
ImageJ green st dev	1	Positive
ImageJ green mode	2	1 Positive and 1 negaitve
ImageJ blue st dev	4	Positive
ImageJ blue mode	2	1 Positive and 1 negaitve

Table 29 Frequency and direction of correlations

Correlations were found between the majority of the objective measures and one or more question. There were five correlations found with both the number of unique words and the compression ratio. The correlations found with the number of unique colours were positive for all questions apart from Question F (How much does this page discourage you to explore the rest of the site?). This finding is consistent with those of Holland (2004). This was also consistent for the

findings of Study One, in relation to Question A (How eye-catching do you find this page?). There were also five correlations found with the compression ratio. The negative correlation shows links between those pages that could be compressed to a lesser degree. This finding is particularly interesting due to web design guidance to create a minimalist design (Nielsen, 1995). However, designs that users perceive to be minimalist may not be the same set that an objective measurement of complexity would predict to be minimalist. The finding regarding compression ratio also contrasts with the finding that for two questions the number of words is negatively correlated. A finding in accordance with that expected from design heuristics. There were four correlations found with the blue standard deviation, suggesting a variance in the level of blue present on the page was important not just the presence of blue. This finding may be domain specific. The relevance of blue to the domain was uncovered through the card sort and laddering activities, and this discovery is covered in Section 5.6.8.

4.6 Conclusion

This chapter has described two related studies that gathered objective measurements of web pages and performed a correlational study between these ratings and the subjective ratings gathered in Chapter 3.

The key conclusions from the research described in this chapter are:

- Tentative associations were observed between a number of the measures and the subjective ratings. The most promising of these were:
    - number of unique colours in the web page,
    - compression ratio (percentage compression possible), and
    - standard deviation of the blue component on the image histogram.

In relation to the number of unique colours present in the web page, the study showed a positive correlation with five of the survey questions and a negative correlation with one of the survey

questions. By reviewing the confidence intervals the null hypothesis could be rejected in each case. This indicates that the more diverse the colour palette of a page, the more positively it will be perceived.

The study showed that the less a page could be compressed the more highly rated it was by the participants, a negative correlation. The compression ratio was used as a proxy for the amount of information in the page, and the study found those pages with more information were rated more highly. Whilst this may appear to be contrary to Nielsen's (1995) guidance regarding minimalist design, it should be remembered that compression ratio is a crude proxy and factors such as the level of detail in images could have a significant impact on the measure. Further investigation could narrow which complexity factors lead to this association.

In relation to the level of blue within the web page, the standard deviation was measured and the larger this value, the more positive the reaction to the web page. A high standard deviation reflects a high level of variance of shades of blue (brightness) used within the page. The blue component measures all occurrences of blue within the page, not just those occurrences that would be perceived as blue by the user, but there is clearly a positive association with this variable and it would still be possible to use it within an automated assessment. The colour identified for this domain has specific domain relevance, which is discussed in more detail in Chapter 5. The existence of this domain specific colour link, shows the importance of identifying colours which have a special association with a topic.

4.6.1 Limitations and further work

In light of these findings, further surveys or experiments should be conducted considering each measurement in isolation. If the associations are found in subsequent studies, software can be produced that will assess web page designs based on these criteria, and allow recommendations to be made regarding optimising the pages for these factors. This ability would allow feedback without the need for user testing. Blue standard deviation appears to be a domain specific

finding: any subsequent work in a different domain should incorporate elicitation of domain specific colours.

The correlational results from Study One found only four correlations, however the results from Study Two found twenty one. The measure that was correlated in the most questions was number of colours. The set of measures used in the two studies described in this chapter did not find consistent results, suggesting the set of objective measures used was not optimum. However, the findings from Study Two, combined with those from Holland (2004), indicate this is an avenue worthy of further exploration.

The limitations with the data collection discussed in Chapter 3 also relate to the findings from this chapter, as some the information was used in the correlational study. These have been discussed in Section 3.8.1.

This study reviewed colour through the overall histogram, and the RGB component histograms. These were sensible starting points as these are the components of direct light as used in computer displays. Now that some level of association is evident the next stage would be to check for correlations in other component models such as the CYM (cyan, magenta, yellow) of reflected light images or non-component hues such as secondary or tertiary colours like purple.

Chapter 5 Using card sorts to elicit visual attributes of Web pages

5.1 Introduction

A key theme within this thesis is the identification of attributes which correlate with subjective opinions of web pages. A well-established technique for identifying attributes is card sorts (Upchurch *et al.*, 2001). In the study described in this chapter, a card sorting technique has been used to elicit information regarding categorisations used by web page users. This information has been used to investigate nomothetic and idiographic preferences within the domain of web design.

The purpose of this chapter is to investigate whether a categorisation task is a suitable elicitation technique to identify visual attributes used to form subjective opinions of web pages. Two sub-questions will also be investigated. The first is what level of agreement exists between individuals. The second is whether sort criteria are revealed in an order linked to their importance to the participant.

In order to address these questions a card sorting activity was performed using a set of 30 cards. The cards contained screen shots of web pages for companies selling hot air balloon experiences. The set of cards was consistent with those used for Study Two, described in Chapter 3. The domain decision has been explored in Section 3.6.2.1. The analysis has been conducted using co-occurrence matrices (Martine & Rugg, 2005) and thematic analysis (Braun & Clarke, 2006). The card sorting activity was followed up with a laddering activity (Hinkle, 1965) for a subset of the participants. The results from this were reviewed to evaluate the potential of this technique for identifying the belief models that affect user opinions of web pages.

5.2 Background

This section discusses the background literature on card sorts and alternative categorisation techniques. It then addresses the use of laddering in conjunction with card sorting. Finally, this section provides details of the experimental design used for the study described in this chapter.

5.2.1 Card sorts

Card sorting is a technique in which participants are given a set of cards, and asked to sort them into categories. The categories can be predefined by the researcher, the technique known as a closed sort, or of the participant's own choosing, the technique known as an open sort (Rugg & McGeorge, 2005). The process can be repeated as many times as needed, until the participant has run out of categories, until the researcher feels they have hit diminishing returns or until the specified time is complete. Closed sorts can be used to determine the level of agreement between participants. Analysis of closed sorts is more time efficient than analysis of open sorts, however, the respondents are tied to the researcher's categorisations, which is not appropriate for all research. Card sorting can be completed in person or online (Righi *et al.*, 2013), and the relative merits of each option are discussed further in Section 5.2.3.2. Card sorting can be used to investigate the level of agreement between users on a particular topic (Rugg & McGeorge, 1997), making it an appropriate technique for the research described in this chapter. Card sorting is used for determining information positioning within the field of Information Architecture (Spencer, 2009). Other uses include identification of users' decision-making processes when browsing web pages (Kodagoda, 2010). Card sorting is originally a technique from psychology, with one of the first reported examples of its use being by Bergström (1983) to investigate the interference effect. It has previously been used to elicit user preferences of music (de Quincey, 2010) and learning preferences (Price, 2004).

One pattern of response to be aware of in relation to card sorts is a tendency for males to sort using two groups and females into more than two (Gerrard & Dickinson, 2005). Therefore, it is

sensible to record demographic information about gender and to analyse results to consider how many binary sorts have been performed by each set of participants.

Research covering shufflers vs spreaders is notably absent from the literature, although it is a concept understood by the card sorting community. Anecdotal evidence suggests the environment the sort is conducted in can make a difference to some individuals, especially if they are unable to use their preferred method for looking through all the cards. Some participants like to be able to place all the cards in view at the same time, whereas other prefer to hold them in their hands and "shuffle" through them. It can be off-putting to participants if their preferred option is not available to them (Rugg, 2013).

This section has provided an overview of card sorting, a technique identified as a suitable candidate for the research described in this chapter. A review of alternative categorisation techniques was undertaken, with each being assessed for suitability; this is summarised below.

5.2.1.1 Alternatives to card sorts

The purpose of including a sorting technique in the research described in this chapter is to elicit information regarding users' categorisations of web pages, their higher level value and beliefs that affect web page opinions, visual attributes that imply belonging to particular categories, and the level of agreement between placement in categories between users, and between user and computer placement into categories.

A number of alternative categorisation techniques exist, including Q-sorts, hierarchical sorts, and all-in-one sorts. Q-sorts, and all-in-one sorts are discussed below.

Q-sorts are a sorting technique based on methodology in The Study of Behaviour (Stephenson, 1953). Q-sorts involve participants comparing and ranking a set of information, known as the Q-sample, according to their own point of view. Participants are encouraged to follow a normal distribution when ranking the information. Q-sorting is often used to find out the different types

of beliefs participants have about the Q-sample. Q-sorts are of use for gathering ratings. In some cases information is gathered about the sorting process itself, this information is often less informative than from other sorting techniques (Dziopa & Ahern, 2011). Some examples of the use of Q-sorts include investigating trust in e-Government (Alsaghier *et al.*, 2009), investigating human perceptions within Human Computer Interaction (Meloche, 1999), and investigating student perceptions towards distance learning (Valenta *et al.*, 2001). Q-sort methodology has been widely adopted and adapted, but not all applications have been true to the original concept (Dziopa & Ahern, 2011), and the method has frequently been merged with quantitative methods such as R-Methodology. The assumption of a normal distribution is not appropriate for every piece of research (Rugg & McGeorge, 1997). One purpose of the sorting activity in the research described in this chapter is to identify the criteria used with no direction from the researcher. Q-sorts use pre-existing criteria, therefore Q-sorts would not be appropriate for the study described in this chapter.

In the field of Information Architecture sorting tasks are used for a number of design tasks including structuring the navigation or knowledge base of a website (Spencer, 2009). All-in-one sorts are often used for this task. An all-in-one sort entails asking respondents to sort all items into logical categories. This process can be conducted with a very large number of cards, with numbers ranging from 30 to 219 (Hannah, 2005). All-in-one sorts involve sorting the cards a single time, and each card can only belong to a single category. For Information Architecture, these are logical restrictions as each piece of content belongs in a single place; however, for the research described in this chapter it was desirable to encourage multiple sorts. This maximised the potential for recording all the sorts a participant could think of.

Card sorts were selected as the most appropriate technique to explore categorisation in the study described in this chapter. The use of this technique alone provides information of the categorisation model used by participants, but the use of card sorts alone does not provide an

opportunity to explore the responses given directly with the participant. Laddering is a technique often used in conjunction with card sorting to identify the underlying reasons for the categorisations chosen. Information about how important a particular criterion is to a respondent, and the reason behind the level of importance, was relevant to this study. It was not clear how evident this information would be from sorting alone, so a decision was made to use laddering in addition to card sorts with a subset of the participants. This enabled the researcher to determine whether subsequent studies should employ the techniques in tandem, or whether card sorting alone provided the information sought in this research.

The following section provides a review of the laddering technique and discusses the suitability for the research described in this chapter.

5.2.2 Laddering

Laddering is a specific form of questioning, in which the researcher continues asking questions concerning the preference and the reasons behind a given preference. It is a flexible technique that has been adopted within domains ranging from knowledge acquisition (Corbridge *et al.*, 1994), advertising (Reynolds & Gutman, 1988) to the psychology of constructs (Bannister & Fransella, 1986). The technique involves a recursive use of "why" questions, with an end goal of discovering the higher-order beliefs of the respondent (Hinkle, 1965; Neimeyer *et al.*, 2001). It can be applied to many different situations, from an individual's preference between two website designs to what their ideal job is. The idea is to ladder upwards to the core reason behind the previous response. The technique has its roots in personal construct theory (Kelly, 1955). Laddering is one way to discover the mental model of the world an individual uses. This information can be difficult to access through traditional methods including questionnaires or interviews, as the participant is not always able to verbalise their world view. The versatility of the method is considered both a strength and a weakness. The line of questioning is dependent on the previous answer, and on the researcher being able to ladder up the hierarchy on-the fly,

often while noting down the response from the previous question. There is also no way to know in advance the line of reasoning that may be discovered, making preparation in advance impractical. In some cases laddering can result in the disclosure of personal information, irrelevant to the focus of the research. A researcher using the technique needs to be aware of this risk, and know to stop laddering upwards at the first sign of a participant going off-topic.

The previous two sections evaluated the techniques used for the study described in this chapter. There are a number of other design decisions which need to be discussed, and this discussion is found in the following section.

5.2.3 Design

The research described in this chapter has been conducted using the same set of web pages as Study Two, described in Chapter 3. This decision enabled a comparison of findings, and conclusions to be drawn regarding the respective strengths and weaknesses of each technique used. Therefore, the domain used for the research described in this chapter was hot air balloon experiences. A full discussion of this decision can be found in Section 3.6.1.2.

5.2.3.2 Data collection method

Card sorting is a well-recognised technique used to elicit information regarding categorization and grouping (Rugg & McGeorge, 1997). The results can be analysed a variety of ways including to understand the categories themselves, the items that tend to be grouped together, and the naming of the categories. This technique is usually conducted in person, but software has also been created to be used for the task. The use of software has some advantages: it does not need a researcher to conduct each experiment, saving time and also meaning experiments can be run simultaneously. The time implication can make a large sample size more realistic in a short timescale. However, software does not always replicate the in-person card sorting experience (de Quincey, 2010), and limitations include the inability to ask questions to the researcher, potential for incorrect sorting and missing the opportunity to follow up on the sorts conducted.

The research described in this chapter used a sorting technique to elicit the categorisations individuals use when looking at web pages. Any software solution for this task needed to allow multiple sorts and enable users to view a full-sized screenshot of each web page. In order to complete a sort, participants needed to be able to view a number of cards at the same time. It was not possible to meet all these requirements using software, so the data collection was conducted in person. This also had the benefit that laddering could be conducted with participants.

The following section describes the card sorting and laddering activities conducted to identify visual attributes that contribute to subjective opinions of web pages.

5.3 Method

The participants, materials and procedure are distinct for the card sorting and laddering activities. Therefore, the method section has been organised to present the two stages of the research as separate activities for clarity.

5.3.1 Part A: Card sorting activity

Participants undertook repeated single-criterion sorts on picture cards of 30 web pages.

5.3.1.1 Participants

Participants were recruited via convenience sampling. An invitation to take part was circulated via email to students, society members, and staff, at Keele University. This set was chosen as it was a convenience sample, and groups of people were identified where a high response rate was predicted. Sorts from 18 participants were gathered. A screening question was used to ensure no participants could understand written Finnish.

Demographic	Categories	Count
Gender	Male	11
	Female	7
	Other / prefer not to say	0
Web design experience	Never designed a website	5
	Designed a few websites	12
	Designed many websites	1
Hot air balloon familiarity	Not at all familiar	12
	Slightly	6
	Somewhat	0
	Moderately	0
	Extremely	0

Table 30 Demographic information about participants

The card sorting activity took place before the online data collection described in Chapter 3 and there was no prior exposure to the sites.

5.3.1.2 Materials

The cards used for the sorting activity were A4 size and contained screen shots of a set of 30 hot air balloon web pages in Finnish. The pictures on the cards were web pages showing a partial screenshot (taken using a browser window of A4 size and without scrolling). The web pages were the same set that was used for the research described in Section 3.6. Information on how the web pages were selected and gathered can be found in Section 3.6.1.2. The images were printed on paper and laminated. Each card had an identification number in the top right hand corner. The identification labels were printed in black on a white background and were of a consistent size and in a consistent font. The identity numbers were assigned arbitrarily. The number of

cards is at the upper bound of that recommended for repeated single criterion sorts (Rugg & McGeorge, 1997), and was dictated by the number used for the correlation study described in Chapter 4, as the two studies used the same set of web pages.

The participants were provided with an information sheet, a consent form, a background questionnaire and an instruction sheet. These can all be found in Appendices X to X respectively. The instruction sheet has been modified from Rugg and McGeorge (2005). The sort information regarding criterion, categories, and assignment of cards to categories was recorded on paper by the researcher.

A pilot study was conducted with two participants. Further details of the pilot can be found in Section 5.3.3 below. A small number of changes were made after the pilot including the addition of two questions after every sort, these have also been discussed in Section 5.3.3. Five cards were replaced after the pilot for content-related reasons such as a high proportion of the content being in English. New screenshots were taken of all of the pages and the final set of cards was reviewed for content by one of the pilot participants.

5.3.1.3 Procedure

Participants were initially given an information sheet about the research and a consent form. Once the consent form had been completed, the instruction sheet and background questionnaire were provided. The four documents can be found in Appendices N to Q respectively. The background questionnaire was then reviewed by the researcher to check the participant was unable to understand written Finnish. The instruction sheet was then provided. The use of an instruction sheet ensured the instructions were consistent for all participants. To ensure each participant understood how to perform a sort a practice run was conducted using a set of cards depicting cars. When the participant has completed a practice sort they were asked if they understood the task. Once the participant confirmed they understood the task, the main sorting activity began.

Participants were asked to sort the cards into categories based on a single criterion. Once this task had been completed, the criterion, category names and cards belonging to each category were recorded on paper by the researcher. After each sort had been recorded participants were asked two questions linked to the sort:

Which category do you prefer and why?

How important is this criterion to you when looking at web pages?

The answers to these questions were recorded next to the sort information. The answers to these questions were followed up in the laddering section for the subset of participants also doing the laddering task.

Participants were not given any particular perspective or scenario to consider, giving them freedom to base their criteria on anything they could think of, without restrictions. In a number of cases, participants asked for clarification on whether the level of importance they placed on the criterion was when looking at hot air balloon web pages, or all web pages. The researcher provided a consistent response to this, which was "Tell me the importance in relation to both scenarios separately".

The recorded card numbers were counted after every sort to ensure no cards had been missed. Due to the size of the cards it was unlikely that they would be misplaced during the sort. If there had been a risk of this, the cards would also have been counted after every sort.

The researcher did not conduct any additional category discovery through dyadic or triadic elicitation as the focus was not to discover the maximum possible criteria, but to discover the criteria that respondents came up with unprompted. All sorts were open sorts as the activity was focussed on the individual's groupings, not which cards belonged to predefined groups.

The card sorting activity was conducted in the same location for every participant. This allowed control of environmental factors such as space available.

5.3.2 Part B: Laddering activity

A laddering task was performed with a subset of the card sorting participants. The task involved asking further questions relating to the sorts performed and the reasons behind the participant's views.

5.3.2.1 Participants

The laddering activity was conducted with 10 participants after they undertook the card sorting activity described in Section 5.3.1 above. Participants were acquired for the card sorting activity using the methods described in Section 5.3.1.1. The first 10 participants that were available for a maximum of 90 minutes, as opposed to 60, as required for the card sorting activity, were asked to undertake the laddering activity in addition to the card sorting activity.

5.3.2.2 Materials

Laddering was conducted based on the results provided during the card sorting task. The researcher used the laddering technique to follow up responses of interest. It was not possible to know what these would be in advance. To assist with the question creation, the researcher had access to a list of standard phrasings which can be found in Appendix T. The researcher chose appropriate, relevant phrasings from the list and incorporated the card, category or criterion details from the preceding sorts. The standard phrasings were modified, when required, to fit the situation. In some cases, cards were presented to the participant relating to the question posed. In any instance where participants were asked to further discuss cards that had been placed in a particular category, the first two cards written down in the category list were used.

5.3.2.3 Procedure

The researcher conducted laddering on concepts, criteria, and reasons for preferred categories based on the responses for each individual using the phrases (or modifications) as shown in

Appendix T. It was not possible to use a standard format as the laddering was tailored for each participant based on their sort results. By asking questions regarding preference, information about the participants' core values was elicited and laddering continued until these were found, or appropriate questions had been exhausted.

5.3.3 Pilot

Each stage of the research described above was piloted with two participants. After the pilot activities each participant was asked to reflect on task they had been asked to perform. This acted as a meta-analysis of the task and required the respondents to step out of the social experiment context and look at the task from a different context.

During the pilot some of the criteria and categories provided had clear positive and negative categories, but in other cases the researcher chose to confirm "which would be the most preferred category and why" during the laddering portion of the study. An example of this was a sort on the criterion of "navigational cues" with categories "navigational cues visible" and "navigational cues not visible". It was logical to infer that "navigational cues visible" would be preferred but it was not definitive from the wording used. In this instance the respondent seemed confused when the question was posed regarding which was their preferred category as they felt it was obvious. In order to reduce this confusion a consistent approach was adopted of asking the question for every criterion. Further laddering was conducted regarding the "why" response if the participant was in the group completing the laddering task. It was also observed in the pilot that when the question "Which category do you prefer and why?" was posed at the end of all the sorts there were some instances where a participant had difficulty remembering what they were thinking when they did the original sort. To combat this memory issue the question was posed directly after the sort with the second pilot participant. This proved to be a more practical timing of such questions, and ensured that preference information was gathered from the participant rather than guessed by the researcher.

During the pilot some sorting was performed based on criteria that the researcher believed to be based on observable visible attributes and it was not clear whether they were of importance to the participant when viewing web pages, examples of this were the colour of the page and whether it had a big picture or not. In the reflective discussion this belief was confirmed. For this reason an additional question was added after each sort asking how important each criterion was to the participant.

Seven web pages were changed after the pilot due to content. In two cases the original page included a video and there was no obvious frame to use for the screen capture. In three cases a video or image did not load correctly when the page was parsed via Google Translate. In two cases the screen shot was reviewed and the majority of the content of the image was in the original language, due to the use of images not text in the original site. Once new sites had been selected the new cards were reviewed by one of the pilot participants.

The images used during the pilot were taken using Chrome, and the browser window had custom colours. One of the pilot participants commented that they found the colour distracting when evaluating the sites. For the final set of images the default browser colours were used.

5.4 Analysis procedure for the card sorting activity

The following sections detail the analysis procedure, the use of independent judges and the coding of sort information that was undertaken.

5.4.1 Quantitative analysis

The number of sorts and the number of binary sorts performed by each participant was recorded. The participants were split into a number of subpopulations, based on their answers to the background questionnaire. Information was provided on the total number of sorts, total number of binary sorts, percentage of binary sorts, mean number of sorts and median number of sorts, for all participants and for each subpopulation.

5.4.2 Qualitative analysis

Exploratory analysis of the sort information was conducted using thematic analysis (Braun & Clarke, 2006). This technique was chosen in place of the more commonly used content analysis, as it offers a more flexible approach, whilst still providing a rich, detailed and complex account of the information. The research described in this chapter sought to identify emergent themes from the data set, and did not seek to report the full data set. Superordinate analysis was performed, taking into account both the superordinate groupings of criteria and gist analysis of categories. The superordinate analysis was informed by the category and criterion names identified through the thematic analysis. Two independent judges, psychology graduates with experience of thematic coding, were used during the qualitative analysis.

Thematic analysis (Braun & Clarke, 2006) using open coding was performed on the raw sort data and the themes, clusters, or units of analysis, were identified. This process was performed independently by the researcher and one of the independent judges. The two sets of identified units of analysis were reviewed and combined before the next stage of the process was conducted. The purpose of this analysis was to identify common or recurring concepts and sorts before performing superordinate and gist analysis.

The criteria and categories within each unit of analysis were reviewed and coded into superordinate groupings and gist categories by the other independent judge. The coding was reviewed by a member of the supervisory team for validity. The coding process at this stage was used to provide an overview of the entire data set for the units of analysis identified during the previous stage, therefore, inductive, as opposed to theoretical, thematic analysis was used. Independent researchers were used for both these stages to prevent the analysis being analyst-driven and influenced by the researcher's pre-conceived coding frame, and the literature on the wider thesis topic. In two instances, a particular sort was identified to belong to more than one theme, and a duplicate sort was added to the sort record with a different SortID, to allow the sort

to be in two superordinate groupings. The two additional sorts were Sort 103 and Sort 104. Once a number of superordinate groupings had been identified the resulting sort information, regarding cards and the categories to which they had been assigned, was inputted to a card sort analysis spread sheet (Spencer, 2009). The spread sheet provided information regarding the level of agreement on card placement within categories respondents. At all stages of the analysis process discussed above, where disagreements occurred, the researcher made the final decision based on information from all parties.

Item-based analysis was conducted, using co-occurrence matrices, to investigate the interaction between different criteria. This process required the creation of a single set of card allocation information for each criterion that had been used by more than one respondent. The single set of card allocation information is referred to as the "combined sort" information from this point onwards. To create the combined sort, the researcher reviewed the level of agreement spread sheet discussed in the previous paragraph. Each card was allocated to the group that had the highest level of agreement on the spread sheet. In any instance where the levels of agreement were the same, the card was labelled with a group name that showed it was unclassified and was also unique. The group name had to be unique, as two cards being unclassified did not mean they were similar. In order to achieve this, the group name was the sort name, the suffix NC and a number (for example: NavigationNC1). The numbering began at 1 for each sort criterion, and incremented each time a "not classified" group was required within that sort criterion. A co-occurrence matrix was then generated for each theme, using the combined sort information for any criterion that occurred more than once. Co-occurrence matrices were also generated for a number of combinations of themes. To generate a co-occurrence matrix, a spread sheet was used (Rice, 2012) which already had the formulae to produce the matrix.

Information regarding the importance of each criterion to the respondent was recorded verbatim. There was no guidance provided concerning the type of information the respondent could

provide. This was intentional, as it allowed the respondent to provide as much information as they wished, and they were not tied to a framework that they would not have considered themselves. For analysis purposes the responses were coded as of the highest importance or lowest importance to each respondent. In order to take inter-judge reliability into account, this process was done by the researcher and the two independent judges. The process was carried out independently, and the results were reviewed and combined to provide results for this concept. This allowed analysis of the timing of the most important and least important criteria being mentioned. The responses between respondents cannot be compared directly as they are from different points of reference.

5.5 Results

The results presented in this section show the quantitative analysis and the superordinate groupings of criteria used by more than one participant. Information about the analysis process used for the card sort information can be found in Section 5.4 above. A sample of card sort responses can be found in Appendix U. For practical reasons the full set of responses have not been included.

5.5.1 Quantitative analysis

The following section covers the results of the exploratory analysis of the card sorting activity. Card sorting results were obtained from 18 participants, resulting in a total of 102 sorts.

Demographic	Categories	Total number of sorts	Mean number of sorts	Median number of sorts	Total number of binary sorts	Mean number of binary sorts	Median number of binary sorts
Overall		102	5.67	5	49	2.72	3
Gender	Male	65	5.91	5	32	2.91	3
	Female	37	5.29	5	17	2.43	2
	Other / prefer not to say	0	0	0	0	0	0
Web design experience	Never designed a website	24	4.8	5	10	2	2
	Designed a few websites	68	5.67	5	33	2.75	3.5
	Designed many websites	10	10	10	6	6	6
Hot air balloon familiarity	Not at all familiar	69	5.75	5	35	2.92	3
	Slightly	33	5.5	5	14	2.33	2
	Somewhat	0	0	0	0	0	0
	Moderately	0	0	0	0	0	0
	Extremely	0	0	0	0	0	0

Table 31 Number of sort information broken down by demographic

The mean and median have both been included in Table 31 above as they each provide useful information in this context. The mean has been provided as it allows all results to be fully taken into account. In the majority of cases the mean will not produce an integer, and in the context of card sorts, information about the middle sort, when ordered, is more appropriate. By rounding the mean information may be lost. For this reason the median has also been included.

5.5.2 Superordinate groups

Criteria used by multiple respondents, after superordinate grouping.

Superordinate Group	Gist Category	Number of sorts
Amount of information in the page	Too much information	6
	Right / manageable level of information	
	Not enough information	
Background	Picture	3
	Solid colour	
	Gradient	
	Texture	
	Other	
	Solid colour	
Colour that stands out first	Blue	3
	White	
	Black	
	Other colours	
Contact details	Contact details present	5
	Contact details not present	
	Possibly	
Logo position	Easy to see	2
	Not easy to see / no logo	
Navigation location	Top	9
	Left	
	Top and left	
	Other	
Overall impression of the site	Good impression	12
	Average impression	
	Not a good impression	
Picture size / quantity	Lots of pictures	4
	Not many pictures	
	Big pictures	
	Small pictures	
	Medium pictures	
	Video	
Price information	Price information present	3
	Price information not present	
Professionalism of the page	Professional	3
	Not professional	
	Unsure	

Readability	No body text	3
	Poor	
	Ok	
Search tool	Search tool present	4
	Search tool not present	
Social media	Social media present	2
	Social media not present	
Unrelated marketing	Unrelated marketing not present	2
	Unrelated marketing present	
	Not sure	
Use of blue on the page	Very blue	3
	Some blue	
	Not very blue	
Use of colour on the page	Can't tell	3
	Not used well	
	Well used	
Would I use the site?	I would use the site	2
	I would not use the site	
	No idea (don't know)	

Table 32 Superordinate groupings of sorts

5.6 Discussion

5.6.1 Quantitative analysis

The total number of sorts performed by males was sixty-five, and thirty-two (49%) of these were binary sorts. The total number of sorts performed by females was thirty-seven, and seventeen (46%) of these were binary sorts. Card sorting literature has previously investigated differences in sorting practices of males compared to females, and found that males are more likely to perform binary sorts (Gerrard & Dickinson, 2005). In the study described in this chapter the proportion of binary sorts performed by males compared to female was higher, however only marginally so. It is possible that additional demographic factors, such as whether the respondent was an expert or novice web designer, also affected the likelihood of producing binary sorts, and this additional demographic information may explain the similar proportions of binary sorts observed between the genders.

The mean number of sorts by individuals with no web design experience was 4.8, for those with some experience it was 5.66, and for the individual with a lot of web design experience it was 10. This suggests that there are differences between the number of categorisations between people with no web design experience and those with some experience. The number of sorts performed by the respondent with lots of web design experience was ten, and is much higher than for the groups with less experience or no experience. When considering the median instead of the mean, the group with no experience and the group with some experience, produced the same median number of sorts, five. If the median had been the only average calculated the difference between the two groups of respondent would have been masked. It is therefore sensible when reporting average card sorting results for groups of respondents, to include both the mean and the median. The result of ten sorts for respondents with lots of web design experience should be treated with caution as there was only one respondent in this category. A larger number of respondents would be required to draw definitive conclusions on the differences between expert and novice web designers. The differences in number of sorts observed between respondents with no web design experience and some or lots of web design experience may be explained by domain specific knowledge. Website designers commonly evaluate sites and are more likely to have existing criteria for evaluation or review. Website designers are also used to making decisions on a number of aspects of design for a page, and this knowledge will have influenced their sorts.

The average number of sorts for those that were not at all familiar with hot air balloons was 5.75, and for those that were slightly familiar it was 5.5. The difference between these two categories represents a fairly small difference in level of knowledge, therefore, conclusions have not been drawn regarding difference in sorting practices between these two groups. The hot air balloon familiarity question was included to identify any experts in the domain. There were no experts within the participants.

Two participants, both male, performed one sort each where all the cards were placed in the same category. The criteria used were*: https?,* and *floating page or not floating page?*. In both cases the participant was part way through the sort before querying whether such sorts were allowed. The researcher encouraged them to complete the sort, as the criterion used was of relevance to the study.

5.6.2 Thematic groups

Thematic analysis was conducted on the card sorting criteria and category results in order to identify themes. The resulting information was used to generate co-occurrence matrices covering each theme. The purpose of the analysis was to identify frequently occurring themes and provide a more detailed account of these themes. Inductive analysis was used to code the data used in the co-occurrence matrices, and the analysis was data driven. The terminology used to reflect upon this information differs between thematic analysis, cluster analysis, and content analysis. The terminology from thematic analysis has been used in this thesis, with groups of information identified as themes.

One common theme identified within the sort results concerned whether functional aspects of the page were present or not. Examples of this theme were: contact details, price information, a search tool, and links to social media. In a large number of cases the presence of these functional aspects was considered a positive thing. The presence of unrelated marketing was also used for sorting and this can be considered the presence of an artefact or not, rather than a function. There were mixed responses to the presence of unrelated marketing, with one respondent viewing this positively, and two respondents viewing it negatively. To one respondent (P5) the inclusion of unrelated marketing implied other companies had a high enough opinion of the site to want their advertising on it. To another respondent (P11) the inclusion of external marketing often led, in their past experience, to a more cluttered page, and they preferred web pages to be less cluttered. Another respondent (P18) mentioned advertising when explaining how the site

design affected their decision to book a trip with the company or not, and they felt that a company without external advertising could be trusted more, as they were not reliant on additional funding through advertising.

A category for cards that did not fit the other categories was used eight times. This category was labelled “unknown”, “don’t know”, “can’t tell”, and “unsure”. In this report it has been referred to as “unknown”. This category was used by one participant four times. Cards that were included in the “unknown” category in any sort were: 1, 2, 5, 6, 7, 8, 13, 14, 15, 17, 19, 20, 25, 26, 27, 28, 30. Cards 2, 7, 14 and 28 were placed in the “unknown” category on two separate occasions. Unknown is a term that can be used in two types of situation, those where the cards falls outside of the range of convenience for the term (Kelly, 1955), and those where something about the card makes it difficult to place into a particular category. Further questions would be needed in order to determine which situation it was, and whether it is related to the image itself, or the categories the participant had chosen.

5.6.3 Thematic analysis of groupings elicited based on co-occurrence

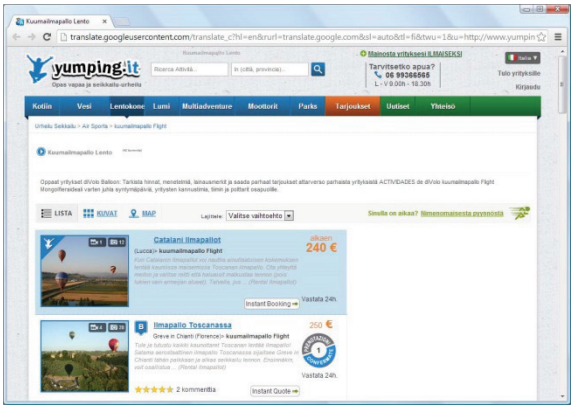
The analysis of co-occurrence matrices has not been covered in detail in the card sorting literature. One possible reason for the absence of literature on this topic is that the analysis varies depending on the purpose of the research. Previous examples of the use of co-occurrence matrices include Website Similarity (Martine & Rugg, 2005) and Folk Biological Classification (Molnar, 2012). Martine and Rugg (2005) used co-occurrence matrices to spot websites with a high level of similarity. The purpose of the research was to investigate websites, and the implications of system design and usability. The analysis focussed on identifying pages with a high level of similarity. The co-occurrence matrices generated allowed the researchers to determine how similar websites were, which could then be used to show whether sites produced from a set of guidelines appear to conform to a house style. The matrices could also be used to spot potential plagiarism between websites, but the researchers cautioned against the use of this

technique alone as a basis for legal action. Molnar (2012) used picture sorting to look at classifications for habitats and to consider the differences of classification between groups of respondents. The images used belonged to particular taxonomies, and a question they sought to answer was how well the existing taxonomies were understood and used by different groups of respondents. The co-occurrence matrices generated showed similar classifications between the various types of experts used, with laymen creating less structured groupings.

In the study described in this chapter, co-occurrence matrices were generated for the following themes: colour, emotion producing, balloon image, website design and functions. A matrix was also generated for all the aforementioned themes combined. In each matrix, cells were highlighted according to the co-occurrence percentage. Dark blue represented 75-100%, medium blue represented 50-74%, light blue represented 25-49% and white represented 0-24%. The full size matrices can be found in Appendix W. In the paragraphs below cards with a high number of instances of high or low co-occurrence have been selected. The number of instances has been shown in brackets after the words *high number*. The number considered to be high has been chosen in the context of the co-occurrence results for each theme individually.

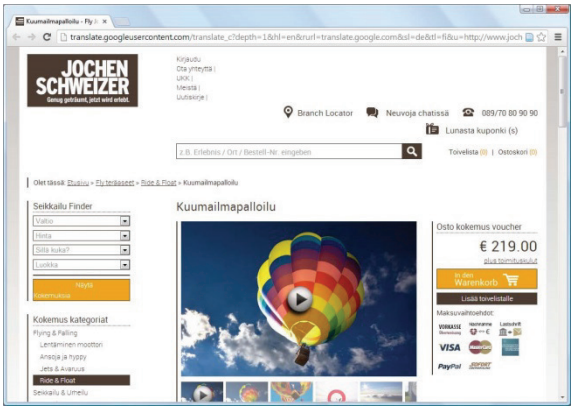
5.6.3.1 Colour

The co-occurrence matrix on the theme of *colour* was generated from three sorts. All of the three sorts had been standardised though the process described in Section 5.4.2. The raw data from nine sorts was used to create the three standardised sorts.



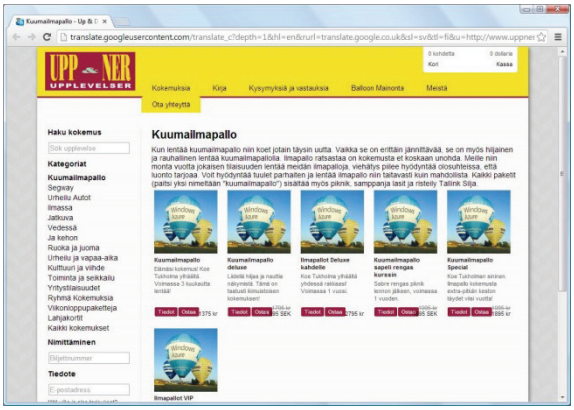
Card 8

Card 11

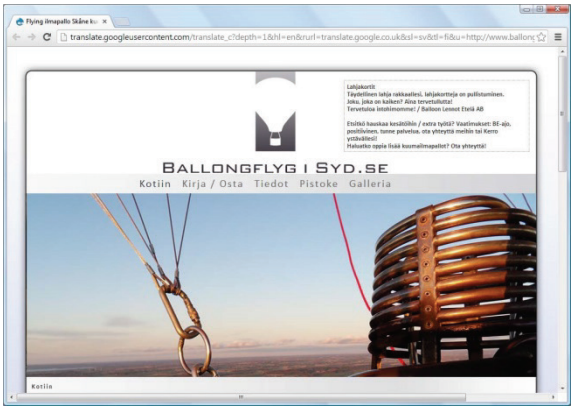


Card 12

Card 25



Card 26

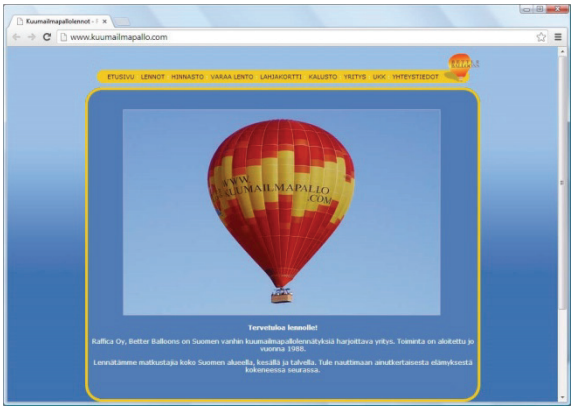


Card 28

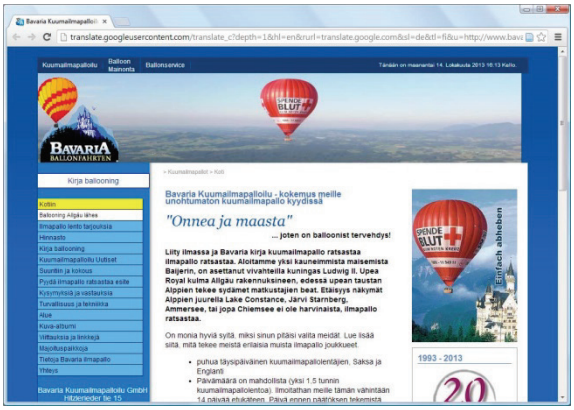
Figure 31 Cards with low co-occurrence for theme of colour

The cards shown in Figure 31 had a high number, between fifteen and twenty-four, of co-occurrence percentages between 0 and 24 for the theme of *colour*. This means that these cards were not usually placed in the same category as other cards for any of the three sorts covered by this theme. The sort names were: *use of colour, use of blue*, and *colour that stands out first*.

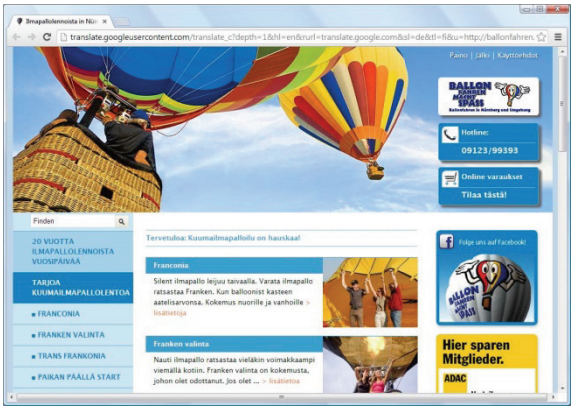
By reflecting on the standardised sort data for this theme, it can be seen that the majority of the cards that exhibited a large number of low co-occurrence percentages were *not categorised* for the sort based on how well colour was used within the page. Non-categorisation occurred when there was no majority category shown in the correlation spread sheet, after standardisation of sort categories had occurred. The only exception to this was Card 8, which was placed in the *not used well* category.



Card 1



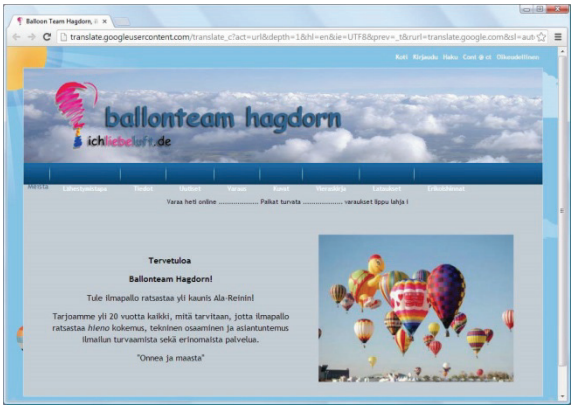
Card 13



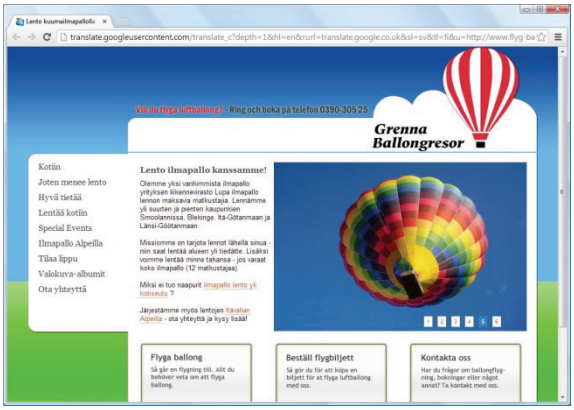
Card 14

Card 15

Card 20



Card 24



Card 27

Figure 32 Cards with high co-occurrence for theme of colour

The cards shown in Figure 32 had a high number (seven) of co-occurrence percentages between 75 and 100 for the theme of *colour*. This means that these cards were often in the same category as each other for all three of the sorts covered by this theme. The sort names were: *use of colour*, *use of blue*, and *colour that stands out first*. It is logical that a card would appear in similar categories when considering the colour that stood out first to the respondent and the level of blue within the page as the two criteria are interlinked. The more interesting finding is that cards were often in the same category in relation to how well they had used colour within the page. This similarity between category placement of a subjective and objective criterion is useful as it allows us to infer what may cause a respondent to determine how well colour has been used. A positive response to this criterion may also lead to a more positive perception of the site overall. The findings need to be compared with the subjective ratings gathered in the research described

in Chapter 3 to investigate this possibility further. The cards identified look at first glance as though they would belong to the *blue* category (when considering *colour that stood out first*). A link between this and *use of colour – well used* is worthy of further exploration, through alternative analysis of the results from this study and in future studies.

All the cards that exhibited a high level of co-occurrence belonged to the following categories within the three criteria used with the theme of colour; *use of blue – very blue, use of colour – well used, colour that stands out first – blue.* The first and last of these categories are interlinked, and this may have affected the results. It is interesting to note that all these cards were also felt to have used colour well. It is not the case that all *very blue* cards also used colour well, but the majority did.

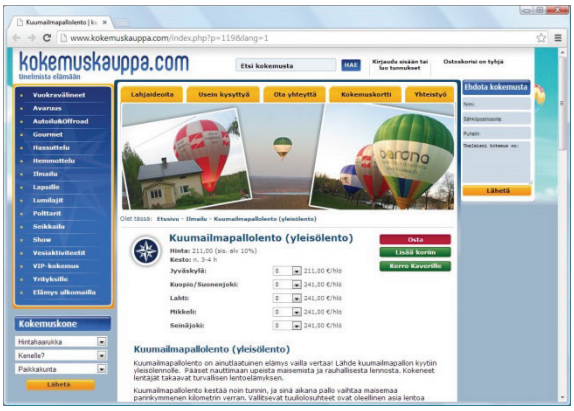
It should be noted that when discussing card category placement in relation to how well the page used colour, the standardisation process resulted in twenty-four cards in the *well used* category, five cards *not categorised,* and only one card in the *not used well* category. The small number of cards in the *not used well* category means any conclusions drawn regarding other colour aspects and how well the colour has been used should be treated with caution. Further research focusing on this theme only, through closed sorts and/or think aloud tasks, would provide a more conclusive sample size.

Card 11 stands out as an anomaly in the colour co-occurrence matrix. Card 11 had 24 instances of a low co-occurrence percentage, the highest count across all the co-occurrence matrices generated. By reflecting on the standardised sort data for this card, it was not categorised for two of the three sorts and given a unique group in each case. Card 11 was not categorised for *use of colour*, a sort which represented how well colour had been used within the page. The underlying features that caused these cards to be non-categorised have not been addressed directly in this study as it was an artefact of the standardisation process, not the verbatim sort results. This card was also not categorised for *colour that stands out first.* It would be interesting to use this card as

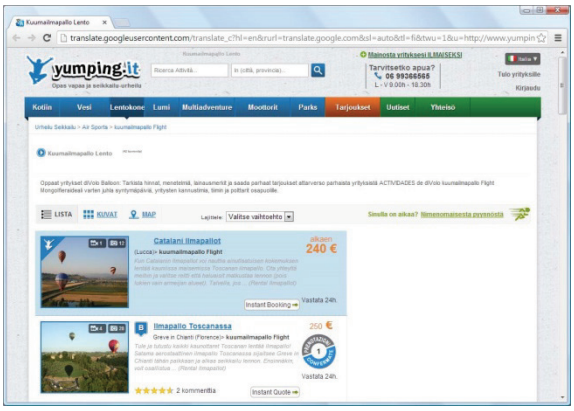
a focus for a think aloud task with a small number of participants to explore this card further. The use of Framing for the task would ensure responses related to the intended theme.

5.6.3.2 Website design

The co-occurrence matrix on the theme of *website design* was generated from fifteen sorts. Six original sorts were used as well as nine sorts that had been standardised though the process described in Section 5.4.2. The raw data from thirty-five sorts was used to create the nine standardised sorts.



Card 6



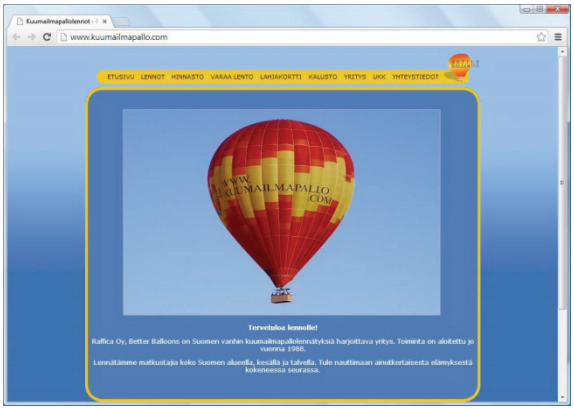
Card 8

Card 25

Figure 33 Cards with low co-occurrence for theme of website design

The cards shown in Figure 33 had a high number, between six and seven, of co-occurrence percentages between 0 and 24 for the theme *website design*. This means that these cards were not usually placed in the same category as other cards for any of the fifteen sorts covered by this

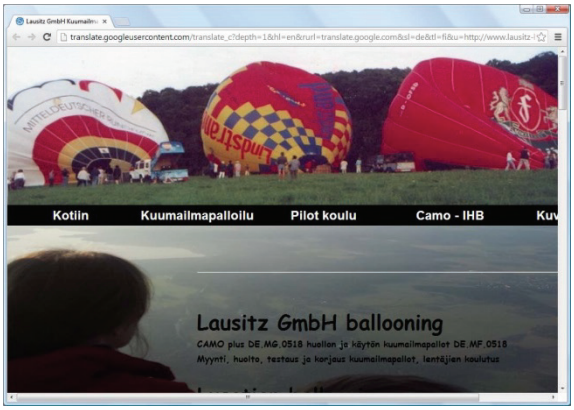
theme. The sort names for the standardised sorts were: *amount of information, readability, professionalism of the page, contact details, navigation location, price information, search tool* and *social media*. The three cards shown in Figure 33 were all allocated to the n*ot categorised* category for at least one sort. As this co-occurrence matrix was generated from fifteen sorts, a relatively large number, the significance of the *not categorised* allocation is smaller than in some of the matrices generated from smaller number of sorts.



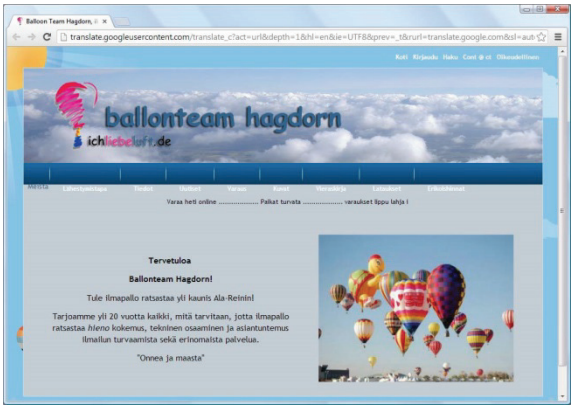
Card 1

Card 4

Card 20



Card 21



Card 24

Figure 34 Cards with high co-occurrence for theme of website design

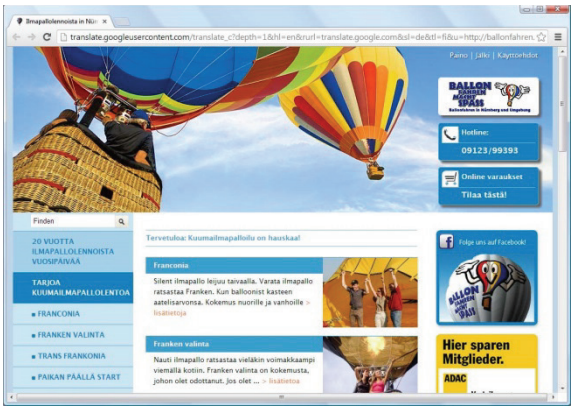
The cards shown in Figure 34 above had a high number, between four and five, of co-occurrence percentages between 75 and 100 for the theme *website design*. This means that these cards were often in the same category as each other for all three of the sorts covered by this theme. The sort names for the standardised sorts were: *amount of information, readability, professionalism of the page, contact details, navigation location, price information, search tool* and *social media*. Similarly to the sorts within the theme of colour, website design covers a number of subjective and objective sort criteria. By referring to the summary sort results for this theme it can be seen that some of the commonly occurring category allocations for the cards shown in Figure 34 were: *would I use the site – I would not use the site, readability – OK, focus of the page – information focussed, professionalism of the page – not professional, ease with which to look at and navigate – simple, image to text ratio – more text than image,* and *contact details, search tool, price information ,and social media – all not present.* One key observation from this list is that the majority of the cards in Figure 34 were missing all the functional aspects of a site identified by respondents. The cards in Figure 34 were sites that the respondents said they would not use, and felt were not professional; both negative subjective opinions. The high level of co-occurrence between cards placed in the negative subjective categories, and the categories reflecting an absence of functionality, implies the functions are used as quality indicators. Another interesting implied link from this co-occurrence matrix is between the identified focus of

the page towards either selling or information and the subjective criteria. There is a trend for pages focussed on information to be perceived negatively when considering subjective criteria.

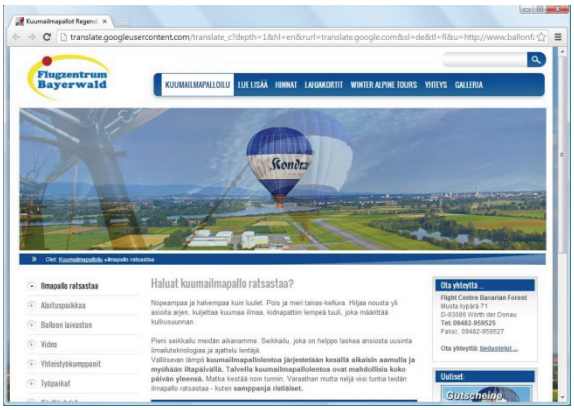
Interestingly, two of the cards in the above table were commented on by a number of participants as their least favourite, or the cards they felt had the worst design.

5.6.3.3 Emotion producing

The co-occurrence matrix on the theme of *emotion producing* was generated from two sorts. One original sort was used as well as one sort that had been standardised though the process described in Section 5.4.2. The raw data from twelve sorts was used to create the standardised sort.



Card 14



Card 16

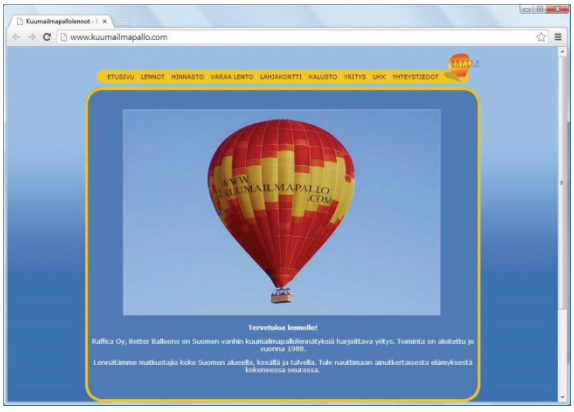
Card 17

Card 20

Figure 35 Cards with low co-occurrence for theme of emotion producing

The cards shown in Figure 35 had a high number, between eleven and twenty-two, of co-occurrence percentages between 0 and 24 for the theme *emotion producing*. This means that these cards were not usually placed in the same category as other cards for any of the two sorts covered by this theme. The sort names were: *overall impression* and *impression of site*.

Card 16 and Card 17 had very high counts of low co-occurrence. By reflecting on the standardised summary of the sorts used for this matrix it shows Card 16 was *not categorised* for *overall impression*, and Card 17 was the only card with the combination of *overall impression – good impression* and *impression of site – mixture.* The original sort results had a much wider selection of responses, and these responses were standardised into a single sort. From Appendix X the original responses can be seen, and the concepts covered within sorts on the topic of overall impression included aesthetic impression, personal preferences towards or against a page, and whether the page was appropriate for the topic. The verbatim sort responses provide much richer data than the standardised sort in this instance, and there are a number of possible avenues for a narrower follow up study from this information.



Card 1

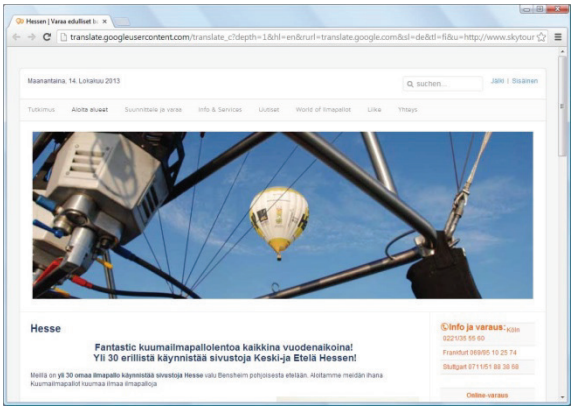
Card 2

Card 4

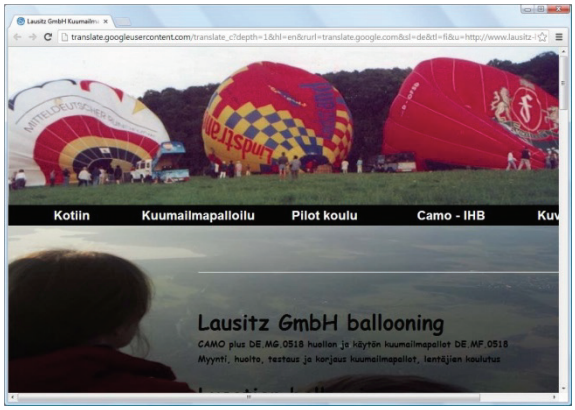
Card 5

Card 15

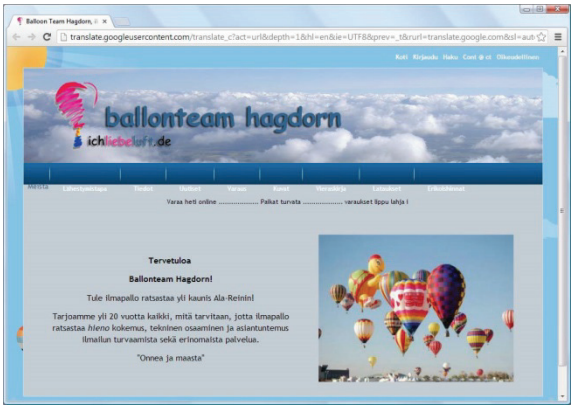
Card 18



Card 19

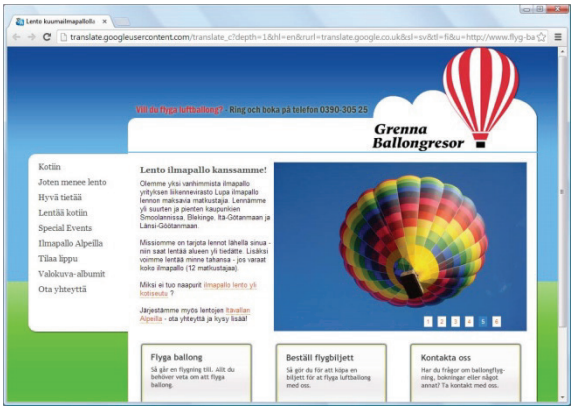


Card 21



Card 24

Card 25



Card 27

Card 30

Figure 36 Cards with high co-occurrence for theme of emotion producing

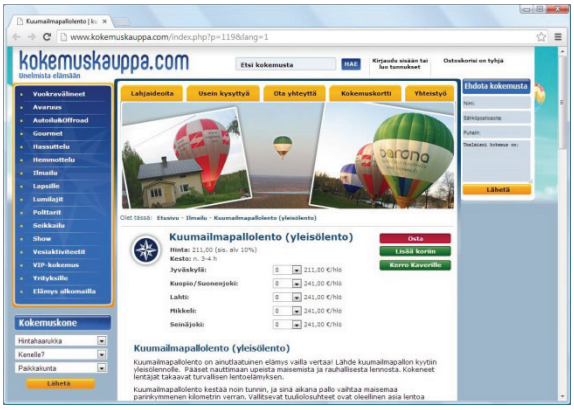
The cards shown in Figure 36 had a high number (twelve) of co-occurrence percentages between 75 and 100 for the theme *emotion producing*. This means that these cards were often placed in the same category as other cards for any of the two sorts covered by this theme. The sort names were: *overall impression* and *impression of site*. In this instance the criteria and categories the cards shown in Figure 36 were allocated to were o*verall impression – not a good impression,* and *impression of site – don't feel commercial at all.*

Information on co-occurrence for the theme of *emotion producing* should be treated with caution as the matrix was generated from two sorts. This small sample size is a threat to validity. The emotion produced by the image was not the only focus of this research, and this topic was not explored any further. It would be appropriate to carry out closed sorts to find out more about this topic.

5.6.3.4 Balloon image

The co-occurrence matrix on the theme of *balloon image* was generated from four sorts. Two original sorts were used as well as two sorts that had been standardised though the process described in Section 5.4.2. The raw data from sixteen sorts was used to create the standardised sorts.

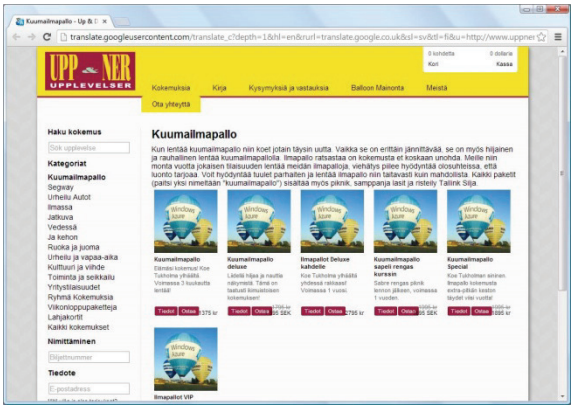
Card 2



Card 6

Card 17

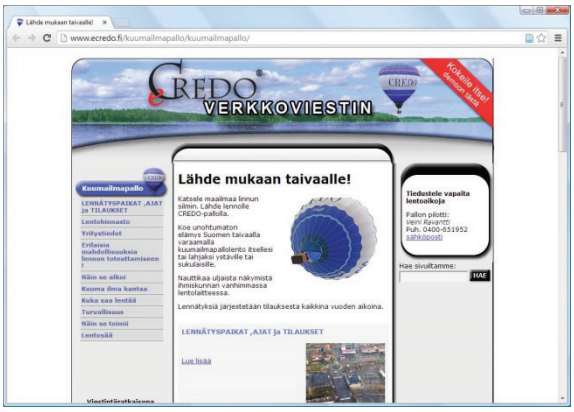
Card 20



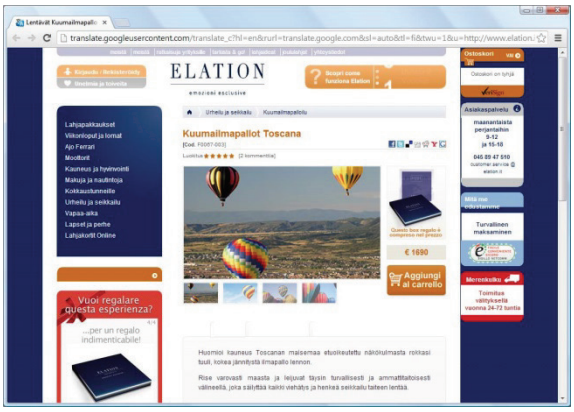
Card 26

Figure 37 Cards with low co-occurrence for theme of balloon image

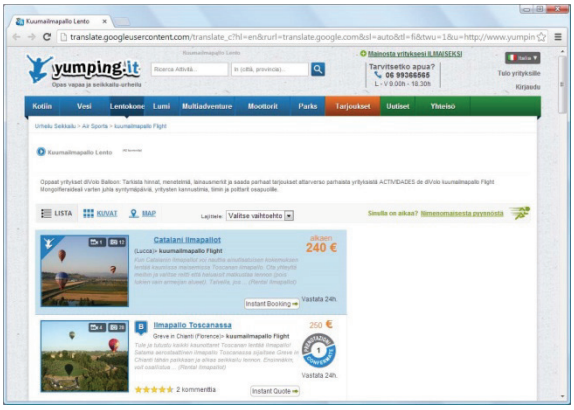
The cards shown in Figure 37 had a high number, between four and eight, of co-occurrence percentages between 0 and 24 for the theme *balloon image*. This means that these cards were not usually placed in the same category as other cards for any of the three sorts covered by this theme. The sort names were: *picture size / quantity, balloon location, content in main section* and *overall impression of the site*. By referring to the summary sort information for the five cards shown in Figure 37, it can be seen that three of the cards were not categorised for at least one of the four sorts, and one other card was placed in the n*ot sure* category for one sort.



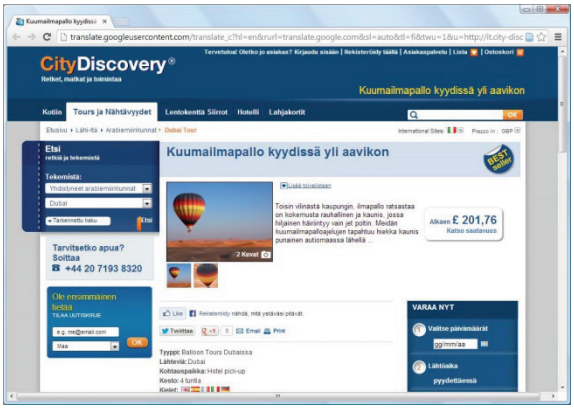
Card 4



Card 7

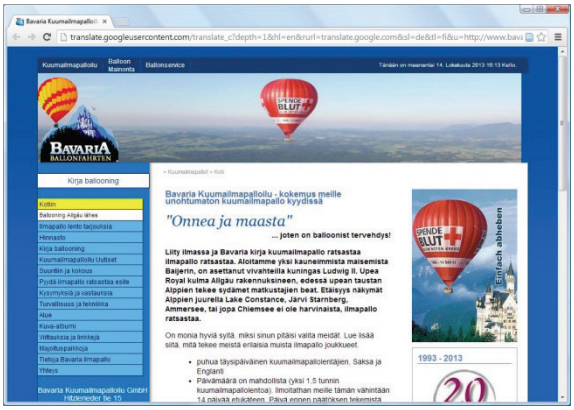


Card 8



Card 9

Card 11



Card 13

Card 25

Card 30

Figure 38 Cards with high co-occurrence for theme of balloon image

The cards shown in Figure 38 had a high number (twelve) of co-occurrence percentages between 75 and 100 for the theme *balloon image*. This means that these cards were often placed in the same category as other cards for any of the four sorts covered by this theme. The sort names were: *picture size / quantity, balloon location, content in main section*, and *overall impression of*

the site. For the eight cards shown in Figure 38 the most common combination of sort allocations was: *overall impression of the site – not a good impression, balloon location – balloons mostly flying, picture size / quantity – lots of pictures* and *content in main section – text and image.* The combination of allocations shown suggests a link between overall impression and how much is on the page, with lots of pictures and content on the main section being both text and image. The trend suggests that pages with a lot of aspects present were perceived negatively. This concept has been investigated in the objective and subjective measurement section of this chapter, with the premise that a complex page will have a low subjective rating.

5.6.3.5 Summary of thematic groupings

The card sorting activity described in this chapter was scoped to use open sorts as an exploratory technique. The co-occurrence matrices described in this chapter were created based on both original sort results, and the results following a standardisation process. When investigating the interaction between different criteria within the same theme, it was sensible to use standardised sorts for the generation of the matrices in order to prevent a higher weighting being given to sorts performed on similar categories. If all the sorts had been included as individual sorts an unintentional high level of co-occurrence would be present for any criteria used by more than one respondent. The use of standardised sorts controlled this.

After standardising similar sorts into a single standardised sort, the co-occurrence matrices were constructed, each with relatively small sample sizes. To combat this, either verbatim sorts or closed sorts performed using categories identified through a pilot sorting activity could be considered. The techniques used show promise for identifying trends between subjective opinions of sites, such as overall impression or how well they used colour, and objective measures, such as the colours used, or level of complexity of the page. The analysis techniques discussed would be more powerful if used to analyse the results of closed sorts based on these subjective and objective criteria. Another observation regarding the use of standardised sorts

was that a result of *Not categorised* after the correlation analysis could have a large affect on a co-occurrence matrix. The use of closed sorts would allow the inclusion of a single category covering: *don't know, not sure or not applicable*. In the standardisation process it was not appropriate to allocate *not categorised* cards to a single category, as it was not possible, after the fact, to be sure why a card was not categorised. A failure to agree between respondents is not the same as a respondent not knowing how to allocate a card, so unique categories had to be used for every instance where a majority category was not found in the correlation analysis.

5.6.4 Preferred category

This section provides a detailed account of a number of themes identified within the preferred category information gathered after each sort. Themes were open coded by the researcher and one of the independent judges.

The information provided by respondents in relation to "Which category do you prefer and why?" covered a wide range of concepts, and provided an unexpected insight into the thought processes involved.

In participants' justification for their preferred category, reference was made to the opinion of the company the participant had formed based on the criteria. Examples of the values participants discussed included professionalism, trust, confidence, and safety. In each case an aspect of the web page led the participant to form an opinion of the company. Some interesting links were provided between web page content and a positive opinion of the page. Examples include: images of balloons in the air were preferred as this gave the participant confidence in the company; the use of a white background to the text content of the page instead of a coloured background led the participant to believe the company was more professional; being able to identify the brand or company easily made one respondent trust the company more; and having a site that looked professional led to one respondent feeling safer contemplating a transaction being conducted through the site. These visual cues are very useful for designers to be aware of

as many of them are fairly simple, and low cost to implement, and could give a company the edge over their competitors.

Another theme that was identified in the category preference information concerned ease of use. Participants preferred web pages that looked easy to use, or had expected functionality. There was also a trend towards web pages that required less effort both to complete tasks and just to look at the page.

There are a small number of examples of the phrasings used being negatively framed, for example: “Don’t like multiple layout systems”, “Don’t have to go searching for the price”, “If there isn’t enough information I will leave the site and look elsewhere”, “Colours not too strong, pictures aren’t dominating…”, “Less of a waste of time…”, and “Don’t want to have to adjust window size to be able to read font”. The question posed to the respondents was positively framed, and in some cases the negatively framed response may have been to counter this. The use of a two part question may mean the framing has been forgotten by the time they answer the second part “and why?”. It may also be that the participant perceives their preferred category as the least objectionable option.

A small number of respondents justified their preferred category through the level of enjoyment they anticipated the activity to deliver, based on the web page. Comments included: “Want to have a great time if you booked a trip”, “Can identify more with going on the activity”, and “Like seeing the ground underneath – makes it seem more exciting”. This suggests users intricately link web page design and the end product a company will deliver. This has serious implications for companies, as the importance placed on the web page, and the emotions they evoke, are difficult to tailor to every user. This links back to the literature on nomothetic and idiographic preferences, and whether you can truly tailor something to appeal to a particular demographic. In situations where a company supplies products with varying purposes, creating a design that evokes the right emotions for all products is a non-trivial task.

5.6.5 Informal observations

Participants often volunteered much more detail than expected. One possible explanation for this was that during the sorting process there isn't much opportunity to explain or justify. The clarification questions at the end allow for this information to be recorded. Some participants felt the need to add qualifying phrases after their statements. It was not clear if this was to justify to themselves or the researcher.

One participant appeared to reflect on the logicalness of their comment in relation to how good the web page design was, and a well-designed site made them trust the company more. They then spontaneously commented that they were aware that the people taking the balloons up in the air were unlikely to be the same people that designed the website, so any implied trust due to the website design was illogical, but it was why they preferred a well-designed web page. This was an example of a participant acknowledging the decisions they make are not always based on sound reasoning.

In two cases participants opted to use the floor for the sorting task as opposed to the desk. In both cases they began by spreading all the cards out on the floor so they could see every card at once. One of the participants did this before a number of their other sorts, and when they were displaying their categories they tended to position the cards so about a quarter of each card was visible once they had been allocated to categories. The other participant only spread the cards out for the first sort, and for their subsequent sorts they would have had enough space available on the desk, but remained on the floor. This participant came up with a large number of categories in some of their sorts, and in order to adequately separate the piles of cards the additional space was useful. The background colour of the desk and the floor were both brown so this difference in sorting location was unlikely to have introduced any colour interference effects.

5.6.6 Clarification questions regarding preference and importance

The use of two questions to clarify the participant's preferred category and the importance of each criterion provided useful insights into the participant's thought processes. This was due to their tendency to add clarification information, rather than answer the question outright. Without these two questions it is conceivable the reason behind the positive association between hot air balloons and being in the sky may not have been discovered with such a clear line of reasoning.

5.6.7 Order of criteria

In order to investigate whether users sort by most or least important criteria first, an analysis of the order of sort importance was undertaken. The researcher and both of the independent judges reviewed the information regarding the level of importance of each criterion given. The results were grouped by respondent. The judges were asked to label the highest and lowest criterion given by each respondent. Multiple criteria could be labelled as highest or lowest, and the results for a respondent could be left blank if it was unclear which was the most or least important. Three separate individuals were used due to the subjective nature of the coding task. A graphical representation of the coding can be found in Appendix Y.

The researcher marked sorts for 13 respondents for both highest and lowest criteria. Independent judge 1 marked all 18 respondents, and Independent judge 2 marked 17 of the 18 respondents. A graphical representation of this information is shown in Figure 39 below.

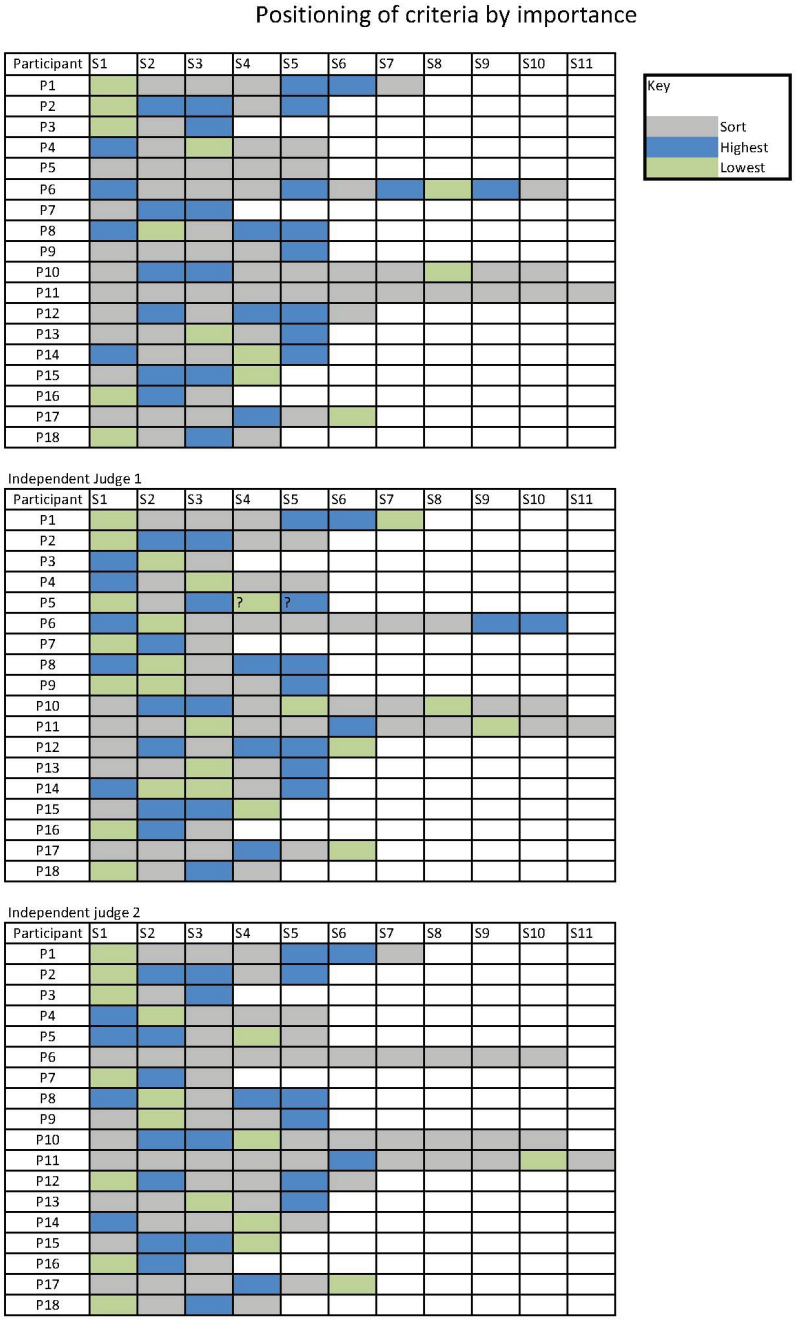


Figure 39 – Positioning of criteria by importance

The results do not show any clear pattern in the location of the high- or low-importance criteria. The results show it is marginally more likely that the criterion of lowest importance would be sorted on in the first sort by a respondent. One possible explanation for the lowest importance

criteria being sorted on first is that respondents are sorting on trivial but noticeably visible aspects of the page. Something that catches your eye immediately when presented with a set of images of web pages may not be something that would be considered when browsing sites in everyday life.

5.6.8 Laddering

One of the most insightful findings from the laddering was a domain specific justification for colour preference. A participant sorted based on how much blue was in the page. In their explanation for their preferred category, blue and white, they stated they preferred this category because they felt it portrayed the sky the best, which suited the topic best. They then went on to say it was nice that the blue content on the page was broken up with another appropriate colour. The researcher followed this statement up in the laddering by asking why they had a preference for blue on the page. They stated blue was the most appropriate colour given the topic, balloons fly in the sky, and the sky is blue. The researcher followed this up asking why balloons being in the sky was preferred to them being on the ground. The participant replied that it was quite important that a website for a trip in a hot air balloon give the impression that the balloons will go in the air. They followed this statement up with an interesting caveat. They stated they did not really know much about ballooning, but their naïve appreciation of the activity was that balloon being in the sky was a good thing. They were aware this was not a particularly technical assessment of the situation, but for an unfamiliar domain it was enough for them to state it as a visual cue for a successful company. This insight is not linked to the participant's core values or beliefs, however this insight provided useful domain specific information that could easily be incorporated into websites for the domain. It demonstrates the importance of asking the user the reasons behind their responses.

One interesting finding from the laddering portion of the study was the importance of efficiency to participants. This core value was revealed during laddering with two different participants. .

In the first case, the participant had included a number of sorts related to the functionality of the web page. They stated that looking for functionality had been a good indicator of how easily they would be able to complete tasks in the past. When asked why a more usable website was important, they stated it would take less time to complete a task. Obviously to this particular participant the time a task takes to complete is very important. They related this to wanting to be able to complete tasks quickly. Their line of reasoning also covered the level of frustration they would feel if a task was taking a long time to complete. The third participant used the criteria of whether the price was visible or not, and when explaining why they preferred the cards with a visible price they volunteered information about a preference to be able to find the information without having to go hunting for it. This led to a laddering question about why time was important to them and they stated they want to use their time to do other things than searching a website for information. They felt that time spent hunting for information they would expect to be able to locate easily on a website was a waste of their time.

Functionality was a common trend within the sorts. In one case a participant appeared to view any issues being able to understand or follow the functionality as a deficiency on their part, rather than down to the website they were using. When asked why they preferred sites with good functionality, they stated that it was important a site was easy to use and had straightforward labelling. When asked why an easy to use site was important to them, they stated it gave them more confidence they would be able to do the task they wanted on the site. When asked why it was important to them to be able to complete the task they wanted on the site, they responded that it was important to be able to browse a site without having to worry about how to do things within the site. This was a diversion from the anticipated response regarding efficiency and the importance of their time, as discussed above. At this point it seemed prudent to stop questioning as a further questioning on their level of confidence would not have been appropriate and was unlikely to elicit additional relevant information. What was particularly striking about the response given by this participant was that their tone and description of the issues suggested they

believed it was their fault if they could not work out how to complete a given task on a website. Most of the other participants that covered ease of task completion referred to a lack of functionality, or inability to locate information, as a problem with the design, not themselves as the user.

5.6.9 Comparison of elicitation techniques

During the card sorting activity a number of participants revealed information about their most and least preferred of the cards. When the experiment was set up a decision was made to only record information about the completed sorts, card placements, and answers to the two additional questions. Therefore, information about most and least preferred cards was not gathered consistently from all participants. Although this information wasn't recorded, similarities between the cards stated and the results from the subjective ratings were identified. A systematic comparison of the results from Chapter 3, and the results from a number of the subjective superordinate groupings was conducted. By comparing the cards with high subjective ratings with the categorisations for *overall impression* it can be seen that four of the eight cards were placed in the *good impression* category. Two of the cards were *not categorised* and two of the cards were in the *not a good impression* category. For results regarding positive impression the results from the two techniques were not consistent. However, when comparing the cards with low subjective ratings it can be seen that all eight cards from the subjective ratings were placed in the *not a good impression* category. This shows both techniques obtained some consistent information. Another comparison was performed between the subjective ratings and the categorisations for *professionalism*. All eight cards with high subjective ratings were placed in the *professional* category. Five of the cards with low subjective ratings were placed in the *not professional* category with the remaining three *not categorised*. The fact that participants made a judgement on the professionalism of the pages, without any understanding of the content, is interesting as it suggests a high level of attention should be paid to the visual appeal of a web page.

Overall, this study shows that the aspects of the pages which contribute to a low rating can be identified using both methods. This finding adds strength to the argument that web page preference is not entirely individual, and there are some underlying processes driving the decision making in this context.

5.7 Conclusions

This chapter has described two interlinked activities to identify visual attributes of web pages and their importance to users. The web pages used were consistent with those from Study Two and this allowed for a comparison of results.

The key conclusions from the research described in this chapter were:

- Impressions can be formed based on visual cues alone
- Both card sorting activities and surveys can be used to identify impressions of web pages
- Aspects of web pages that create negative impressions can be identified more reliably than aspects that create a positive impression
- A sorting activity combined with laddering is a suitable technique to elicit domain specific visual cues in web pages
- Users revealed indicators of distrust more frequently than indicators of trust

The implications of these are discussed below.

The research presented in this chapter identified common themes within the categorisations used including website design, functionality, colour, and emotion. Aspects of design and functionality feature heavily within web design and usability literature, however, the impact of colour and the importance of creating the right emotional response are less prevalent.

Trends were identified between respondents regarding preferences for clear and obvious functionality within a web page. The large number of sorts performed relating to emotional reactions to the web pages showed that visual aspects of the page clearly have an impact on

users. Participants were also able to make judgements on whether they would pursue a site without being able to understand any of the textual content.

This research demonstrates it is possible to identify features that create a negative impression of a page from visual aspects alone. This is significant because negative impressions can be formed before a user has had the opportunity to read any of the textual content on the page. This shows that poor design can overcome good written content.

In this study responses from participants about elements or features that invoked a negative reaction were more consistent across individuals than responses about features that provoked positive reactions. Therefore, it is more useful to web designers to understand the elements of a web page that invoke a negative impression in several respondents rather than those that create a positive impression in a few respondents.

The study described in this chapter revealed domain specific preferences relating to the colour of the web site. It is important for a developer to be aware of any such domain specific associations. The techniques used in this chapter enabled such domain specific features to be determined, through the use of card sorts and laddering to identify preferred categories and the reasons behind these preferences.

In the laddering portion of the study, the number of mentions of indicators of distrust exceeded the number of mentions of indicators of trust. Designers need to be aware that indicators of distrust are considered relevant to users, identify what they are and seek to minimise user's exposure to them.

The data collection in this study gathered a large amount of information with a vast number of analysis options available, therefore, it was not feasible to analyse every component in-depth. This chapter has presented the most relevant and interesting results, however, it also raises a number of new questions and opens up many avenues for further research.

The findings regarding criteria importance, and the stage at which the highest or lowest criteria are sorted on found no patterns or trends of note.

5.7.1 Limitations and further work

The sample size used for the card sorting activity, although acceptable for this data collection technique, was still relatively small. Some of the results discussed are very specific to the individuals participating, and the impact of individual unique preferences should be considered. All the participants were university educated. Although this had not been the target population, it occurred due to the sampling method used. The findings should be considered with these limitations in mind.

The card sorting activity described in this chapter was exploratory in nature. In order to address some of the individual findings of the research, further studies could be conducted, each with a narrower focus. Two such areas of exploration are the impact of colour, and the perceived level of web design of a web page.

One area which warrants further investigation is the level of agreement between individuals on card placement in categories. A study using closed sorts would be appropriate for a study into the level of agreement between individual while co-occurrence matrices would be appropriate for analysing the results of the closed sorts. From the set of criteria found in the study described in this chapter, key sorts can be identified where further investigation would provide useful insights into the categorisation process. The set of criteria would depend on the purpose of the investigation.

The theme of colour was identified during the sorting activity described in this chapter. The literature on colour psychology suggests that colour impacts visual processing in a number of interesting and unexpected ways. Therefore, this is an area to consider in more detail.

The findings from the sorting activity described in this chapter suggest that the topic or context of the site can have implicit colour associations. In this case it appears blue was viewed positively as at least one respondent stated that in their mind they linked it to balloons being in the sky. This suggests that, for any domain, it is worth investigating whether there are any colour links that would subconsciously send out a positive or negative message to users. Suitable elicitation techniques to gain this insight include the think aloud technique, and card sorting followed by laddering.

A worthy avenue for a future study is the impact of colour, in isolation. This could be conducted through closed sorts on the three criteria associated with colour; *use of colour*, *use of blue* and *colour that stands out first*. All three of these criteria rely on respondents input to allocate cards into each category, as they have a subjective element to the card placement. The most subjective of the criteria in relation to colour is *use of colour*, and in the research described earlier in this thesis the card allocations for this criterion were done through standardisation and responses coded to conform to the standardised categories. This process was conducted thoroughly and verified, but any standardisation process has an element of interpretation which cannot completely replicate the thought processes of the participants. Therefore, asking respondents directly, using the exact categories verbatim, removes this interpretation. For the study described here, the purpose was to explore respondent perceptions and groupings, and for this reason no closed sorts were conducted. The use of open sorts led to a large number of criteria being discovered as the number of respondents using the same criteria as another respondent was low. The use of closed sorts on *use of colour* would provide validation of the findings regarding this topic, and would allow responses to be gathered from a larger sample than was used in the study described in this chapter.

There is scope for additional thematic analysis of the card sorting and laddering results, as many interesting questions arose from the analysis conducted using this technique. As part of the

laddering activity, justifications were gathered regarding card placement in categories, where the researcher felt further insight would be useful. Through a systematic analysis of this information additional visual attributes may present themselves.

The similarity between the results concerning high and low subjective opinions between the two techniques is a sensible area for further work.

The use of closed sorts would allow further investigation into the level of agreement across participants on placement of cards into categories. Therefore, this is an interesting field of research which requires further exploration into the subjective and objective measures found here.

Chapter 6 Conclusion

The research presented in this thesis was concerned with subjective opinions of web pages and the ability to predict these opinions without the need for user interaction. This was investigated through three different studies.

6.1 Key findings and implications

6.1.1 Impressions are formed quickly and can be based on visual attributes alone

The literature regarding the time taken to make decisions about web pages suggests that decisions are made before a user has had the opportunity to read the textual content of a page.

In Chapter 5 it was shown that users could react emotionally to pages where no readable content was present, meaning visual cues alone can lead to such reactions. Furthermore, Chapter 5 illustrated that a higher level of consistency was demonstrated across users in factors that created a negative reaction than in factors that created a positive reaction. Therefore, it is more effective to address negative reaction provoking factors than to introduce positive ones.

An implication of this is that in terms of retaining users, identifying and minimising negative reaction provoking factors should be the primary focus, with improving positive reaction provoking factors being treated as a secondary concern.

The findings concerning the importance of very early impressions of web pages were consistent in each study, and with Lindgaard *et al*. (2006), Kim and Fesenmaier (2008), Papachristos and Avouris (2011), and can be generalised.

6.1.2 Proxies for reactions to visual attributes

Visual attributes lead to impressions being formed. The findings showed that high-rated pages tended to have more colours. High-rated pages also had a high level of information complexity,

therefore, only a low level of compression. These trends match those found by Reinecke *et al.* (2013).

This similarity in results implies that there are underlying trends that can be measured objectively. Although the research presented has not determined if the set of measures used is the optimum set of objective measures to use for prediction of subjective opinions, the research did uncover a number of trends, which should be further explored in the future.

6.1.3 Encourage is not polar opposite of discourage, in the context of web design

In Chapter 3 encouragement and discouragement were found to be two distinct concepts and therefore, increasing encouragement is not the same as decreasing discouragement.

An implication of this is that designers and researchers need to decide which factor they are interested in.

The pattern between responses for encouragement and discouragement warrants further research to untangle the underlying thought processes. The ability to review a website for the level of discouragement is a useful approach for web design.

The findings concerning encourage and discourage were consistent for both sets of web sites, and do not appear to be specific to the domains used, therefore, these findings can be generalised.

6.1.4 The colour blue has positive associations in the context of hot air balloons

In Chapter 5 one participant offered an interesting justification for their colour preferences when looking at the cards. They explained that they preferred the cards with blue present, as hot air balloons are supposed to fly in the sky, so the colour blue is a good sign. For this participant, the topic of the web site led to a specific colour preference. When comparing this finding with the correlation study in Chapter 4, it can be seen that the blue standard deviation and blue mode were positively correlated with the subjective ratings six times.

The implication of this is that it is essential to understand any domain specific context surrounding the focus or purpose of the web page, and whether this leads to associations with particular visual cues. In the example used for this study, the domain of hot air balloons linked to the colour blue in users' minds and web pages using the colour blue had an implied positive association with the domain. Domain specific contexts and features can also be a self-perpetuating cycle; as more web pages adopt the feature people expect to see it in web pages for the domain, and the presence of the features reinforces the positive association.

6.1.5 Similarity between card sort and questionnaire results

The two techniques used, card sorts and an online survey, provided consistent results regarding impressions of web pages. This is worth being aware of as depending on the situation, one technique may be more appropriate, for example, surveys can be conducted with large numbers of people relatively quickly, but are unlikely to provide information about why a particular opinion has been formed. A card sorting activity provides rich information about the topic, but is likely to cost more in terms of time and analysis. If the goal of the evaluation is to determine overall impressions of the pages, either technique could be used.

The similarity between findings from questionnaire and card sorts was really interesting, and points to some underlying phenomenon. The discovery of similar information from two very different techniques, lends external validity to the findings.

6.2 Recommendations

The key findings and implications have lead to two sets of recommendations, one for web site developers, and one for researchers.

6.2.1 Recommendations for website developers

Previous research has shown that impressions of web pages are difficult to change once formed. Therefore, it is important to ensure that the first impression is favourable.

Visual aspects of web pages are processed and evaluated before the text content. Therefore, it is important to understand which visual aspects of a page lead to a positive or a negative impression.

A higher level of agreement between participants was observed regarding negative impressions of web pages, than positive impressions. Therefore, further work is required to understand which aspects contribute to this, and developers should be aware of this work.

For web sites where users will be required to enter personal details or payment details, trust plays an important role. Users can determine a level of trust in a web site from visual cues alone. Therefore, It is important to be aware of research into this topic, and to incorporate elements that provoke trust, and reduce or remove elements that provoke distrust, within web sites.

Encouragement and discouragement are not polar opposites of a single scale. By trying to encourage users to explore the rest of your site, this will not automatically reduce aspects that discourage users. Therefore, these two factors need to be addressed separately. Know what your focus is: is it to minimise discouragement or maximise encouragement. It is also important to be aware of research into this topic.

Colour associations were identified to relate to the domain focus of the web sites. Therefore, it is important to understand whether any context specific preferences exist, and to incorporate them where possible.

6.2.2 Recommendations for researchers

When asking questions about levels of encouragement or discouragement, it is important to realise these are two distinct concepts. Therefore, these two concepts should not be placed as opposite ends of a Likert or Likert-style scale. It is possible that different aspects of a web site contribute to each impression, and there is a need to ask additional questions to isolate the aspects of a page contributing to each impression.

Information regarding users' impressions of websites was consistent between the two techniques used, card sorts and questionnaires. Therefore, either technique could be used, depending on other factors such as cost or time available.

Card sorting allows you to identify categorisations used by users, and when coupled with laddering the reasons behind these groupings can be gathered. Therefore, the use of card sorting and laddering is a suitable combination of techniques for identifying visual attributes of web pages that contribute to users' impressions.

This research identified a number of areas where better understanding would be beneficial to web site developers. These included visual attributes that lead to very early negative impressions, and visual attributes that provoke trust and distrust.

Existing web site research frequently uses a set of web sites chosen by the researcher, and does not always take into consideration whether a user would have chosen to remain on a site if they did not have to. This applies to research into first impressions as well as those including task completion. This is an area of research that deserves further investigation.

6.3 Further work

This research was exploratory in nature, and it identified a number of interesting avenues for further research. A prime example would be a focussed study regarding user decisions to stay or leave a page. Closed card sorts would be a very suitable mechanism for exploring this, particularly with the use of framing, to ensure participants placed themselves in the role of a user, as opposed to a web developer.

One main limitation of the work was due to the choice of images for Study One, any future comparison studies should use consistent domains, and ensure any sample better reflects a sample of web pages. To enable a solid comparison between designs the study design needs to be consistent, and for practical reasons, and sue to the different research questions posed in each

study, this research did not have an adequate level of consistency for generalisations across domains.

The use of card sorting and laddering opened up a number of interesting research avenues. Studies of this type are more appropriate for exploratory studies, similar to the one described in this thesis, unless the target population is well defined.

6.3 Summary and conclusions

The research set out with the aim of demonstrating that it was possible to use software tools to objectively predict user reactions to web page designs. The measurements used were chosen as practical initial metrics, based on the literature, that were readily available using existing software tools. These initial measurements have demonstrated a number of associations between the objective metrics and user reactions. This shows that the underlying concept behind the research is valid and that a software tool could be constructed to automatically evaluate web pages based on objective intrinsic factors.

References

ALLEN, I.E. and SEAMAN, C.A., 2007. Likert scales and data analyses. *Quality Progress,* **40**(7), pp. 64-65.

ALSAGHIER, H., FORD, M., NGUYEN, A. and HEXEL, R., 2009. Conceptualising citizen's trust in e-government: Application of Q methodology. *Electronic Journal of E-government,* **7**(4), pp. 295-310.

ALTUN, A. and CAKAN, M., 2006. Undergraduate students' academic achievement, field dependent/independent cognitive styles and attitude toward computers. *Journal of Educational Technology and Society,* **9**(1), pp. 289.

ANDERSON, C.A. and DILL, K.E., 2000. Video games and aggressive thoughts, feelings, and behavior in the laboratory and in life. *Journal of Personality and Social Psychology,* **78**(4), pp. 772.

ANDRADE, E.B. and ARIELY, D., 2009. The enduring impact of transient emotions on decision making. *Organizational behavior and human decision processes,* **109**(1), pp. 1-8.

ARANYI, G., VAN SCHAIK, P. and BARKER, P., 2012. Using think-aloud and psychometrics to explore users' experience with a news Web site. *Interacting with Computers,* **24**(2), pp. 69-77.

ASLAM, M.M., 2006. Are you selling the right colour? A cross-cultural review of colour as a marketing cue. *Journal of Marketing Communications,* **12**(1), pp. 15-30.

BABIN, B.J., HARDESTY, D.M. and SUTER, T.A., 2003. Color and shopping intentions: The intervening effect of price fairness and perceived affect. *Journal of Business Research,* **56**(7), pp. 541-551.

BANNISTER, D. and FRANSELLA, F., 1986. Inquiring Man: The Psychology of Personal Constructs (3rd Edn) Croom Helm.

BATTLESON, B., BOOTH, A. and WEINTROP, J., 2001. Usability testing of an academic library web site: a case study. *The Journal of Academic Librarianship,* **27**(3), pp. 188-198.

BEM, S.L., 1974. The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology,* **42**(2), pp. 155.

BILAL, D., 2000. Children's use of the Yahooligans! Web search engine: I. Cognitive, physical, and affective behaviors on fact-based search tasks. *Journal of the American Society for information Science,* **51**(7), pp. 646-665.

BORGES, J. and LEVENE, M., 2006. Ranking pages by topology and popularity within web sites. *World Wide Web,* **9**(3), pp. 301-316.

BOTTOMLEY, P.A. and DOYLE, J.R., 2006. The interactive effects of colors and products on perceptions of brand logo appropriateness. *Marketing Theory,* **6**(1), pp. 63-83.

BRAUN, V. and CLARKE, V., 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology,* **3**(2), pp. 77-101.

BREWSTER, S., MCADAM, C., MCDONALD, J. and MACIVER, J., 2012. Rethinking camera user interfaces, *IS&T/SPIE Electronic Imaging* 2012, International Society for Optics and Photonics, pp. 82990K-82990K-9.

BRIGGS, P., 2002-last update. *Turning into digital goldfish*. Available: http://news.bbc.co.uk/1/hi/1834682.stm [July, 5th, 2004].

CALITZ, A. and SCHEEPERS, B., 2002. A comparative analysis of e-commerce website development using two implementation methods, *Proceedings of the 2002 annual research conference of the South African institute of computer scientists and information technologists on Enablement through technology* 2002, South African Institute for Computer Scientists and Information Technologists, pp. 251-251.

CALLAHAN, E., 2005. Cultural similarities and differences in the design of university web sites. *Journal of Computer-Mediated Communication,* **11**(1), pp. 239-273.

CHEN, A.C., GERMAN, C. and W ZAIDEL, D., 1997. Brain asymmetry and facial attractiveness: Facial beauty is not simply in the eye of the beholder. *Neuropsychologia,* **35**(4), pp. 471-476.

CHEN, S.C. and DHILLON, G.S., 2003. Interpreting dimensions of consumer trust in e-commerce. *Information Technology and Management,* **4**(2-3), pp. 303-318.

COE, R. and FITZ-GIBBON, C.T., 1998. School effectiveness research: Criticisms and recommendations. *Oxford Review of Education,* **24**(4), pp. 421-438.

COOKE, L., 2010. Assessing concurrent think-aloud protocol as a usability test method: A technical communication approach. *IEEE Transactions on Professional Communication,* **53**(3), pp. 202-215.

CORBRIDGE, C., RUGG, G., MAJOR, N., SHADBOLT, N. and BURTON, A., 1994. Laddering: Technique and tool use in knowledge acquisition. *Knowledge Acquisition,* **6**(3), pp. 315-341.

CROWLEY, A.E., 1993. The two-dimensional impact of color on shopping. *Marketing Letters,* **4**(1), pp. 59-69.

CYR, D., BONANNI, C., BOWES, J. and ILSEVER, J., 2005. Beyond trust: Web site design preferences across cultures. *Journal of Global Information Management,* **13**(4), pp. 25-54.

DE QUINCEY, E., 2010. *Software support for comparison of media across domains.* Keele University.

DZIOPA, F. and AHERN, K., 2011. A systematic literature review of the applications of Q-technique and its methodology. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences,* **7**(2), pp. 39-55.

ERICSSON, K.A. and SIMON, H.A., 1985. *Protocol analysis.* MIT press.

EVANS, J.S.B., 2008. Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology,* **59**, pp. 255-278.

GELO, O., BRAAKMANN, D. and BENETKA, G., 2009. Quantitative and qualitative research: Beyond the debate. *Integrative Psychological & Behavioral Science,* **42**(3), pp. 266-290.

GERRARD, S. and DICKINSON, J., 2005. Women's working wardrobes: A study using card sorts. *Expert Systems,* **22**(3), pp. 108-114.

GOODMAN, L.A., 1961. Snowball sampling. *The Annals of Mathematical Statistics,* **32**(1), pp. 148-170.

HANNAH, S., 2005. *Sorting out card sorting: Comparing methods for information architects, usability specialists, and other practitioners*, University of Oregon.

HAUGEN, E., 1950. The analysis of linguistic borrowing. *Language,* **26**(2), pp. 210-231.

HERMEKING, M., 2005. Culture and internet consumption: Contributions from cross-cultural marketing and advertising research. *Journal of Computer-Mediated Communication,* **11**(1), pp. 192-216.

HINKLE, D.N., 1965. *The change of personal constructs from the viewpoint of a theory of construct implications*, Ohio State University Columbus.

HOLLAND, J. and MENZEL BAKER, S., 2001. Customer participation in creating site brand loyalty. *Journal of Interactive Marketing,* **15**(4), pp. 34-45.

HOLLAND, N., 2004. *An investigation of correlations between subjective and objective criteria concerning websites.* MSc edn. Keele University.

HOPKINS, W.G., 2013-last update, *A new view of statistics*, [Online]. Accessed from http://www.sportsci.org/resource/stats/ [July, 7, 2013].

HUDSON, W., 2013. *Card sorting*. In: SOEGAARD, M. and DAM, R.F. (Eds.) (2013). *The Encyclopedia of Human-Computer Interaction (2nd Ed.)*. [Online]

HURD, A., 2001. *Using card sorts to elicit cross-cultural perceptions of web page quality: A study of students of English.* MSc edn. Keele University.

KAHNEMAN, D., 1973. *Attention and effort.* Englewood Cliffs: Prentice Hall.

KAHNEMAN, D., 2011. *Thinking fast and slow.* St. Ives: Penguin books.

KELLY, G.A., 1955. *The psychology of personal constructs. Vol. 2.* New York: W.W. Norton and Company.

KHOUW, N., 2002. The meaning of color for gender. *Color Matters*. [Online]

KIM, H. and FESENMAIER, D.R., 2008. Persuasive design of destination web sites: An analysis of first impression. *Journal of Travel Research,* **47**(1), pp. 3-13.

KNOTT, D.R., 2003. *Fibonacci numbers and the Golden Section*. [Online]

KODAGODA, N., WONG, B. and KHAN, N., 2010. Open-card sort to explain why low-literate users abandon their web searches early: Proceedings of the 24th BCS Interaction Specialist Group Conference 2010, *British Computer Society*, pp. 433-442.

KORIAT, A., LICHTENSTEIN, S. and FISCHHOFF, B., 1980. Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory,* **6**(2), pp. 107.

KREITLER, H. and KREITLER, S., 1972. *Psychology of the arts.* Duke University Press.

LABRECQUE, L.I. and MILNE, G.R., 2012. Exciting red and competent blue: The importance of color in marketing. *Journal of the Academy of Marketing Science,* **40**(5), pp. 711-727.

LASER, J., 2012. A Link to the Past: An Examination of JS Bach and György Kurtág's compositions using the golden ratio and Fibonacci Sequence. *2012 NCUR*.

LAVIE, T. and TRACTINSKY, N., 2004. Assessing dimensions of perceived visual aesthetics of web sites. *International journal of human-computer studies,* **60**(3), pp. 269-298.

LINACRE, J., 1998. Visual analogue scales. *Rasch Measurements Transactions,* **12**, pp. 639.

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), pp. 115-126.

LIVIO, M., 2002-last update,=. *The golden ratio and aesthetics.* Accessed from: http://plus.maths.org/content/golden-ratio-and-aesthetics [July 17, 2006].

MACKIEWICZ, J., 2006. Audience perceptions of fonts in projected PowerPoint text slides, *International Professional Communication Conference*, *IEEE*, pp. 68-76.

MADDEN, T.J., HEWETT, K. and ROTH, M.S., 2000. Managing images in different cultures: a cross-national study of color meanings and preferences. *Journal of international marketing,* **8**(4), pp. 90-107.

MARTINE, G. and RUGG, G., 2005. That site looks 88.46% familiar: quantifying similarity of Web page design. *Expert Systems,* **22**(3), pp. 115-120.

MCCULLOCH, W.S. and PITTS, W., 1943. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics,* **5**(4), pp. 115-133.

MELOCHE, J.A., 1999. Q Methodology as a research methodology for human computer interaction, *Interfaces for the Global Community, '***99**, pp. 149-152

MICHAILIDOU, E., HARPER, S. and BECHHOFER, S., 2008. Visual complexity and aesthetic perception of web pages, *Proceedings of the 26th annual ACM international conference on Design of communication* 2008, ACM, pp. 215-224.

MITRA, A., WILLYARD, J., PLATT, C.A. and PARSONS, M., 2005. Exploring web usage and selection criteria among male and female students, *Journal of Computer-Mediated Communication,* **10**(3), pp. 00-00.

MOLNÁR, Z., 2012. Classification of pasture habitats by Hungarian herders in a steppe landscape (Hungary), *Journal of Ethnobiology and Ethnomedicine,* **8**(1), pp. 28 - 46.

NEIMEYER, A., ANDERSON, A. and LESLIE STOCKTON, R., 2001. Snakes versus ladders: a validation of laddering technique as a measure of hierarchical structure. *Journal of Constructivist Psychology,* **14**(2), pp. 85-105.

NGO, D.C.L. and CH'NG, E., 2001. Screen design: Composing with dynamic symmetry, *Displays,* **22**(4), pp. 115-124.

NIELSEN, J., 1995-last update. *10 usability heuristics for user interface design: Homepage of Nielsen Norman Group,* [Online]. Accessed from: http://www.nngroup.com/articles/ten-usability-heuristics/ [July 27, 2006].

NIELSEN, J., 1999. User interface directions for the web. *Communications of the ACM,* **42**(1), pp. 65-72.

NISBETT, R.E. and WILSON, T.D., 1977. The halo effect: Evidence for unconscious alteration of judgments. *Journal of personality and social psychology,* **35**(4), pp. 250.

NORMAN, D.A., 2002. *The design of everyday things.* New York: Basic books.

O'CASS, A. and FENECH, T., 2003. Web retailing adoption: Exploring the nature of internet users web retailing behaviour. *Journal of Retailing and Consumer services,* **10**(2), pp. 81-94.

PAPACHRISTOS, E. and AVOURIS, N., 2011. Are first impressions about websites only related to visual appeal? *Human-Computer Interaction–INTERACT 2011.* Springer, pp. 489-496.

PAYNE, J.W., 1976. Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance,* **16**(2), pp. 366-387.

PRICE, L., 2004. Individual differences in learning: Cognitive control, cognitive style, and learning style. *Educational Psychology,* **24**(5), pp. 681-698.

PUTZ, J.F., 1995. The golden section and the piano sonatas of Mozart. *Mathematics Magazine*, pp. 275-282.

REINECKE, K., YEH, T., MIRATRIX, L., MARDIKO, R., ZHAO, Y., LIU, J. and GAJOS, K.Z., 2013. Predicting users' first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* 2013, ACM, pp. 2049-2058.

REINHART, A.L., HARING, S.H., LEVIN, J.R., PATALL, E.A. and ROBINSON, D.H., 2013. Models of not-so-good behavior: Yet another way to squeeze causality and recommendations for practice out of correlational data. *Journal of Educational Psychology,* **105**(1), pp. 241.

REYNOLDS, T.J. and GUTMAN, J., 1988. Laddering theory, method, analysis, and interpretation. *Journal of Advertising Research,* **28**(1), pp. 11-31.

RICE, M., 2012-last update. *Co-occurrence matrix announcement: Homepage of informoire*, [Online]. Accessed from http://www.informoire.com/taxonomy/co-occurrence-matrix-announcement/ [October 28, 2013].

RIGHI, C., JAMES, J., BEASLEY, M., DAY, D.L., FOX, J.E., GIEBER, J., HOWE, C. and RUBY, L., 2013. Card sort analysis best practices. *Journal of Usability Studies,* **8**(3), pp. 69-89.

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

RODRIGUEZ, P., TAN, S. and GKANTSIDIS, C., 2006. On the feasibility of commercial, legal P2P content distribution. *ACM SIGCOMM Computer Communication Review,* **36**(1), pp. 75-78.

ROSEN, D.E. and PURINTON, E., 2004. Website design: Viewing the web as a cognitive landscape. *Journal of Business Research,* **57**(7), pp. 787-794.

RUGG, G., 2013. *Blind spot: Why we fail to see the solution right in front of us.* USA: HarperCollins.

RUGG, G. and MCGEORGE, P., 1997. The sorting techniques: A tutorial paper on card sorts, picture sorts and item sorts. *Expert Systems,* **14**(2), pp. 80-93.

SAYOOD, K., 2012. *Introduction to data compression.* Newnes.

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

SHANNON, C.E. and WEAVER, W., 1949. *The mathematical theory of communication*. Urbana: University of Illinois Press.

SHEA, D., 2006-last update. *CSS zen garden*. Accessed from http://www.csszengarden.com/ [December 3, 2006].

SHEA, D. and HOLZSCHLAG, M.E., 2005. *The zen of CSS Design: Visual enlightenment for the web.* USA: Pearson Education.

SHNEIDERMAN, B. and PLAISANT, C., 2010. *Designing the user interface (5th Ed.).* USA: Pearson Education.

SPENCER, D., 2009. *Card sorting: Designing usable categories.* New York: Rosenfeld Media.

STANOVICH, K.E. and WEST, R.F., 2000. Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences,* **23**(5), pp. 645-665.

STANOVICH, K.E. and WEST, R.F., 2008. On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology,* **94**(4), pp. 672.

STATCOUNTER, 1999. *Visitor analysis for your website*. Accessed from http://gs.statcounter.com [September 5, 2013].

STEPHENSON, W., 1953. *The study of behavior; Q-technique and its methodology*. Chicago: University of Chicago Press

TOMBROS, A., RUTHVEN, I. and JOSE, J.M., 2005. How users assess web pages for information seeking. *Journal of the American Society for Information Science and Technology,* **56**(4), pp. 327-344.

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), pp. 1071-1083.

TRACTINSKY, N., KATZ, A. and IKAR, D., 2000. What is beautiful is usable. *Interacting with Computers,* **13**(2), pp. 127-145.

TUCH, A.N., BARGAS-AVILA, J.A. and OPWIS, K., 2010. Symmetry and aesthetics in website design: It's a man's business. *Computers in Human Behavior,* **26**(6), pp. 1831-1837.

TUCH, A.N., BARGAS-AVILA, J.A., OPWIS, K. and WILHELM, F.H., 2009. Visual complexity of websites: Effects on users' experience, physiology, performance, and memory. *International Journal of Human-Computer Studies,* **67**(9), pp. 703-715.

UPCHURCH, L., RUGG, G. and KITCHENHAM, B., 2001. Using card sorts to elicit web page quality attributes. *IEEE Software,* **18**(4), pp. 84-89.

VALENTA, A., THERRIAULT, D., DIETER, M. and MRTEK, R., 2001. Identifying student attitudes and learning styles in distance education. *Journal of Asynchronous Learning Networks,* **5**(2), pp. 111-127.

VAN DEN HAAK, MAAIKE J, DE JONG, M.D. and SCHELLENS, P.J., 2009. Evaluating municipal websites: A methodological comparison of three think-aloud variants. *Government Information Quarterly,* **26**(1), pp. 193-202.

VAN SCHAIK, P. and LING, J., 2003. The effects of screen ratio and order on information retrieval in web pages. *Displays,* **24**(4), pp. 187-195.

WAGNER, P., 2005. *An undergraduate project.* MSc edn. Keele University.

WEINSCHENK, S., 2011. *100 things every designer needs to know about people.* Berkley: Pearson Education.

WHITAKER, J., 2005. *An undergraduate project.* MSc edn. Keele University.

WHITFIELD, T. and WHILTSHIRE, T., 1990. Color psychology: A critical review. *Genetic, social, and general psychology monographs,*116, pp. 387-412.

YAMAGISHI, T. and YAMAGISHI, M., 1994. Trust and commitment in the United States and Japan. *Motivation and Emotion,* **18**(2), pp. 129-166.

ZAIDEL, D.W., AARDE, S.M. and BAIG, K., 2005. Appearance of symmetry, beauty, and health in human faces. *Brain and Cognition,* **57**(3), pp. 261-263.

ZAIDEL, D.W. and COHEN, J.A., 2005. The face, beauty, and symmetry: Perceiving asymmetry in beautiful faces. *International Journal of Neuroscience,* **115**(8), pp. 1165-1173.

ZAJONC, R.B., 2001. Mere exposure: A gateway to the subliminal. *Current Directions in Psychological Science,* **10**(6), pp. 224-228.

*ZHANG, G., EDDY PATUWO, B. and Y HU, M., 1998. Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting,* **14**(1), pp. 35-62.

Appendix A Images used in Study One

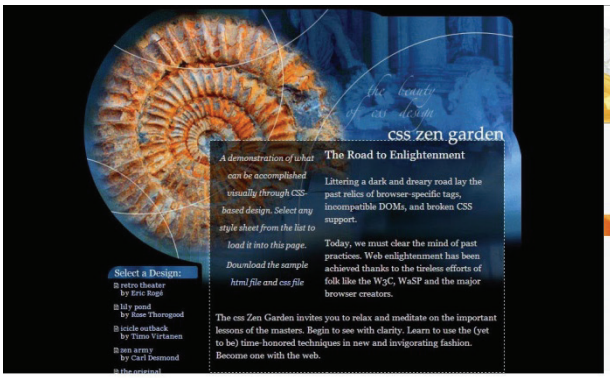


Image 1

Image 2

Image 3

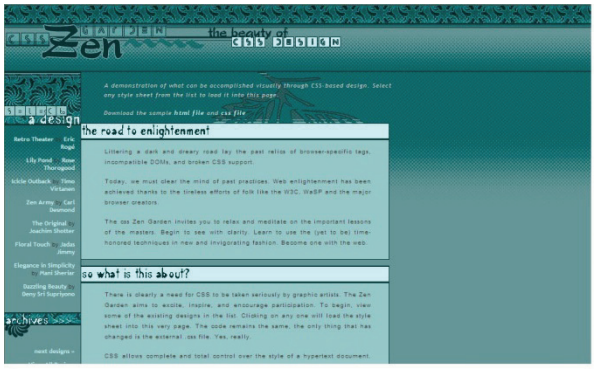


Image 4

Image 5

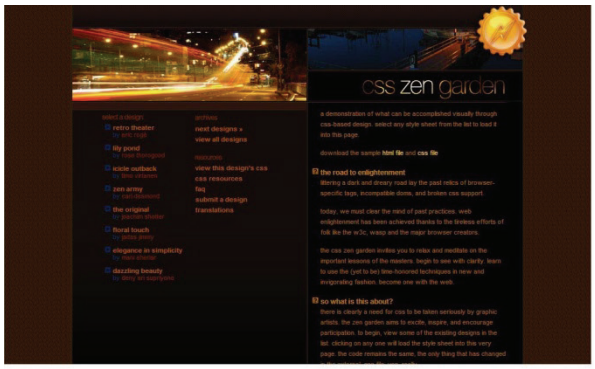


Image 6

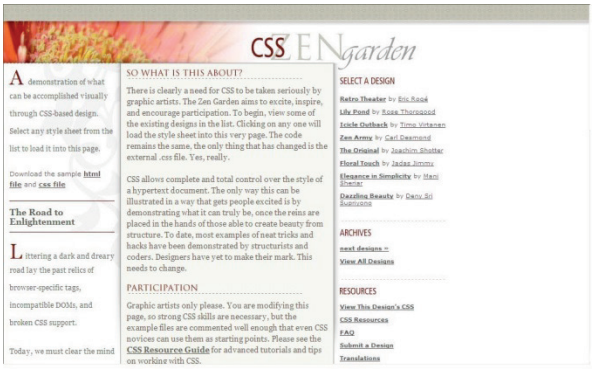


Image 7

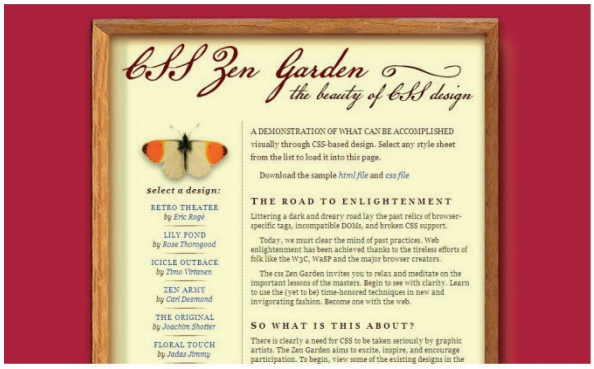


Image 8

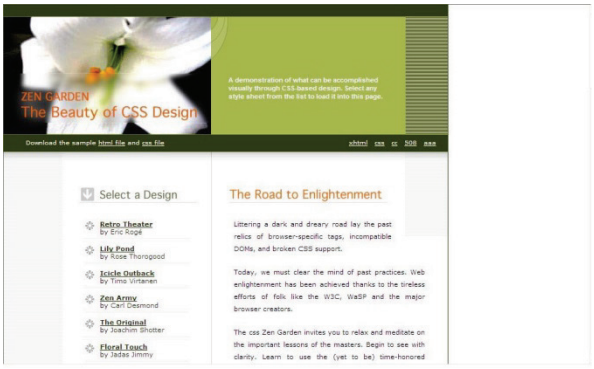


Image 9

Image 10

Image 11

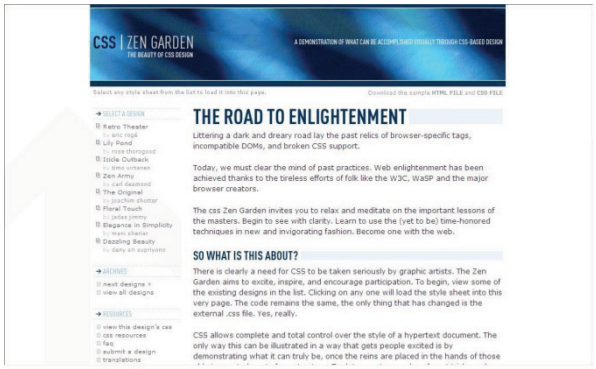


Image 12

Image 13

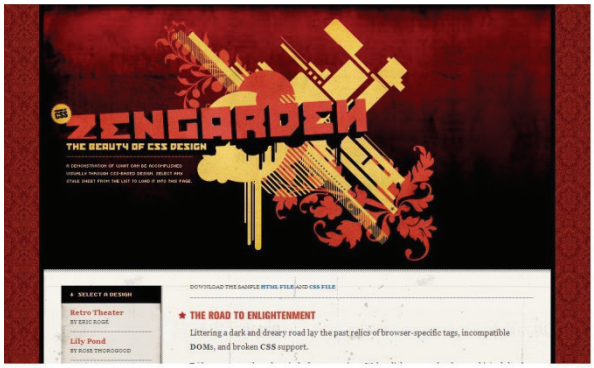


Image 14

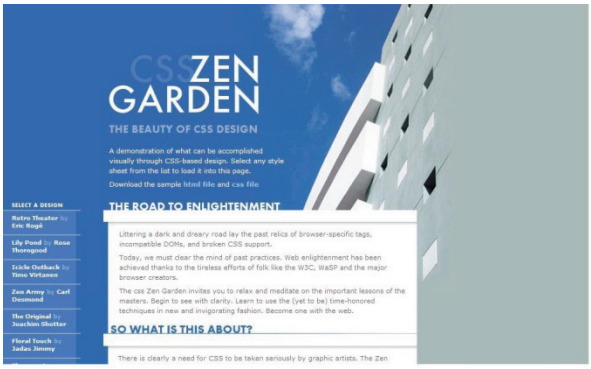


Image 15

Image 16

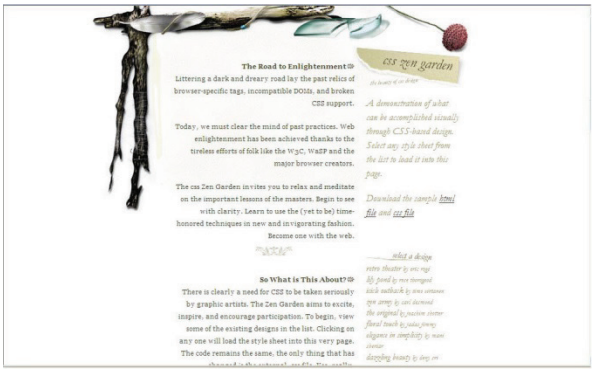


Image 17

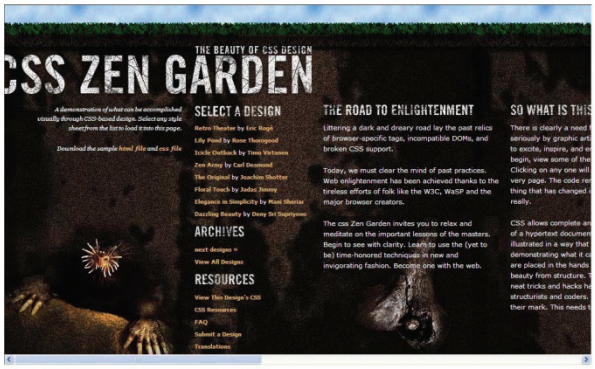


Image 18

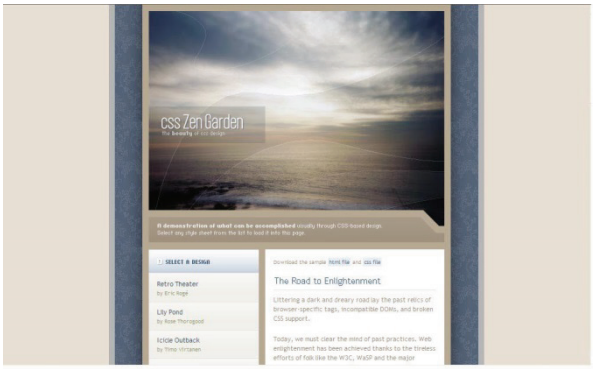


Image 19

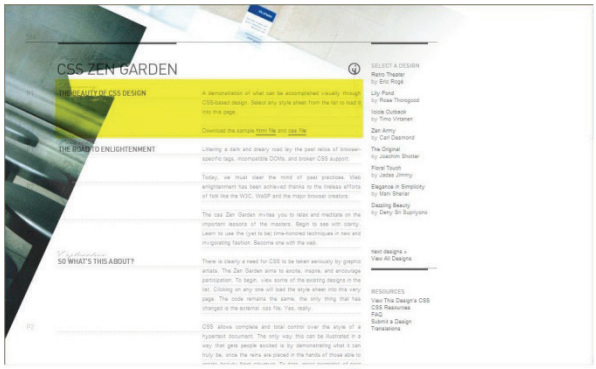


Image 20

Image 21

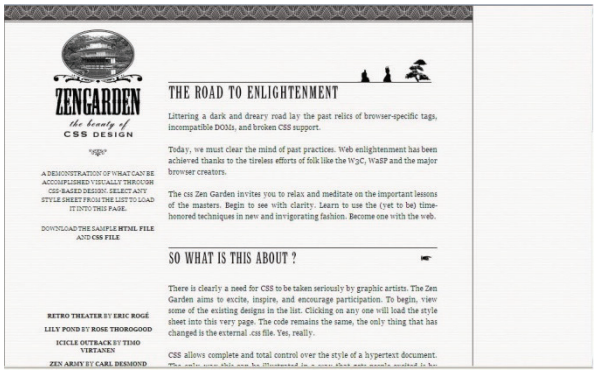


Image 22

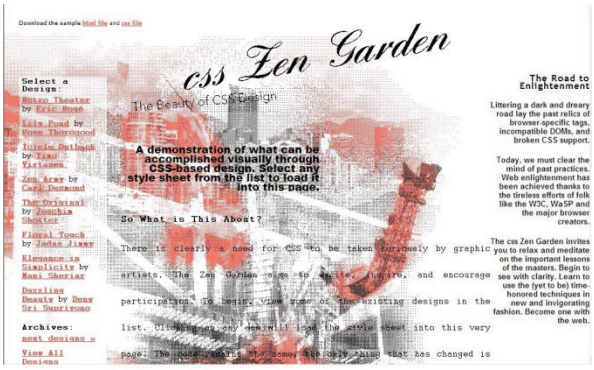


Image 23

Image 24

Appendix B Background questions for Study One

Name:

Age: 17 and Under
18 - 21
22 – 24
25 – 29
30 – 39
40 – 49
50 – 59
60 and above

Gender: Male
Female

Web design experience - where the term 'web site' covers 2 or more linked pages:

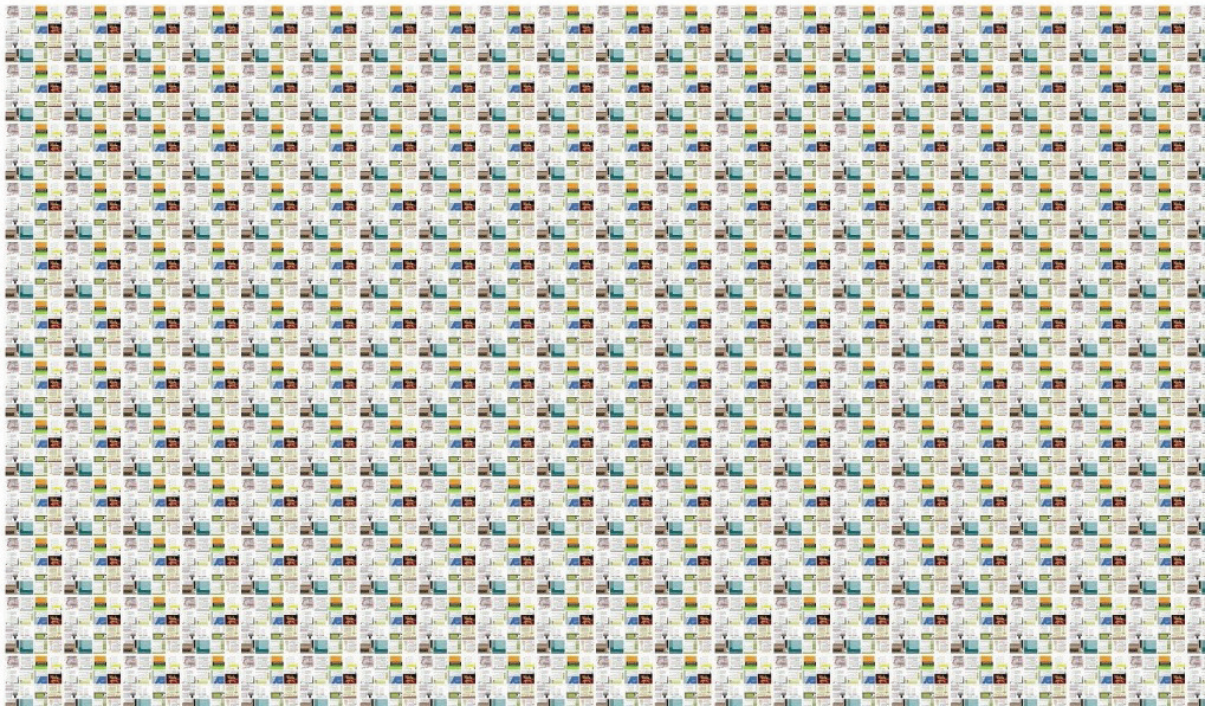
Have never designed a web site before
Have designed one web site
Have designed a few (2 - 10) web sites
Have designed many web sites (10 +) not as a job
Have designed many web sites (10 +) as a job

Web use

Use the Internet daily
Use the Internet a few times a week
Use the Internet weekly
Use the Internet a few times a month
Use the Internet monthly
Use the Internet rarely

Nationality:

Appendix C Masking image used in Study One

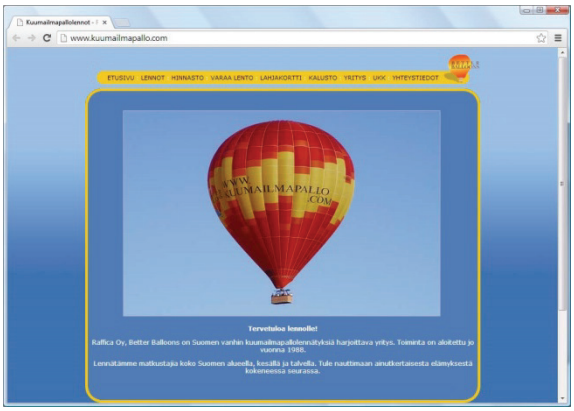


Appendix D List of web pages used in Study Two

1. http://www.kuumailmapallo.com/
2. http://www.ilmailuliitto.fi/index.php?mid=69
3. http://www.aerohot.fi/
4. http://www.ecredo.fi/kuumailmapallo/kuumailmapallo/
5. http://www.aeronaut.fi/
6. http://www.kokemuskauppa.com/index.php?p=119&lang=1
7. http://translate.googleusercontent.com/translate_c?hl=en&rurl=translate.google.com&sl=auto&tl=fi&twu=1&u=http://www.elation.it/in-mongolfiera-sopra-la-toscana_xp367.htm&usg=ALkJrhjvIfNEnMqP6SzRKVFtZS1GvQJryg
8. http://translate.googleusercontent.com/translate_c?hl=en&rurl=translate.google.com&sl=auto&tl=fi&twu=1&u=http://www.yumping.it/volo-mongolfiera&usg=ALkJrhhNn23vexjZ-4BaMC33-9FmpNnAkA
9. http://translate.googleusercontent.com/translate_c?hl=en&rurl=translate.google.com&sl=auto&tl=fi&twu=1&u=http://www.city-discovery.com/it/ID2924_Giro_in_mongolfiera_ad_aria_calda_sul_deserto&usg=ALkJrhiVMp-AQHUdNlBWO2mw5jBta1_VDA
10. http://translate.googleusercontent.com/translate_c?act=url&depth=1&hl=en&ie=UTF8&prev=_t&rurl=translate.google.com&sl=auto&tl=fi&u=http://www.ballonfahrten.com/&usg=ALkJrhhO2i71HLTNXB_v0n7rlgJvE-57Xg
11. http://translate.googleusercontent.com/translate_c?depth=1&hl=en&rurl=translate.google.com&sl=de&tl=fi&u=http://www.ballonservice.de/&usg=ALkJrhhWjANJImNb_df6lcoaoC97TT0Cww
12. http://translate.googleusercontent.com/translate_c?depth=1&hl=en&rurl=translate.google.com&sl=de&tl=fi&u=http://www.jochen-schweizer.de/geschenke/ballonfahrt,default,pd.html&usg=ALkJrhhzdkyfpULdckyQG0HtBBQIxaRK5w
13. http://translate.googleusercontent.com/translate_c?depth=1&hl=en&rurl=translate.google.com&sl=de&tl=fi&u=http://www.bavaria-ballon.de/&usg=ALkJrhj4ixJDzOuIdBcm750GQ4pNLhxk8g
14. http://translate.googleusercontent.com/translate_c?depth=1&hl=en&rurl=translate.google.com&sl=de&tl=fi&u=http://ballonfahren.de/index.html&usg=ALkJrhgaSEvDe_2Cc3QWZ4GdQpEtLUKM2Q
15. http://translate.googleusercontent.com/translate_c?depth=1&hl=en&rurl=translate.google.com&sl=de&tl=fi&u=http://www.sky-fun.de/ballonfahrten/&usg=ALkJrhgYJxaB3Jg9BM2QYNiFFV_mLjbEYQ
16. http://translate.googleusercontent.com/translate_c?depth=1&hl=en&rurl=translate.google.com&sl=de&tl=fi&u=http://www.ballonfahrten-regensburg.de/&usg=ALkJrhhjSQUD9T-pmMl5xHxZrt-xann8Ww
17. http://translate.googleusercontent.com/translate_c?depth=1&hl=en&rurl=translate.google.com&sl=de&tl=fi&u=http://www.eifel-ballooning.de/&usg=ALkJrhjO-HnP0sjTirlygf03wgAmGIRv-A

18. http://translate.googleusercontent.com/translate_c?depth=1&hl=en&rurl=translate.google.com&sl=de&tl=fi&u=http://www.skytours-ballooning.de/region/ballonfahrt-hessen.html&usg=ALkJrhi_WjKUKsVnj3ZukSPxQCxxmxa33A
19. http://translate.googleusercontent.com/translate_c?depth=1&hl=en&rurl=translate.google.com&sl=de&tl=fi&u=http://www.ballonsport-muellheim.de/index.html&usg=ALkJrhhNR5iGm5qFvsaqSdHsDfGNcQl8Sg
20. http://translate.googleusercontent.com/translate_c?depth=1&hl=en&rurl=translate.google.com&sl=de&tl=fi&u=http://www.lausitz-ballonfahrten.de/Startseite.htm%3Fsize%3D1600-772&usg=ALkJrhig4Sq-28Ma0sr_hRjTZf_dquqW7A
21. http://translate.googleusercontent.com/translate_c?depth=1&hl=en&rurl=translate.google.com&sl=de&tl=fi&u=http://www.ballonreisen.de/&usg=ALkJrhiDzGb3HS_r-IDsDVspmcPGMQJnCQ
22. http://translate.googleusercontent.com/translate_c?depth=1&hl=en&rurl=translate.google.com&sl=de&tl=fi&u=http://www.aeroballooning.de/ballonfahrt-hamburg.html&usg=ALkJrhjtj3MQiZAVWqohAjaV-wJH_RDOPQ
23. http://translate.googleusercontent.com/translate_c?act=url&depth=1&hl=en&ie=UTF8&prev=_t&rurl=translate.google.com&sl=auto&tl=fi&u=http://www.ichliebeluft.de/&usg=ALkJrhgmbUZcdrT_RxpYCqDh9SuuZU8Scg
24. http://translate.googleusercontent.com/translate_c?depth=1&hl=en&rurl=translate.google.co.uk&sl=auto&tl=fi&u=http://www.johnbalon.ro/&usg=ALkJrhjKowZBseSg57CynirDGfK9QbS2Bg
25. http://translate.googleusercontent.com/translate_c?depth=1&hl=en&rurl=translate.google.co.uk&sl=sv&tl=fi&u=http://www.uppner.se/upplevelser/luftballong&usg=ALkJrhhqpcXyoY38JCcH5UBxtpk3ZGmlUA
26. http://translate.googleusercontent.com/translate_c?depth=1&hl=en&rurl=translate.google.co.uk&sl=sv&tl=fi&u=http://www.flyg-ballong.nu/&usg=ALkJrhgZOxgO21yPd_0euRLKJ-9NsaiOKQ
27. http://translate.googleusercontent.com/translate_c?depth=1&hl=en&rurl=translate.google.co.uk&sl=sv&tl=fi&u=http://www.ballongflygisyd.se/&usg=ALkJrhjT28BzT04gZemYrNQcVni_q4KNmQ
28. http://translate.googleusercontent.com/translate_c?depth=1&hl=en&rurl=translate.google.co.uk&sl=sv&tl=fi&u=http://www.farochflyg.se/&usg=ALkJrhgTHVHraetPRRULoMDoF1YuIreAZA
29. http://translate.googleusercontent.com/translate_c?depth=1&hl=en&rurl=translate.google.co.uk&sl=ca&tl=fi&u=http://www.globubolg.com/index.html&usg=ALkJrhjUhPtY0USelD7Ok3E-KJKmgCdozQ
30. http://translate.googleusercontent.com/translate_c?depth=1&hl=en&rurl=translate.google.co.uk&sl=ca&tl=fi&u=http://www.ballooning.es/ec/&usg=ALkJrhj-kGEF99pEIotAclK8Ioc0NngrgQ

Appendix E Cards used in Study Two



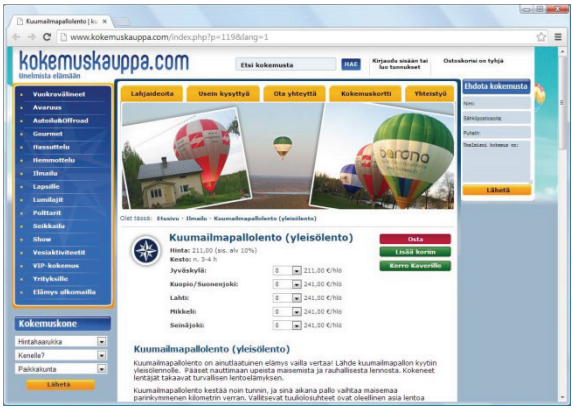
Card 1

Card 2

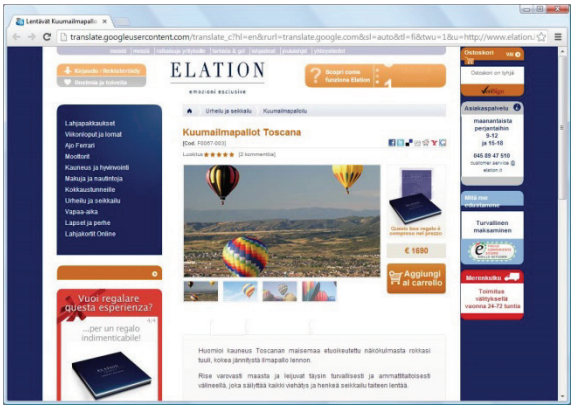
Card 3

Card 4

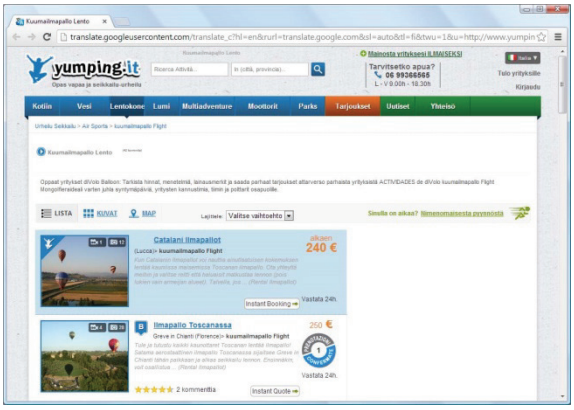
Card 5



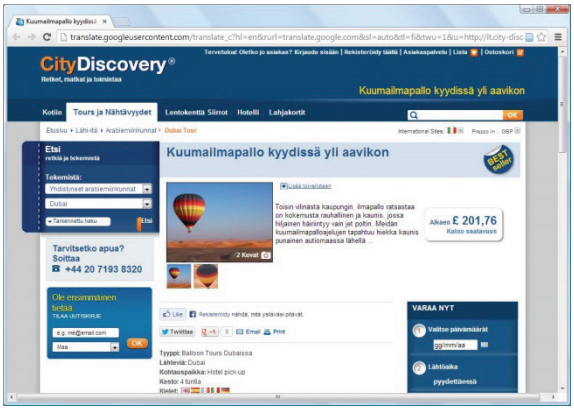
Card 6



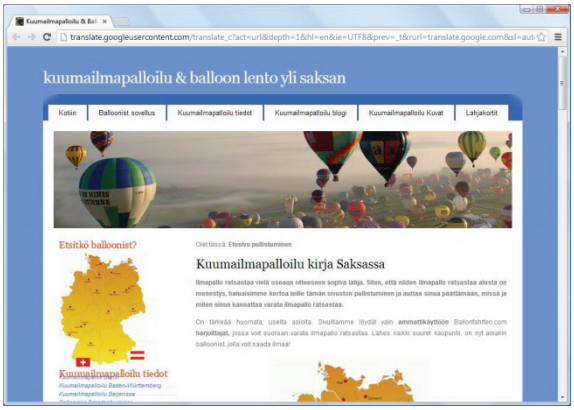
Card 7



Card 8

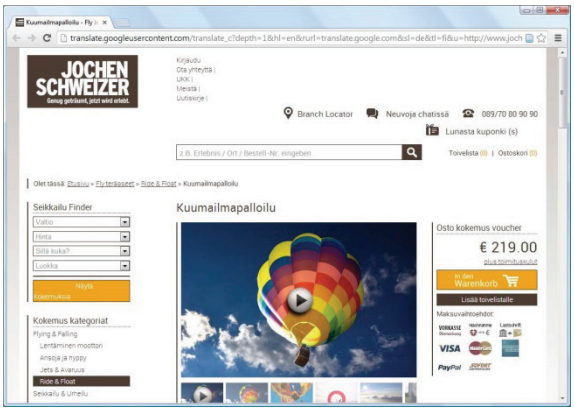


Card 9

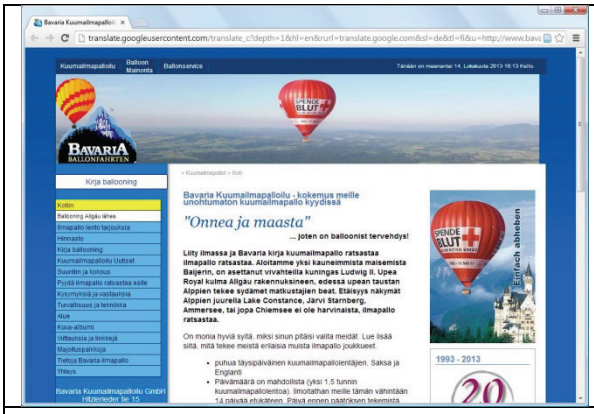


Card 10

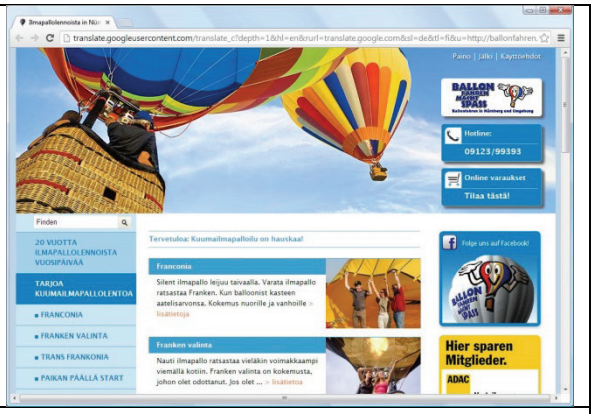
Card 11



Card 12

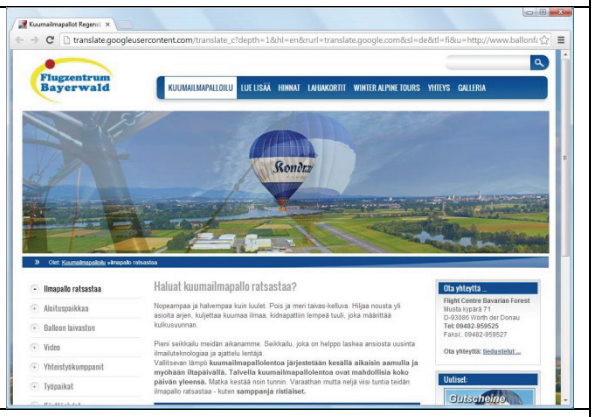


Card 13



Card 14

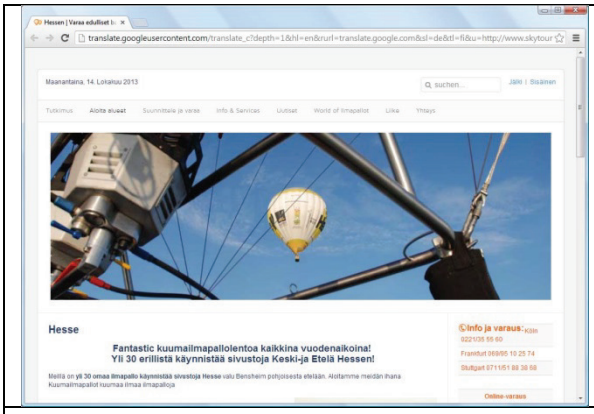
Card 15



Card 16

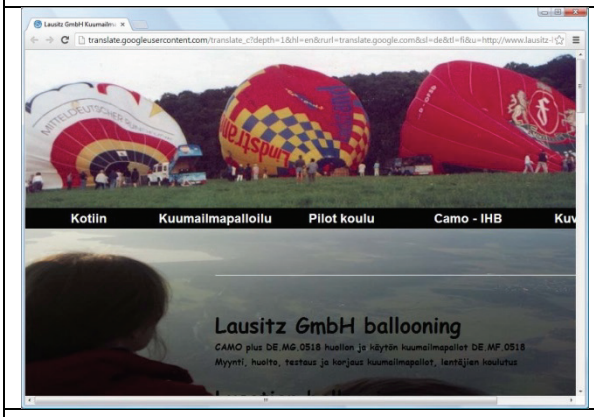
Card 17

Card 18

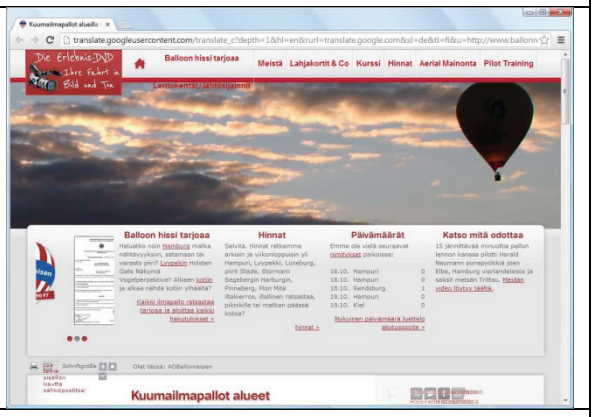


Card 19

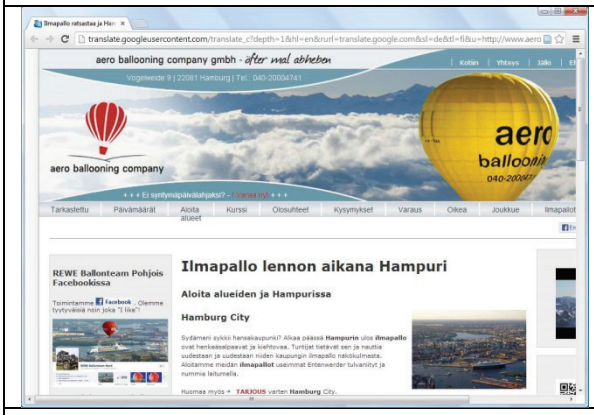
Card 20



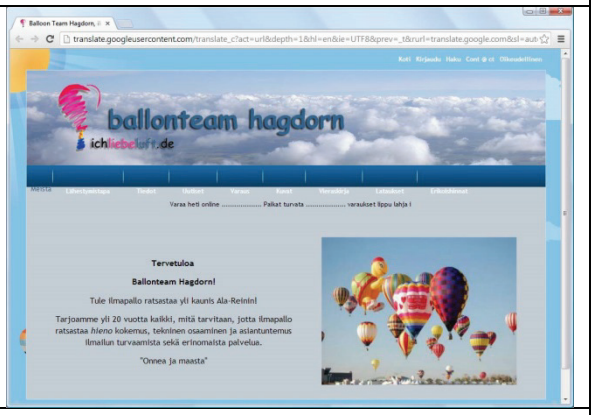
Card 21



Card 22

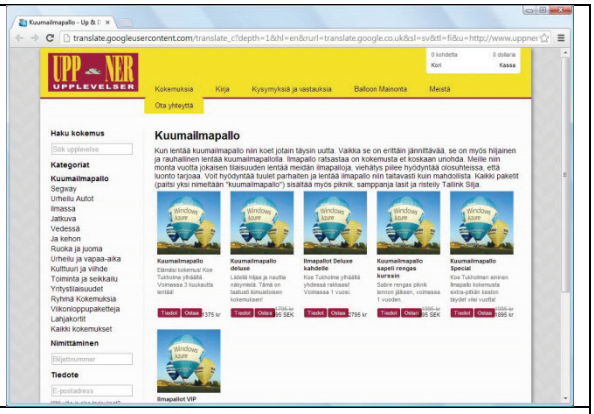


Card 23

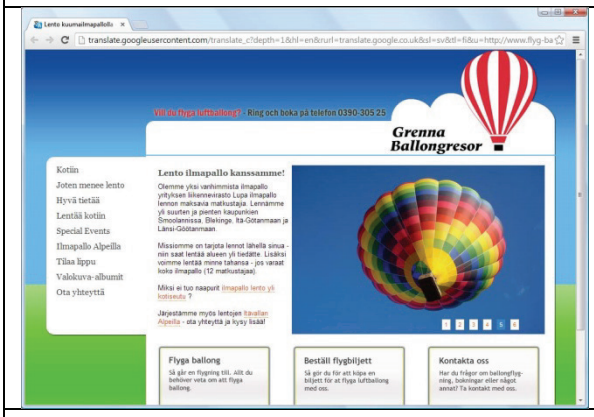


Card 24

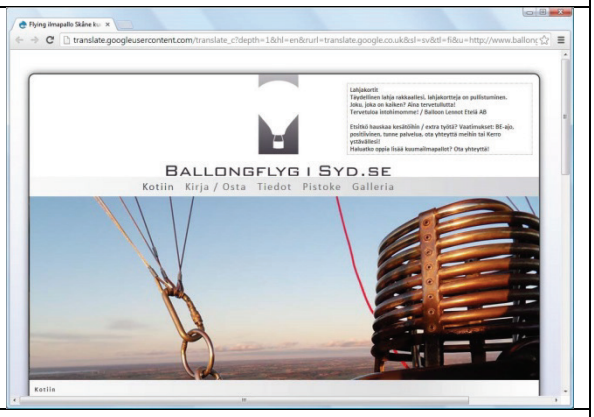
Card 25



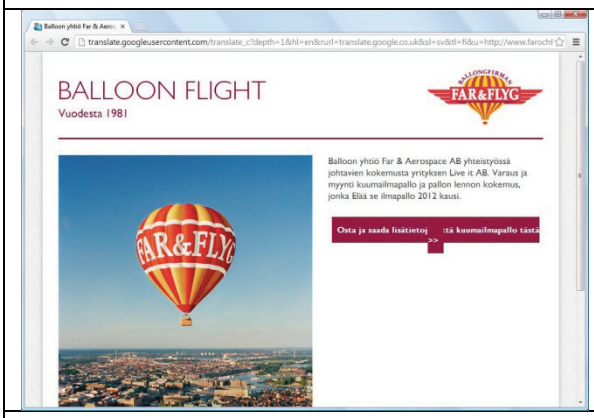
Card 26



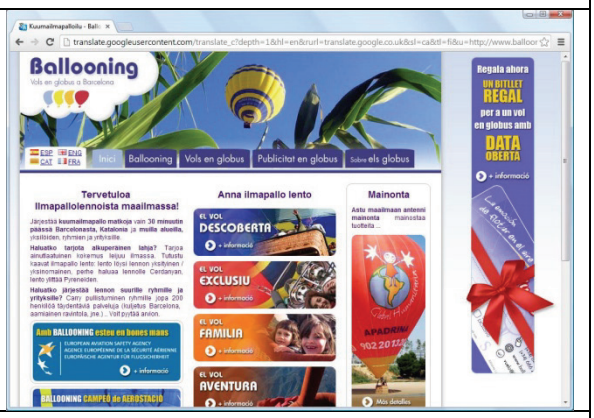
Card 27



Card 28



Card 29



Card 30

Appendix F Invitation to take part Study Two: Online data collection

I'm doing my PhD in Computer Science at Keele University. I am conducting an online survey. I'm trying to get as many people as possible to complete this. It should take no more than 15 minutes and involves looking at some images and answering questions about them.

If you are willing to help please go to
http://www.ccit.co.uk/webmetrics/
where you will find more information about the study and be able to take part.

If you could post a link to this group on your own profile I would be really grateful.

Thanks,

Nikki

My contact details:

Nikki Williams
CR 40
Colin Reeves Building
Keele University
Staffordshire
ST5 5BG
E-mail: n.k.williams@keele.ac.uk
Tel: 01782 733253

Contact details for the Research Governance Officer at Keele University (if you do not wish to contact the researcher directly):

Nicola Leighton
Research Governance Officer
Research & Enterprise Services
Dorothy Hodgkin Building
Keele University
ST5 5BG
E-mail: n.leighton@uso.keele.ac.uk
Tel: 01782 733306

Appendix G Consent form for Study Two: Online data collection

CONSENT FORM

Title of Project: A study of website aesthetics Part A
Name of Principal Investigator: Nikki Williams

Please tick box

1	I confirm that I have read and understand the information sheet for the above study and have had the opportunity to ask questions.	☐
2	I understand that my participation is voluntary and that I am free to withdraw at any time.	☐
3	I agree to take part in this study.	☐
4	I understand that data collected about me during this study will be made anonymous before publication.	☐
5	I agree to allow the data collected to be used for future research projects.	☐
6	I agree to be contacted about possible participation in future research projects.	☐

Contact Details

Nikki Williams
CR 40
Colin Reeves Building
Keele University
Staffordshire
ST5 5BG
E-mail: n.k.williams@keele.ac.uk
Tel: 01782 733253

Appendix H Information sheet for Study Two: Online data collection

Information Sheet

Study Title: A study of website aesthetics Part A

Aims of the Research
This research aims to identify predictors for users' opinions of websites. In the first part of the study users' opinions need to be gathered along with various objective measures. These objective measures include, but are not limited to, the word count, the colour balance on the page, the number of images, the background colour of the page, and the most used colour within the page. These results will then be analysed for correlations, to see if predictors exist. In the second part, information regarding user's grouping of websites will be gathered, analysed and used to complement the first section findings.

Invitation
You are being invited to consider taking part in the research study: A study of website aesthetics. This project is being undertaken by Nikki Williams (a PhD student within Computing at Keele University).

Before you decide whether or not you wish to take part, it is important for you to understand why this research is being done and what it will involve. Please take time to read this information carefully and discuss it with friends and relatives if you wish. Ask us if there is anything that is unclear or if you would like more information.

By taking part online you are giving consent for your information and responses to be used for this research.

Why have I been chosen?
Participants have been gathered by inviting a group of people to take part, then asking them to invite their friends, and so on. The mediums for delivering these invitations include using email and social networking. There are no restrictions on who can take part, except that you do not speak Finnish to a level where you could understand written content at a glance.

Do I have to take part?
You are free to decide whether you wish to take part or not. If you do decide to take part you will be asked to complete the online permission form. You are free to withdraw from this study at any time and without giving reasons.

What will happen if I take part?
Participants will be asked to complete a brief background questionnaire. Participants will then be shown 30 web pages for hot air balloon companies and asked to give their ratings in relation to six factors. The questions will appear one at a time at the top of the screen *5 seconds* after the web page is shown. The page will remain visible for *30 seconds* before vanishing. The task should take no more than 15 minutes.

If I take part, what do I have to do?
You will be expected to complete the background questionnaire and answer six questions about all 30 sites. Each question will be answered using a slider on a scale.

What are the benefits (if any) of taking part?

There are no obvious benefits to participants.

What are the risks (if any) of taking part?
There are no obvious risks to participants.

What if there is a problem?

If you have a concern about any aspect of this study, you may speak to the researcher who will do their best to answer your questions. You should contact Nikki Williams on n.k.williams@keele.ac.uk. Alternatively, if you do not wish to contact the researcher you may contact Nicola Leighton, Research Governance Officer on 01782 733306 or n.leighton@uso.keele.ac.uk

If you remain unhappy about the research and/or wish to raise a complaint about any aspect of the way that you have been approached or treated during the course of the study please write to Nicola Leighton who is the University's contact for complaints regarding research at the following address:-

Nicola Leighton
Research Governance Officer
Research & Enterprise Services
Dorothy Hodgkin Building
Keele University
ST5 5BG
E-mail: n.leighton@uso.keele.ac.uk
Tel: 01782 733306

How will information about me be used?
The data will be collected electronically and stored on a database. The data will be analysed using statistical software. The data will be retained for use in the second part of this study, and potentially for future studies for up to five years after this study. Data will be made anonymous before analysis, with names being stored in a separate, password protected, file. The background information provided will be used to group results.

Who will have access to information about me?
All personal details *(eg consent forms)* and data obtained relating to the research will be stored on a password protected computer. The data collected will be anonymous. Your personal details will remain confidential and will not be linked to your responses. In accordance with Keele University Guidelines all data and consent forms will be retained by the researcher for 5 years after which they will be securely disposed of.

Who is funding and organising the research?
The research is being undertaken as part of a PhD in Computer Science.

Contact for further information

Nikki Williams
CR 40
Colin Reeves Building
Keele University
Staffordshire
ST5 5BG
E-mail: n.k.williams@keele.ac.uk
Tel: 01782 733253

Gordon Rugg
CR 102
Colin Reeves Building
Keele University
Staffordshire
ST5 5BG
Email: g.rugg@cs.keele.ac.uk
Tel: 01782 733410

Appendix I Ethical approval letter

RESEARCH AND ENTERPRISE SERVICES

18th September 2013

Nikki Williams
CR40
Colin Reeves Building
Keele University

Dear Nikki,

Re: A study of website aesthetics

Thank you for submitting your application amendment for review.

I am pleased to inform you that your amendment has been approved by the Ethics Review Panel.

If the fieldwork goes beyond the date stated in your application, you must notify the Ethical Review Panel via the ERP administrator at uso.erps@keele.ac.uk stating ERP1 in the subject line of the e-mail.

If there are any other amendments to your study you must submit an 'application to amend study' form to the ERP administrator stating ERP1 in the subject line of the e-mail. This form is available via http://www.keele.ac.uk/researchsupport/researchethics/

If you have any queries, please do not hesitate to contact me via the ERP administrator on uso.erps@keele.ac.uk stating ERP1 in the subject line of the e-mail.

Yours sincerely

pp

Dr Jackie Waterfield
Chair – Ethical Review Panel

CC RI Manager
Supervisor

Research and Enterprise Services, Keele University, Staffordshire, ST5 5BG, UK
Telephone: + 44 (0)1782 734466 Fax: + 44 (0)1782 733740

Appendix J Mean and standard deviation of subjective ratings for each image in Study One, displayed by question

Mean and standard deviation by timing for Question A						
How eye-catching do you find this page?						
Image	Mean	Mean a	Mean b	St Dev	St Dev a	St Dev b
1	68.44	71.70	64.59	22.64	20.43	24.71
2	54.40	61.13	46.46	23.89	22.64	23.13
3	45.09	49.65	39.85	23.78	24.22	22.43
4	45.67	49.14	41.66	24.98	23.96	25.84
5	48.10	52.71	42.77	25.60	26.00	24.37
6	55.27	55.74	54.72	24.11	24.50	23.95
7	41.80	46.54	36.05	20.83	20.52	19.98
8	60.03	62.11	57.65	23.48	23.35	23.70
9	53.04	56.30	49.18	25.01	23.75	26.20
10	69.69	67.48	72.23	23.79	22.37	25.37
11	57.59	61.00	53.68	25.67	23.52	27.71
12	39.14	41.36	36.53	23.55	22.37	24.93
13	54.81	62.40	46.28	23.12	21.58	22.01
14	67.35	69.00	65.39	21.23	18.67	24.03
15	57.74	59.07	56.25	25.45	26.15	24.88
16	44.25	48.96	38.69	23.68	22.53	24.07
17	44.64	48.72	40.25	24.32	23.09	25.12
18	52.94	55.67	49.79	25.04	27.93	21.14
19	51.35	55.44	46.75	25.42	24.24	26.22
20	37.44	41.22	33.37	22.28	22.26	21.85
21	59.84	62.98	56.13	24.70	25.23	23.84
22	31.10	37.36	24.00	20.66	21.23	17.71
23	39.55	39.62	39.49	24.52	25.67	23.66
24	52.57	54.71	50.10	21.10	19.80	22.52

Mean and standard deviation by timing for Question B						
How visually attractive do you find this page?						
Image	Mean	Mean a	Mean b	St Dev	St Dev a	St Dev b
1	64.60	66.96	61.82	23.89	23.62	24.22
2	58.24	66.65	48.31	24.68	21.55	24.67
3	58.60	65.67	50.44	23.49	20.84	23.98
4	40.52	46.70	33.37	24.86	23.89	24.33
5	48.81	55.29	41.13	25.20	24.95	23.58
6	54.76	56.98	52.15	25.64	24.87	26.59
7	46.33	50.85	41.00	22.00	21.76	21.34
8	57.53	61.04	53.50	25.60	26.53	24.19
9	56.63	61.70	50.80	23.60	20.60	25.67
10	58.81	58.02	59.70	26.39	24.68	28.49
11	60.05	62.13	57.53	23.96	22.61	25.57
12	40.18	41.04	39.13	25.15	23.51	27.28
13	57.64	63.36	51.20	24.40	24.06	23.42
14	55.94	55.63	56.32	25.92	25.69	26.53
15	56.75	61.44	51.33	25.38	24.30	25.82
16	47.64	54.07	40.05	23.31	23.02	21.55
17	49.02	53.57	44.03	22.66	23.43	20.94
18	41.84	44.66	38.67	25.75	26.92	24.31
19	59.07	63.24	54.28	22.54	20.80	23.74
20	31.06	33.76	28.23	20.20	20.45	19.79
21	41.24	40.98	41.54	24.49	25.25	23.93
22	36.20	40.05	31.84	22.71	23.95	20.68
23	27.54	27.30	27.77	23.38	23.68	23.40
24	53.65	58.02	48.62	21.99	19.55	23.77

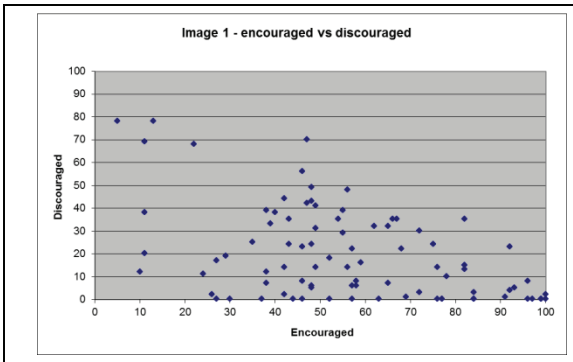
Mean and standard deviation by timing for Question C						
How much does this page encourage you to explore the rest of the site?						
Image	Mean	Mean a	Mean b	St Dev	St Dev a	St Dev b
1	56.27	58.59	53.60	24.04	23.91	24.20
2	48.36	54.09	41.78	23.64	23.05	22.84
3	49.87	56.22	42.18	22.77	21.03	22.68
4	36.12	40.84	30.51	22.22	22.27	21.09
5	47.24	53.16	40.24	24.40	24.98	22.01
6	49.51	51.84	46.87	24.64	24.64	24.69
7	43.49	49.00	37.00	23.24	22.78	22.34
8	52.88	58.20	46.90	24.79	25.54	22.78
9	51.55	54.15	48.49	25.23	22.37	28.22
10	51.19	48.64	54.05	28.13	26.22	30.22
11	52.44	54.31	50.28	23.63	23.27	24.17
12	42.60	45.09	39.66	27.37	26.87	28.02
13	50.80	54.51	46.51	23.60	22.80	24.07
14	50.06	47.67	53.06	26.85	27.43	26.17
15	49.33	50.00	48.59	22.57	22.66	22.73
16	45.54	50.98	39.11	24.42	22.33	25.51
17	41.16	42.74	39.45	23.68	24.13	23.36
18	39.18	38.89	39.53	25.27	26.84	23.64
19	49.51	50.49	48.38	23.20	22.90	23.79
20	28.12	29.95	26.03	17.52	19.67	14.69
21	40.20	40.79	39.51	24.29	24.12	24.80
22	36.41	42.86	28.92	22.59	22.50	20.55
23	23.72	23.14	24.23	22.52	20.57	24.40
24	46.24	50.53	41.28	21.93	21.49	21.65

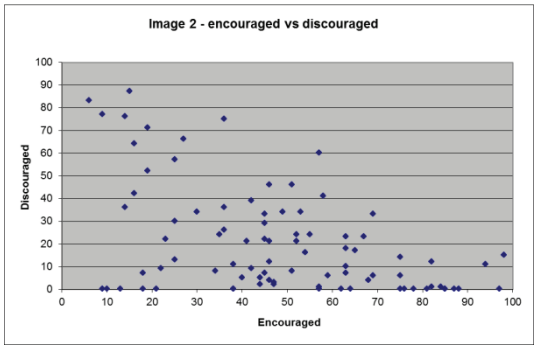
Mean and standard deviation by timing for Question D						
How well designed is this page?						
Image	Mean	Mean a	Mean b	St Dev	St Dev a	St Dev b
1	55.78	57.37	53.95	22.48	21.64	23.54
2	51.86	57.74	45.10	23.72	21.13	24.98
3	54.92	61.93	46.82	22.13	19.90	22.03
4	37.13	42.00	31.17	20.91	19.87	20.87
5	49.44	55.31	42.67	24.23	24.64	22.18
6	48.42	47.31	49.69	24.28	24.82	23.90
7	49.69	56.65	41.26	22.36	21.09	21.16
8	50.70	54.78	46.00	25.19	26.26	23.35
9	52.51	55.33	49.28	26.16	23.26	29.11
10	52.02	47.89	56.78	25.28	22.92	27.27
11	53.14	54.24	51.87	23.46	23.70	23.42
12	46.89	50.43	42.61	27.51	26.35	28.61
13	51.13	55.67	46.03	22.28	22.23	21.48
14	49.51	49.93	49.03	24.47	24.99	24.19
15	51.87	51.93	51.80	22.77	20.53	25.32
16	48.64	54.54	41.30	24.50	23.48	24.04
17	44.54	47.11	41.70	22.15	22.04	22.21
18	41.76	43.35	40.00	23.55	25.40	21.51
19	48.91	50.93	46.63	22.94	23.59	22.27
20	29.73	33.16	26.39	18.31	20.62	15.27
21	40.71	39.86	41.66	23.69	24.22	23.39
22	41.38	47.56	34.19	21.28	20.21	20.46
23	23.28	22.88	23.62	20.87	21.18	20.89
24	49.29	54.00	43.97	21.84	22.70	19.78

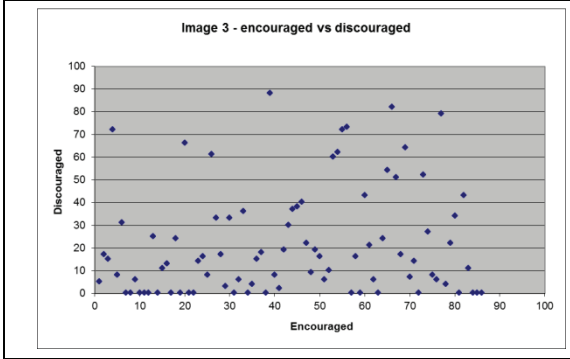
Mean and standard deviation by timing for Question E						
How good is this page as an advertisement for the website?						
Image	Mean	Mean a	Mean b	St Dev	St Dev a	St Dev b
1	55.47	59.82	50.58	25.31	23.75	26.40
2	49.36	53.76	44.03	22.42	21.24	22.93
3	47.66	53.89	40.31	25.34	24.25	24.90
4	34.64	38.80	29.70	22.77	21.15	23.91
5	42.72	47.78	36.74	22.50	23.63	19.74
6	45.37	44.43	46.44	25.10	25.53	24.90
7	40.92	45.54	35.32	21.75	21.98	20.35
8	48.13	51.33	44.45	23.75	23.75	23.50
9	50.42	50.30	50.56	24.82	21.04	28.95
10	52.24	49.85	55.00	26.75	26.04	27.62
11	51.37	51.87	50.79	26.20	25.92	26.84
12	39.21	39.25	39.16	24.70	23.56	26.32
13	49.91	56.56	42.43	22.07	19.86	22.27
14	48.86	48.19	49.65	25.43	24.85	26.43
15	49.24	51.20	47.03	24.32	20.68	27.98
16	42.49	46.44	37.68	23.90	24.51	22.53
17	40.06	41.44	38.54	23.00	23.34	22.83
18	36.93	35.31	38.94	24.16	26.62	20.88
19	47.32	51.09	42.97	24.45	24.31	24.20
20	28.68	31.08	26.00	18.02	19.61	15.90
21	38.22	39.05	37.26	24.24	24.74	23.98
22	31.23	35.58	26.16	19.36	18.07	19.82
23	21.71	19.91	23.51	18.54	16.82	20.20
24	47.35	51.36	42.61	20.93	20.69	20.47

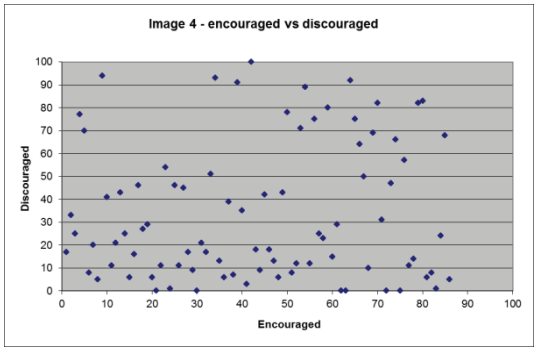
Mean and standard deviation by timing for Question F						
How much does this page discourage you to explore the rest of the site?						
Image	Mean	Mean a	Mean b	St Dev	St Dev a	St Dev b
1	24.882	24.865	24.903	19.512	18.639	20.819
2	26.206	24.568	28.161	23.116	20.266	26.330
3	28.508	24.818	32.313	23.253	21.214	24.947
4	35.815	31.295	41.189	29.245	26.948	31.281
5	33.176	31.472	35.094	23.876	24.858	22.962
6	33.871	35.757	31.758	27.289	28.470	26.176
7	31.959	29.054	34.944	26.097	22.297	29.523
8	29.275	28.861	29.727	23.368	24.937	21.905
9	27.290	30.171	24.324	24.760	23.085	26.388
10	30.903	32.238	29.033	27.337	26.548	28.758
11	28.387	28.462	28.306	25.319	24.169	26.854
12	34.836	32.462	37.559	25.922	22.952	29.073
13	25.794	22.371	29.424	21.582	17.453	25.004
14	33.125	34.842	31.206	26.969	26.635	27.608
15	29.059	27.694	30.594	25.242	22.806	28.023
16	33.000	32.684	33.353	23.277	22.397	24.558
17	32.222	35.568	28.686	25.716	28.377	22.435
18	40.696	42.070	39.056	30.607	32.699	28.276
19	28.065	28.070	28.059	24.234	22.359	26.763
20	50.367	51.628	48.861	29.748	27.500	32.564
21	43.154	47.488	38.351	30.252	31.873	27.993
22	40.104	36.325	44.189	29.665	27.736	31.486
23	61.788	64.444	58.371	28.858	28.946	28.799
24	30.986	29.081	33.121	25.096	24.661	25.786

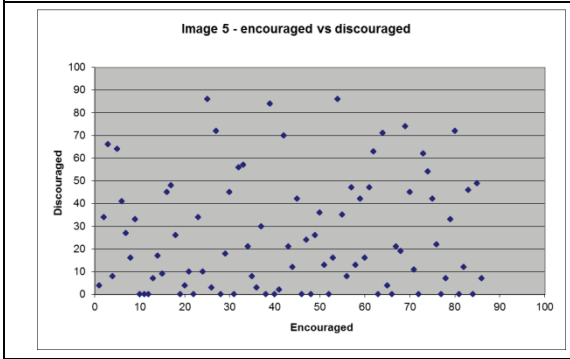
Appendix K Scatter graphs of encourage and discourage for each image for Study One

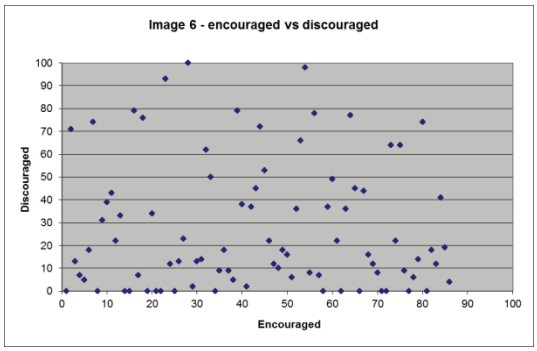


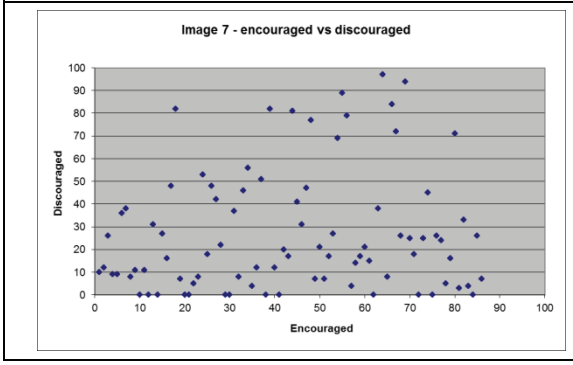


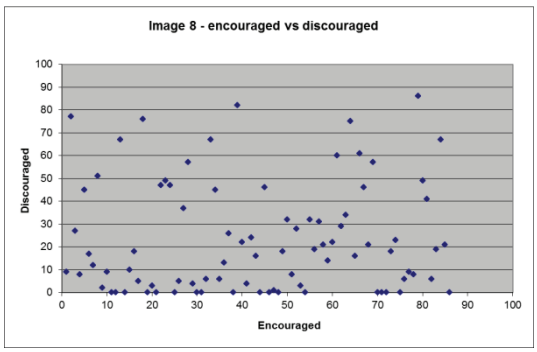












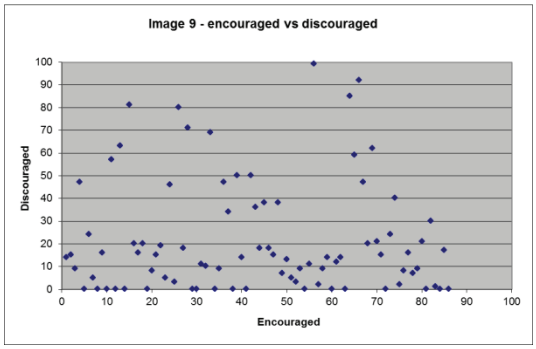


Image 9 - encouraged vs discouraged
Discouraged
Encouraged

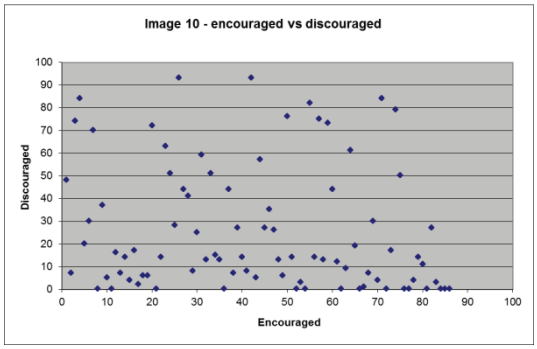


Image 10 - encouraged vs discouraged
Discouraged
Encouraged

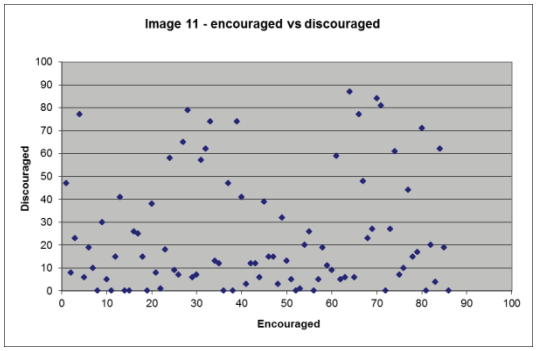


Image 11 - encouraged vs discouraged
Discouraged
Encouraged

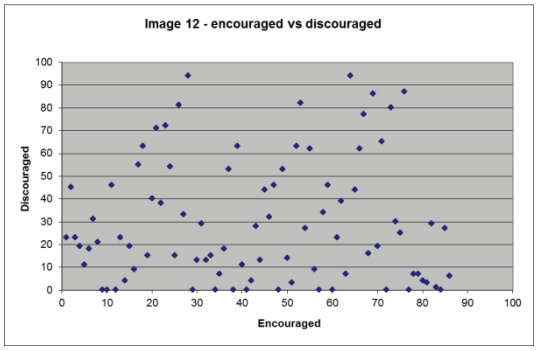


Image 12 - encouraged vs discouraged
Discouraged
Encouraged

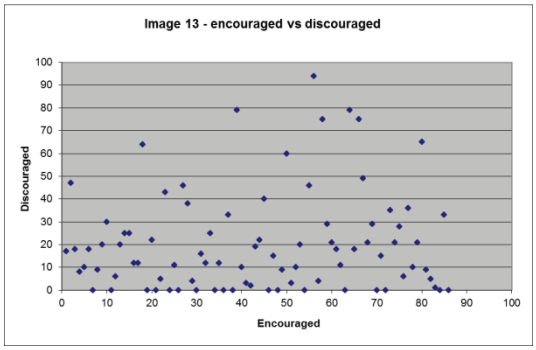


Image 13 - encouraged vs discouraged
Discouraged
Encouraged

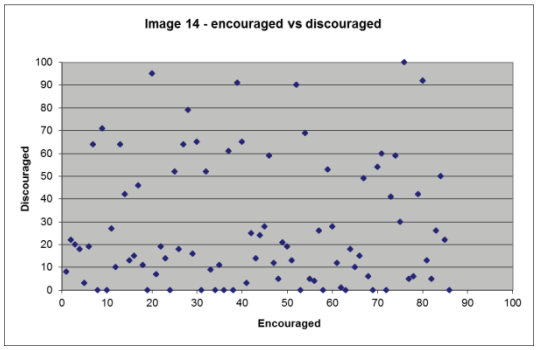


Image 14 - encouraged vs discouraged
Discouraged
Encouraged

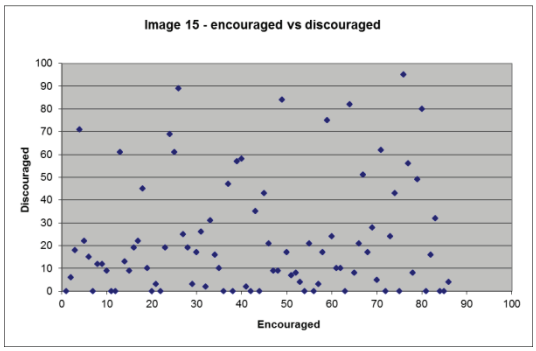


Image 15 - encouraged vs discouraged
Discouraged
Encouraged

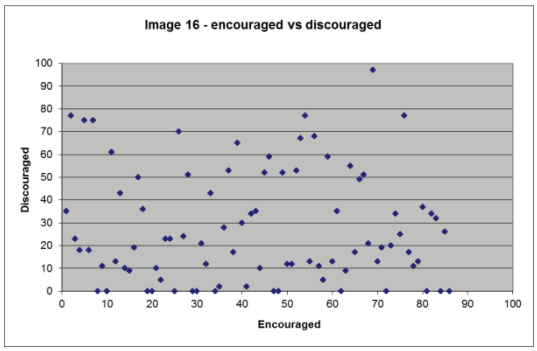


Image 16 - encouraged vs discouraged
Discouraged
Encouraged

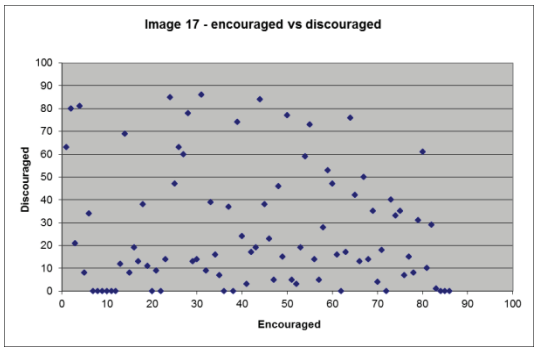


Image 17 - encouraged vs discouraged
Discouraged
Encouraged

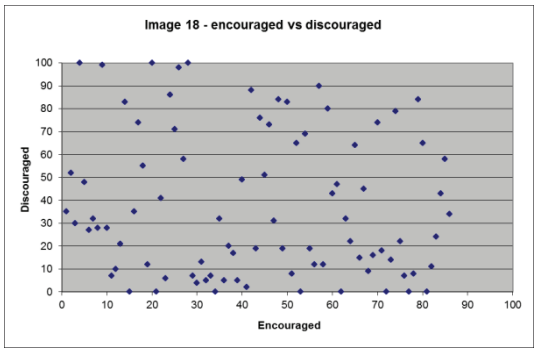


Image 18 - encouraged vs discouraged
Discouraged
Encouraged

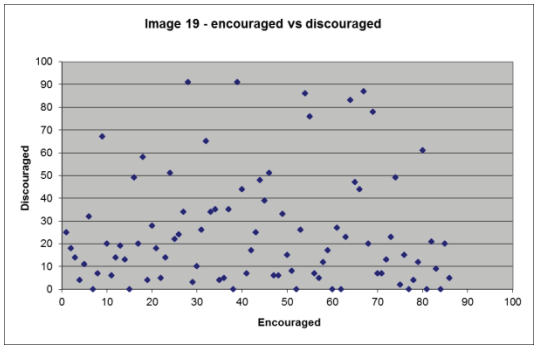


Image 19 - encouraged vs discouraged
Discouraged
Encouraged
0 10 20 30 40 50 60 70 80 90 100

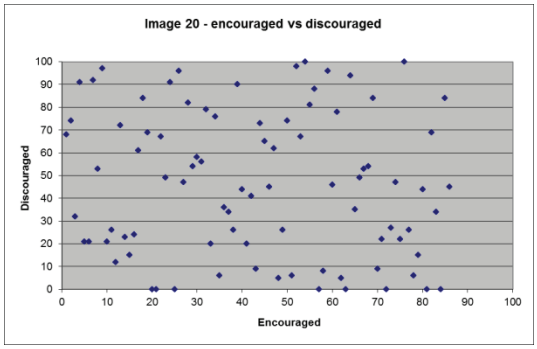


Image 20 - encouraged vs discouraged
Discouraged
Encouraged
0 10 20 30 40 50 60 70 80 90 100

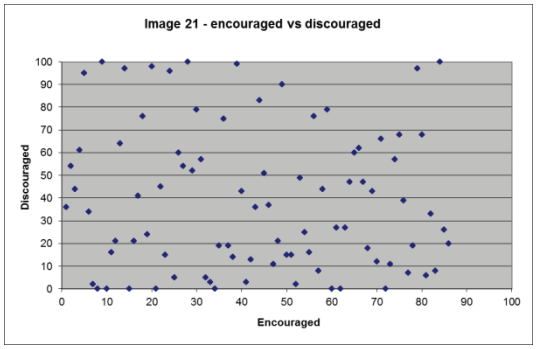


Image 21 - encouraged vs discouraged
Discouraged
Encouraged
0 10 20 30 40 50 60 70 80 90 100

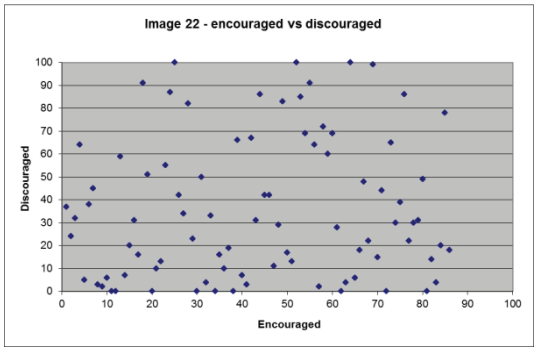


Image 22 - encouraged vs discouraged
Discouraged
Encouraged
0 10 20 30 40 50 60 70 80 90 100

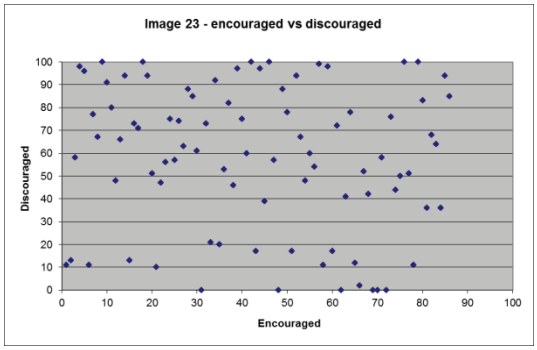


Image 23 - encouraged vs discouraged
Discouraged
Encouraged
0 10 20 30 40 50 60 70 80 90 100

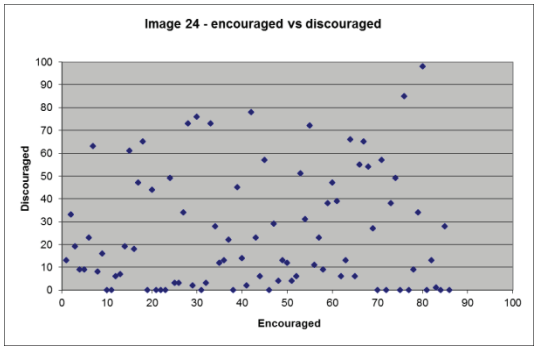


Image 24 - encouraged vs discouraged
Discouraged
Encouraged
0 10 20 30 40 50 60 70 80 90 100

Appendix L Study One mean subjective ratings by viewing condition

Images in rank order for viewing condition A

TimeA	Question a		Question b		Question c		Question d		Question e		Question f	
Rank	Website	Score	Website	Score	Website	Score	Website	Score	Website	Score	Website	Score
1	1	71.70	1	66.96	1	58.59	3	61.93	1	59.82	23	64.44
2	14	69.00	2	66.65	8	58.20	2	57.74	13	56.56	20	51.63
3	10	67.48	3	65.67	3	56.22	1	57.37	3	53.89	21	47.49
4	21	62.98	13	63.36	13	54.51	7	56.65	2	53.76	18	42.07
5	13	62.40	19	63.24	11	54.31	13	55.67	11	51.87	22	36.33
6	8	62.11	11	62.13	9	54.15	9	55.33	24	51.36	6	35.76
7	2	61.13	9	61.70	2	54.09	5	55.31	8	51.33	17	35.57
8	11	61.00	15	61.44	5	53.16	8	54.78	15	51.20	14	34.84
9	15	59.07	8	61.04	6	51.84	16	54.54	19	51.09	16	32.68
10	9	56.30	10	58.02	16	50.98	11	54.24	9	50.30	12	32.46
11	6	55.74	24	58.02	24	50.53	24	54.00	10	49.85	10	32.24
12	18	55.67	6	56.98	19	50.49	15	51.93	14	48.19	5	31.47
13	19	55.44	14	55.63	15	50.00	19	50.93	5	47.78	4	31.30
14	24	54.71	5	55.29	7	49.00	12	50.43	16	46.44	9	30.17
15	5	52.71	16	54.07	10	48.64	14	49.93	7	45.54	24	29.08
16	3	49.65	17	53.57	14	47.67	10	47.89	6	44.43	7	29.05
17	4	49.14	7	50.85	12	45.09	22	47.56	17	41.44	8	28.86
18	16	48.96	4	46.70	22	42.86	6	47.31	12	39.25	11	28.46
19	17	48.72	18	44.66	17	42.74	17	47.11	21	39.05	19	28.07
20	7	46.54	12	41.04	4	40.84	18	43.35	4	38.80	15	27.69
21	12	41.36	21	40.98	21	40.79	4	42.00	22	35.58	1	24.86
22	20	41.22	22	40.05	18	38.89	21	39.86	18	35.31	3	24.82
23	23	39.62	20	33.76	20	29.95	20	33.16	20	31.08	2	24.57
24	22	37.36	23	27.30	23	23.14	23	22.88	23	19.91	13	22.37

Images in rank order for viewing condition B

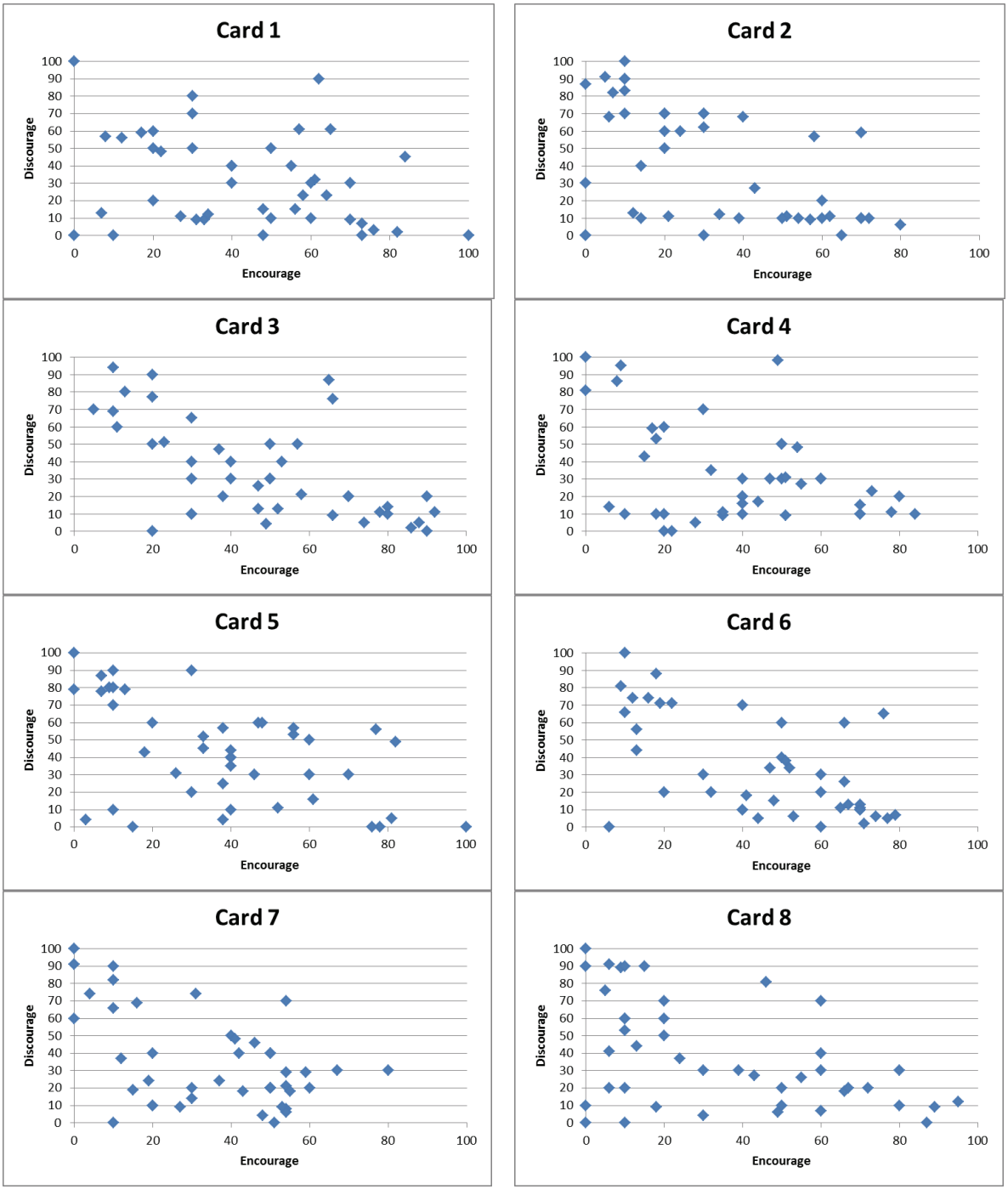
TimeB	Question a		Question b		Question c		Question d		Question e		Question f	
Rank	Website	Score	Website	Score	Website	Score	Website	Score	Website	Score	Website	Score
1	10	72.23	1	61.82	10	54.05	10	56.78	10	55.00	23	58.37
2	14	65.39	10	59.70	1	53.60	1	53.95	11	50.79	20	48.86
3	1	64.59	11	57.53	14	53.06	11	51.87	1	50.58	22	44.19
4	8	57.65	14	56.32	11	50.28	15	51.80	9	50.56	4	41.19
5	15	56.25	19	54.28	15	48.59	6	49.69	14	49.65	18	39.06
6	21	56.13	8	53.50	9	48.49	9	49.28	15	47.03	21	38.35
7	6	54.72	6	52.15	19	48.38	14	49.03	6	46.44	12	37.56
8	11	53.68	15	51.33	8	46.90	3	46.82	8	44.45	5	35.09
9	24	50.10	13	51.20	6	46.87	19	46.63	2	44.03	7	34.94
10	18	49.79	9	50.80	13	46.51	13	46.03	19	42.97	16	33.35
11	9	49.18	3	50.44	3	42.18	8	46.00	24	42.61	24	33.12
12	19	46.75	24	48.62	2	41.78	2	45.10	13	42.43	3	32.31
13	2	46.46	2	48.31	24	41.28	24	43.97	3	40.31	6	31.76
14	13	46.28	17	44.03	5	40.24	5	42.67	12	39.16	14	31.21
15	5	42.77	21	41.54	12	39.66	12	42.61	18	38.94	15	30.59
16	4	41.66	5	41.13	18	39.53	17	41.70	17	38.54	8	29.73
17	17	40.25	7	41.00	21	39.51	21	41.66	16	37.68	13	29.42
18	3	39.85	16	40.05	17	39.45	16	41.30	21	37.26	10	29.03
19	23	39.49	12	39.13	16	39.11	7	41.26	5	36.74	17	28.69
20	16	38.69	18	38.67	7	37.00	18	40.00	7	35.32	11	28.31
21	12	36.53	4	33.37	4	30.51	22	34.19	4	29.70	2	28.16
22	7	36.05	22	31.84	22	28.92	4	31.17	22	26.16	19	28.06
23	20	33.37	20	28.23	20	26.03	20	26.39	20	26.00	1	24.90
24	22	24.00	23	27.77	23	24.23	23	23.62	23	23.51	9	24.32

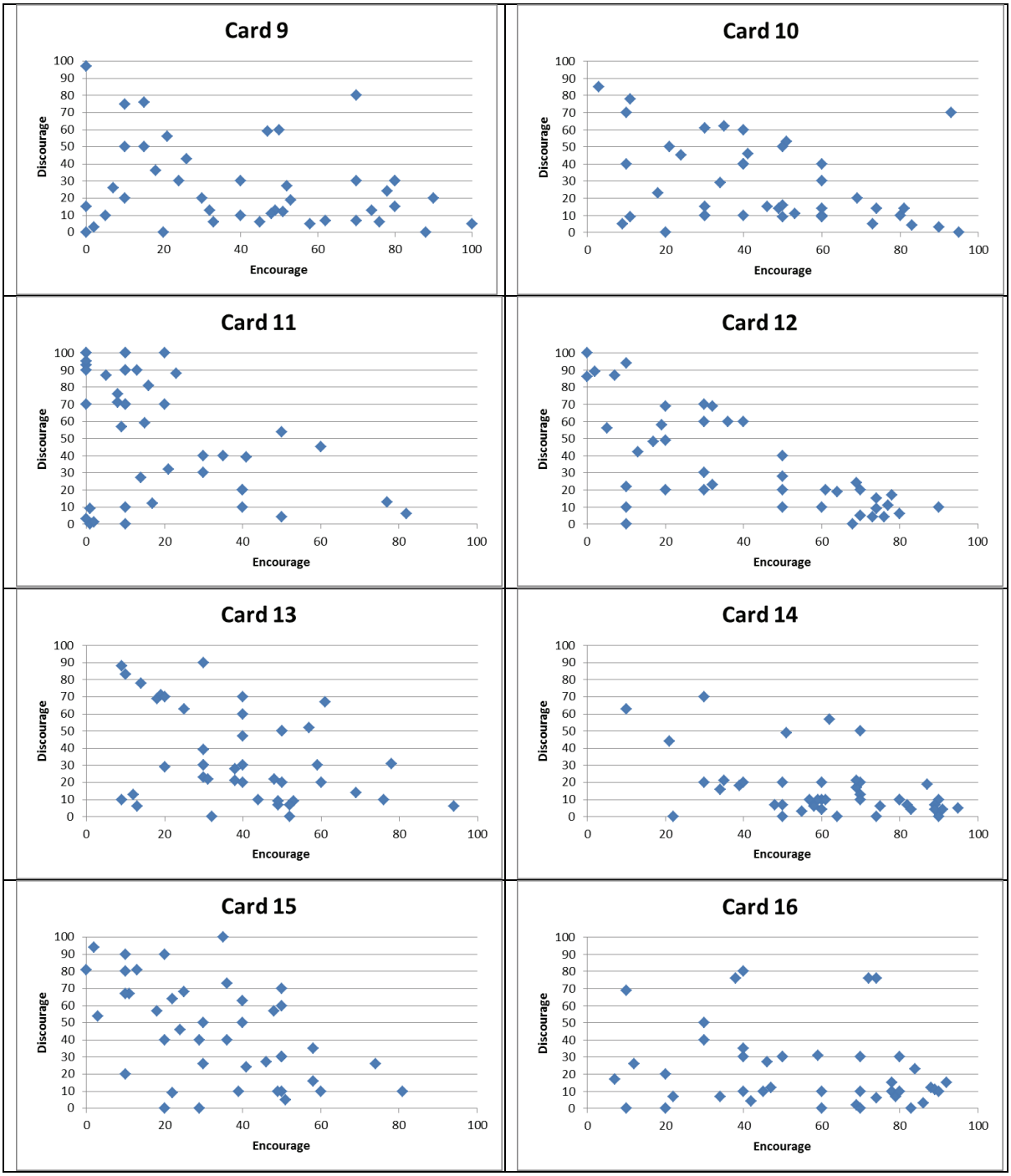
Appendix M Study Two mean subjective ratings by gender

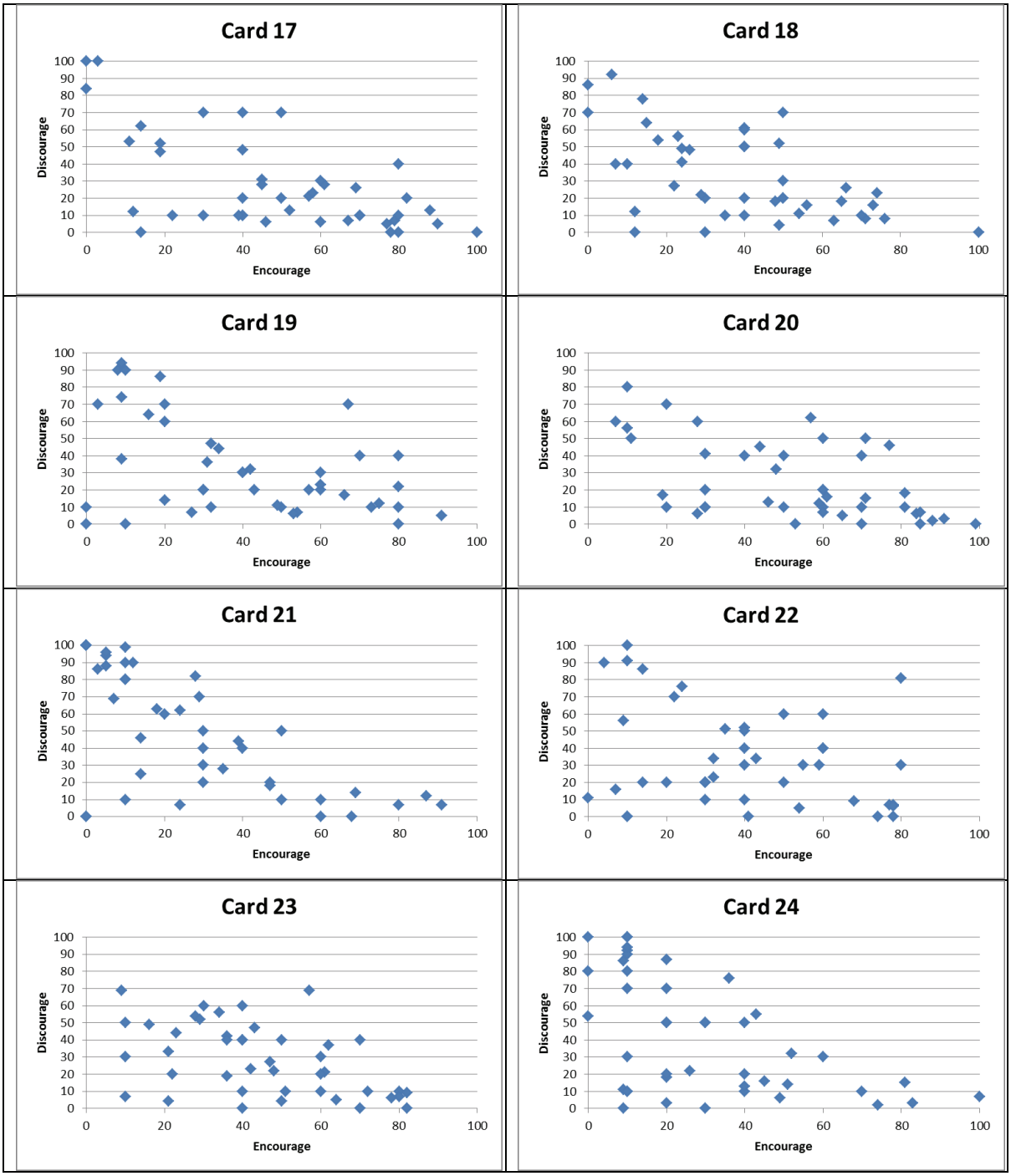
Mean subjective ratings from females						
Card ID	Mean QA	Mean QB	Mean QC	Mean QD	Mean QE	Mean QF
1	57.86	53.18	42.64	40.23	39.23	43.95
2	26.05	27.82	30.45	37.82	27.95	42.23
3	63.64	60.14	52.64	52.36	52.41	30.86
4	39.09	35.45	38.86	42.77	41.95	37.50
5	42.05	36.95	30.55	28.77	31.41	45.95
6	43.86	39.00	45.50	47.68	45.36	32.09
7	36.09	33.82	35.41	34.50	34.64	32.91
8	24.09	27.45	33.05	36.05	31.82	43.32
9	38.50	34.27	41.77	41.41	42.41	30.68
10	62.00	55.86	55.27	53.50	55.73	24.95
11	19.27	16.09	21.23	20.09	19.36	48.59
12	39.00	38.68	38.23	38.45	39.45	42.18
13	43.59	38.68	42.41	39.55	40.59	30.64
14	78.82	70.77	66.91	62.86	70.36	13.55
15	37.64	36.55	34.23	31.95	39.59	44.32
16	58.27	54.82	56.18	49.73	53.14	28.09
17	61.68	49.91	52.41	53.05	54.86	22.36
18	43.09	39.09	44.00	46.23	47.82	34.27
19	43.50	38.05	35.32	36.18	31.23	42.41
20	65.73	60.45	56.64	58.45	62.68	20.86
21	45.00	35.95	30.50	28.36	31.86	51.36
22	46.77	51.55	45.68	44.14	44.50	30.59
23	51.64	47.55	44.59	46.50	45.36	32.23
24	46.05	41.00	36.59	33.41	32.09	37.73
25	46.18	46.82	38.09	33.23	39.77	38.55
26	21.00	18.32	18.50	22.45	22.18	58.09
27	61.41	61.95	54.95	57.55	59.36	22.64
28	59.73	55.45	46.14	45.45	48.82	32.73
29	54.91	48.14	43.14	41.41	39.27	31.32
30	62.86	45.32	43.14	44.18	46.59	36.14
Mean	47.31	43.30	41.83	41.61	42.39	35.44
St Dev	14.05	12.55	10.54	10.16	11.50	9.59

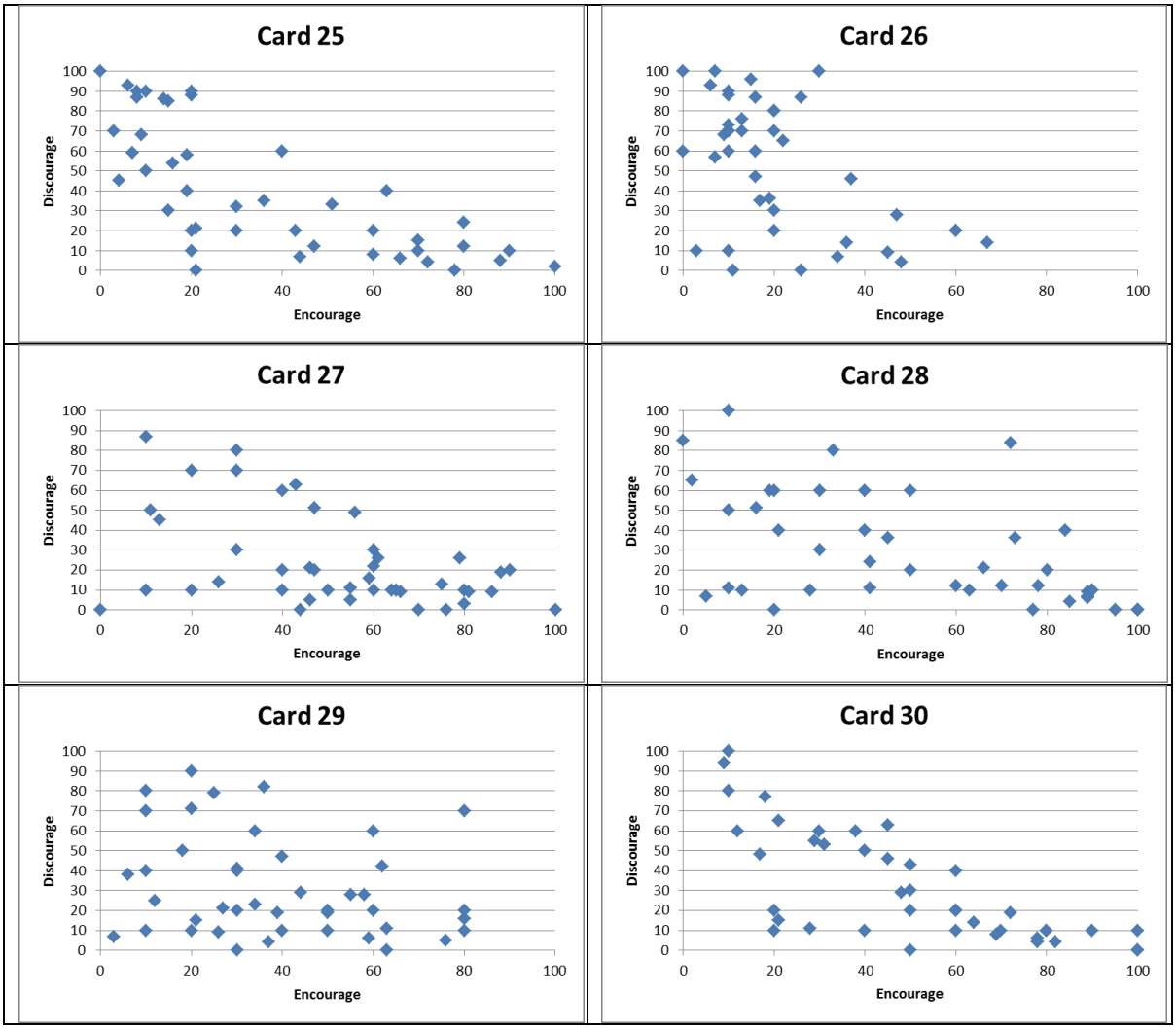
Mean subjective ratings from males						
Card ID	Mean QA	Mean QB	Mean QC	Mean QD	Mean QE	Mean QF
1	51.65	45.45	44.10	45.00	43.55	19.80
2	34.05	36.50	35.40	42.60	39.35	34.90
3	63.80	46.00	42.05	43.20	55.65	40.60
4	45.60	46.35	41.70	47.05	51.15	24.50
5	58.65	54.15	48.80	47.85	53.45	37.45
6	37.05	39.65	47.30	50.70	51.00	33.10
7	35.55	39.50	36.85	41.90	35.00	36.15
8	30.00	32.10	37.25	47.35	33.75	34.90
9	35.80	38.40	42.40	47.75	43.10	23.15
10	39.00	37.45	35.40	44.80	42.65	33.75
11	15.85	16.90	19.05	19.00	23.35	54.65
12	37.05	39.30	43.40	47.40	42.15	28.55
13	37.60	42.15	38.35	33.65	39.10	41.00
14	59.85	63.10	56.15	61.35	56.80	19.40
15	35.15	38.50	33.75	35.65	40.35	46.30
16	58.90	57.65	53.10	59.50	60.65	17.45
17	61.25	47.90	47.70	48.20	59.70	30.25
18	38.30	43.25	35.65	48.60	50.10	31.15
19	44.95	46.65	44.45	44.80	45.25	23.30
20	49.05	47.40	49.75	52.50	50.70	30.70
21	53.55	26.10	24.55	21.70	33.40	46.15
22	36.75	37.95	31.75	35.15	39.70	39.10
23	46.45	41.80	46.90	42.35	50.55	24.20
24	27.20	25.55	23.10	20.25	23.75	43.45
25	49.60	43.60	37.25	36.20	40.40	38.05
26	25.60	21.55	22.75	34.15	33.05	50.25
27	54.90	54.35	49.50	50.65	56.10	22.40
28	56.35	61.75	55.70	59.40	57.50	27.15
29	40.70	39.60	36.80	31.05	40.40	32.95
30	58.85	53.65	57.30	55.80	50.35	25.95
Mean	43.97	42.14	40.61	43.19	44.73	33.02
St Dev	11.93	10.71	9.86	10.71	9.76	9.35

Appendix N Scatter graphs for encourage and discourage from Study Two




Card 9
Discourage
Encourage
Card 10
Discourage
Encourage
Card 11
Discourage
Encourage
Card 12
Discourage
Encourage
Card 13
Discourage
Encourage
Card 14
Discourage
Encourage
Card 15
Discourage
Encourage
Card 16
Discourage
Encourage


Card 17
Card 18
Card 19
Card 20
Card 21
Card 22
Card 23
Card 24
Discourage
Encourage
0
10
20
30
40
50
60
70
80
90
100


Card 25
Discourage
Encourage
Card 26
Discourage
Encourage
Card 27
Discourage
Encourage
Card 28
Discourage
Encourage
Card 29
Discourage
Encourage
Card 30
Discourage
Encourage

Appendix O Invitation to take part Study Two: Card sorting and laddering activities

I'm doing my PhD in Computer Science at Keele University. I am conducting a card sorting activity and a laddering activity. The card sorting activity involves sorting a set of 30 images of hot air balloon websites by criteria and into categories of your choosing. I will then record the criteria used and which cards were assigned to each category. This task will take no more than 1 hour. If you are able to spare 90 minutes the card sorting activity will be followed by a laddering activity where I will ask you some further questions relating to your sorting such as the reasons for choosing the criteria for sorting, the categories used and the allocation of cards to them. The laddering will take no more than 30 minutes.

If you are willing to help please get in touch so we can arrange a convenient time and place. Please include in your response whether you are available for 1 hour or 90 minutes.

Thanks,

Nikki

My contact details:

Nikki Williams
CR 40
Colin Reeves Building
Keele University
Staffordshire
ST5 5BG
E-mail: n.k.williams@keele.ac.uk
Tel: 01782 733253

Contact details for the Research Governance Officer at Keele University (if you do not wish to contact the researcher directly):

Nicola Leighton
Research Governance Officer
Research & Enterprise Services
Dorothy Hodgkin Building
Keele University
ST5 5BG
E-mail: n.leighton@uso.keele.ac.uk
Tel: 01782 733306

Appendix P Consent form for Study Two: Card sorting and laddering activities

CONSENT FORM

Title of Project: A study of website aesthetics Part A
Name of Principal Investigator: Nikki Williams

Please tick box

1	I confirm that I have read and understand the information sheet for the above study and have had the opportunity to ask questions.	☐
2	I understand that my participation is voluntary and that I am free to withdraw at any time.	☐
3	I agree to take part in this study.	☐
4	I understand that data collected about me during this study will be made anonymous before publication.	☐
5	I agree to allow the data collected to be used for future research projects.	☐
6	I agree to be contacted about possible participation in future research projects.	☐

Contact Details

Nikki Williams
CR 40
Colin Reeves Building
Keele University
Staffordshire
ST5 5BG
E-mail: n.k.williams@keele.ac.uk
Tel: 01782 733253

Appendix Q Information sheet for Study Two: Card sorting and laddering activities

Information Sheet

Study Title: A study of website aesthetics Part B

Aims of the Research
This research aims to identify predictors for users' opinions of websites. In the first part of the study users' opinions need to be gathered along with various objective measures. These objective measures include, but are not limited to, the word count, the colour balance on the page, the number of images, the background colour of the page, and the most used colour within the page. These results will then be analysed for correlations, to see if predictors exist. In the second part, information regarding user's grouping of websites will be gathered, analysed and used to complement the first section findings.

Invitation
You are being invited to consider taking part in the research study: A study of website aesthetics. This project is being undertaken by Nikki Williams (a PhD student within Computing at Keele University).

Before you decide whether or not you wish to take part, it is important for you to understand why this research is being done and what it will involve. Please take time to read this information carefully and discuss it with friends and relatives if you wish. Ask us if there is anything that is unclear or if you would like more information.

Why have I been chosen?
Participants have been gathered by inviting a group of people to take part, then asking them to invite their friends, and so on. The mediums for delivering these invitations include using email and social networking. There are no restrictions on who can take part, except that you do not speak Finnish to a level where you could understand written content at a glance.

Do I have to take part?
You are free to decide whether you wish to take part or not. If you do decide to take part you will be asked to sign two consent forms, one is for you to keep and the other is for our records. You are free to withdraw from this study at any time and without giving reasons.

What will happen if I take part?
Participants will be asked to complete a brief background questionnaire. Participants will then be asked to sort 30 images depicting web pages for hot air balloon companies into groups of your choice. The sorting activity will take no more than 60 minutes. For participants available for up to 90 minutes some further questions will be asked regarding the reasons behind the groupings.

If I take part, what do I have to do?
You will be expected to complete the card-sorting task.

What are the benefits (if any) of taking part?
There are no obvious benefits to participants.

What are the risks (if any) of taking part?
There are no obvious risks to participants.

What if there is a problem?

If you have a concern about any aspect of this study, you may wish to speak to the researcher who will do their best to answer your questions. You should contact Nikki Williams on n.k.williams@keele.ac.uk. Alternatively, if you do not wish to contact the researcher you may contact Nicola Leighton, Research Governance Officer on 01782 733306 or n.leighton@uso.keele.ac.uk

If you remain unhappy about the research and/or wish to raise a complaint about any aspect of the way that you have been approached or treated during the course of the study please write to Nicola Leighton who is the University's contact for complaints regarding research at the following address:-

Nicola Leighton
Research Governance Officer
Research & Enterprise Services
Dorothy Hodgkin Building
Keele University
ST5 5BG
E-mail: n.leighton@uso.keele.ac.uk
Tel: 01782 733306

How will information about me be used?
The data will be collected electronically and stored on a database. The data will be analysed using statistical software. The data will be retained for use in the second part of this study, and potentially for future studies for up to five years after this study. Data will be made anonymous before analysis, with names being stored in a separate, password protected, file. The background information provided will be used to group results.

Who will have access to information about me?
All personal details *(eg consent forms)* and data obtained relating to the research will be stored on a password protected computer. The data collected will be anonymous. Your personal details will remain confidential and will not be linked to your responses. In accordance with Keele University Guidelines all data and consent forms will be retained by the researcher for 5 years after which they will be securely disposed of.

Who is funding and organising the research?
The research is being undertaken as part of a PhD in Computer Science.

Contact for further information

Nikki Williams
CR 40
Colin Reeves Building
Keele University
Staffordshire
ST5 5BG
E-mail: n.k.williams@keele.ac.uk
Tel: 01782 733253

Gordon Rugg
CR 102
Colin Reeves Building
Keele University
Staffordshire
ST5 5BG
Email: g.rugg@cs.keele.ac.uk
Tel: 01782 733410

Appendix R Background questions for study 2

Background Questions

Participant ID _____

What is your gender?

☐Male

☐Female

☐Other / Prefer not to say

What is your level of web design experience?

☐Never designed a website

☐Designed a few web sites (1-10)

☐Designed many websites

Do you speak Finnish to a level where you could understand written content at a glance?

☐Yes

☐No

How familiar do you consider yourself to be with hot air balloons?

☐Not at all familiar

☐Slightly familiar

☐Somewhat familiar

☐Moderately familiar

☐Extremely familiar

Appendix S Card sort instructions

Card Sort Instructions

You will be given some cards to sort. Each card will have a picture of a web page screenshot on it.

I would like you to sort the cards into groups, using one criterion at a time. When you have finished sorting, please tell me what the criterion was for that sort, and what the groups were into which you sorted the cards, so that I can record this. Once this has been done, I would like you to sort the cards again, using a different criterion, and then to keep on sorting them until you have run out of criteria.

For example, if the task was sorting different types of car, your first criterion might be 'place of manufacture' and the groups might be 'American', 'British', 'French' etc.; the second criterion might be 'cost', with the groups being 'expensive', 'medium' and 'cheap'.

You are welcome to use any criteria you like, and any groups you like, including 'don't know', 'not sure' and 'not applicable'. The main thing is to use only one criterion in each sort – please don't lump two or more in together. If you're not sure about something, just ask.

You may have noticed that the cards are numbered: this is for convenience when recording the results. The numbering is random, so please don't use that as a criterion for sorting!

If you have any comments or questions, then please say, and I will sort them out.

Thank you for your help.

Appendix T Standard phrasings for laddering

Used for every sort:

Which category do you prefer and why?

How important is the criterion to you when looking at web pages?

Additional phrases relating to sort information:

Why did you place this card into the ----- category?

What are the features that made this card -----?

What was it about this card that made you put it in the ----- category?

Could you tell me what about a card would lead you to place it in the ----- category?

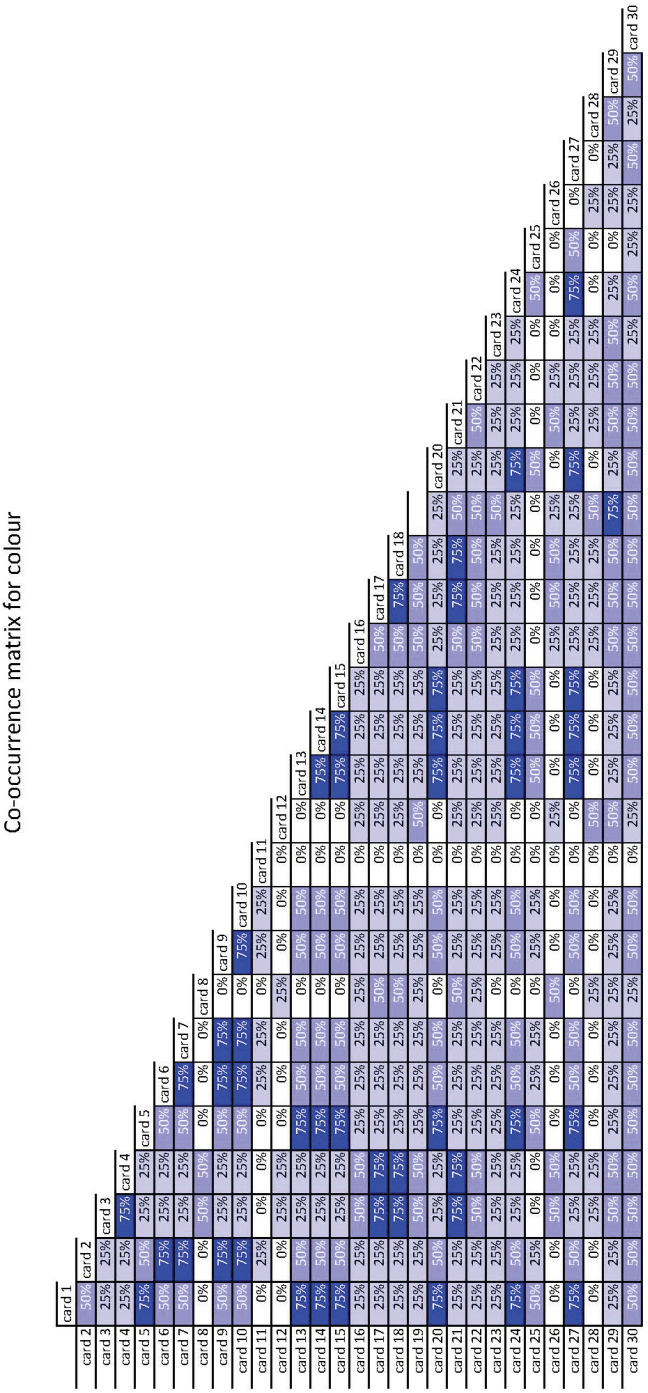
Unrelated to sort information:

Which of these two web pages do you prefer and why?

Sample of card sort responses

SortID	Participant	Sort	Criteria	Category including preference	Why?
86	15	1	Professionalism of design of site	*Professionally designed	Look nicer, look more functional and easier to use
86	15	1	Professionalism of design of site	Amateur designed	
87	15	2	Focussed on selling balloon trips?	Selling focussed	
87	15	2	Focussed on selling balloon trips?	Information focussed	
87	15	2	Focussed on selling balloon trips?	*Can't really tell / not sure	Like the fact I'm not being sold something outright. Shows information and selling. Information based only is negative as they seem less focussed which implies less professional and I wouldn't buy from them
88	15	3	How air balloony they felt (do the aesthetics relate to the topic)	*Very balloony	Makes the topic seem interesting. Engages me with the topic in a positive way. Sites are well designed for balloons (fit for purpose)
88	15	3	How air balloony they felt (do the aesthetics relate to the topic)	Somewhat balloony	
88	15	3	How air balloony they felt (do the aesthetics relate to the topic)	Not very balloony	
88	15	4	How much blue is in the page	Lots of blue	
88	15	4	How much blue is in the page	*Blue and white (50-50)	Felt it portrayed sky best, suited the topic best. Nice to break the blue up with another appropriate colour.
88	15	4	How much blue is in the page	Not a lot of blue	

Appendix V Co-occurrence matrices



Co-occurrence matrix for emotion producing

	card 1																												
card 2	100%	card 2																											
card 3	50%	50%	card 3																										
card 4	100%	100%	50%	card 4																									
card 5	100%	100%	50%	100%	card 5																								
card 6	50%	50%	0%	50%	50%	card 6																							
card 7	50%	50%	0%	50%	50%	100%	card 7																						
card 8	50%	50%	0%	50%	50%	100%	100%	card 8																					
card 9	50%	50%	0%	50%	50%	100%	100%	100%	card 9																				
card 10	50%	50%	100%	50%	50%	0%	0%	0%	0%	card 10																			
card 11	100%	100%	50%	100%	100%	50%	50%	50%	50%	50%	card 11																		
card 12	50%	50%	0%	50%	50%	100%	100%	100%	100%	0%	50%	card 12																	
card 13	50%	50%	0%	50%	50%	50%	50%	50%	50%	0%	50%	50%	card 13																
card 14	0%	0%	50%	0%	0%	50%	50%	50%	50%	50%	0%	50%	0%	card 14															
card 15	100%	100%	50%	100%	100%	50%	50%	50%	50%	50%	100%	50%	50%	0%	card 15														
card 16	0%	0%	50%	0%	0%	0%	0%	0%	0%	50%	0%	0%	50%	50%	0%	card 16													
card 17	0%	0%	0%	0%	0%	50%	50%	50%	50%	0%	0%	50%	0%	50%	0%	0%	card 17												
card 18	100%	100%	50%	100%	100%	50%	50%	50%	50%	50%	100%	50%	50%	0%	100%	0%	0%	card 18											
card 19	100%	100%	50%	100%	100%	50%	50%	50%	50%	50%	100%	50%	50%	0%	100%	0%	0%	100%	card 19										
card 20	50%	50%	50%	50%	50%	0%	0%	0%	0%	50%	50%	0%	0%	0%	50%	0%	0%	50%	50%	card 20									
card 21	100%	100%	50%	100%	100%	50%	50%	50%	50%	50%	100%	50%	50%	0%	100%	0%	0%	100%	100%	50%	card 21								
card 22	50%	50%	0%	50%	50%	50%	50%	50%	50%	0%	50%	50%	100%	0%	50%	50%	0%	50%	50%	0%	50%	card 22							
card 23	50%	50%	100%	50%	50%	0%	0%	0%	0%	100%	50%	0%	0%	50%	50%	50%	0%	50%	50%	50%	50%	0%	card 23						
card 24	100%	100%	50%	100%	100%	50%	50%	50%	50%	50%	100%	50%	50%	0%	100%	0%	0%	100%	100%	50%	100%	50%	50%	card 24					
card 25	100%	100%	50%	100%	100%	50%	50%	50%	50%	50%	100%	50%	50%	0%	100%	0%	0%	100%	100%	50%	100%	50%	50%	100%	card 25				
card 26	50%	50%	0%	50%	50%	100%	100%	100%	100%	0%	50%	100%	50%	50%	50%	0%	50%	50%	50%	0%	50%	50%	0%	50%	50%	card 26			
card 27	100%	100%	50%	100%	100%	50%	50%	50%	50%	50%	100%	50%	50%	0%	100%	0%	0%	100%	100%	50%	100%	50%	50%	100%	100%	50%	card 27		
card 28	50%	50%	100%	50%	50%	0%	0%	0%	0%	100%	50%	0%	0%	50%	50%	50%	0%	50%	50%	50%	50%	0%	100%	50%	50%	0%	50%	card 28	
card 29	50%	50%	100%	50%	50%	0%	0%	0%	0%	100%	50%	0%	0%	50%	50%	50%	0%	50%	50%	50%	50%	0%	100%	50%	50%	0%	50%	100%	card 29
card 30	100%	100%	50%	100%	100%	50%	50%	50%	50%	50%	100%	50%	50%	0%	100%	0%	0%	100%	100%	50%	100%	50%	50%	100%	100%	50%	100%	50%	50%

Co-occurrence matrix for balloon image

	card 1																												
card 2	25%	card 2																											
card 3	25%	25%	card 3																										
card 4	50%	25%	50%	card 4																									
card 5	50%	50%	50%	50%	card 5																								
card 6	25%	75%	50%	50%	50%	card 6																							
card 7	75%	25%	50%	75%	50%	50%	card 7																						
card 8	50%	25%	50%	100%	50%	50%	75%	card 8																					
card 9	50%	25%	50%	100%	50%	50%	75%	100%	card 9																				
card 10	25%	25%	75%	25%	50%	25%	25%	25%	25%	card 10																			
card 11	50%	25%	50%	100%	50%	50%	75%	100%	100%	25%	card 11																		
card 12	75%	25%	25%	50%	50%	25%	75%	50%	50%	25%	50%	card 12																	
card 13	50%	25%	50%	100%	50%	50%	75%	100%	100%	25%	100%	50%	card 13																
card 14	25%	0%	75%	75%	25%	25%	50%	75%	75%	50%	75%	25%	75%	card 14															
card 15	50%	25%	25%	75%	50%	25%	50%	75%	75%	25%	75%	50%	75%	50%	card 15														
card 16	25%	25%	75%	25%	50%	25%	25%	25%	25%	75%	25%	25%	25%	50%	25%	card 16													
card 17	25%	0%	25%	50%	25%	0%	25%	50%	50%	25%	50%	25%	50%	50%	50%	25%	card 17												
card 18	50%	25%	50%	100%	50%	50%	75%	100%	100%	25%	100%	50%	100%	75%	75%	25%	50%	card 18											
card 19	75%	25%	25%	50%	50%	25%	75%	50%	50%	25%	50%	75%	50%	25%	50%	25%	25%	50%	card 19										
card 20	0%	25%	25%	50%	0%	50%	25%	50%	50%	0%	50%	0%	50%	50%	25%	0%	25%	50%	0%	card 20									
card 21	50%	50%	0%	25%	25%	50%	50%	25%	25%	0%	25%	50%	25%	0%	25%	0%	0%	25%	50%	25%	card 21								
card 22	50%	50%	50%	50%	75%	50%	50%	50%	50%	50%	50%	50%	50%	25%	50%	50%	25%	50%	50%	0%	50%	card 22							
card 23	25%	0%	50%	50%	25%	0%	25%	50%	50%	50%	50%	25%	50%	75%	50%	50%	50%	50%	25%	25%	25%	50%	card 23						
card 24	50%	25%	25%	75%	50%	25%	50%	75%	75%	25%	75%	50%	75%	50%	75%	25%	50%	75%	50%	25%	25%	50%	50%	card 24					
card 25	75%	25%	50%	75%	50%	50%	100%	75%	75%	25%	75%	75%	75%	50%	50%	25%	25%	75%	75%	25%	50%	50%	25%	50%	card 25				
card 26	50%	25%	25%	50%	25%	50%	75%	50%	50%	0%	50%	50%	50%	25%	25%	0%	0%	50%	50%	25%	50%	25%	0%	25%	75%	card 26			
card 27	50%	25%	25%	75%	50%	25%	50%	75%	75%	25%	75%	50%	75%	50%	75%	25%	50%	75%	50%	25%	25%	50%	50%	100%	50%	25%	card 27		
card 28	50%	0%	50%	25%	25%	0%	50%	25%	25%	50%	25%	50%	25%	50%	25%	50%	25%	25%	50%	0%	50%	50%	75%	25%	50%	25%	25%	card 28	
card 29	50%	0%	50%	25%	25%	0%	50%	25%	25%	50%	25%	50%	25%	50%	25%	50%	25%	25%	50%	0%	25%	25%	50%	50%	50%	25%	50%	75%	card 29
card 30	50%	50%	75%	75%	75%	75%	75%	75%	75%	50%	75%	50%	75%	50%	50%	50%	25%	75%	50%	25%	25%	75%	25%	50%	75%	50%	50%	25%	25%	card 30

Co-occurence matrix for website design

	card 1	card 2	card 3	card 4	card 5	card 6	card 7	card 8	card 9	card 10	card 11	card 12	card 13	card 14	card 15	card 16	card 17	card 18	card 19	card 20	card 21	card 22	card 23	card 24	card 25	card 26	card 27	card 28	card 29	card 30
card 2	47%																													
card 3	76%	41%																												
card 4	47%	59%	29%																											
card 5	71%	47%	76%	47%																										
card 6	29%	47%	29%	47%	35%																									
card 7	35%	35%	35%	53%	47%	59%																								
card 8	18%	24%	24%	35%	35%	41%	53%																							
card 9	29%	29%	47%	47%	53%	59%	71%	65%																						
card 10	59%	59%	71%	47%	76%	41%	53%	47%	59%																					
card 11	41%	35%	35%	53%	53%	29%	59%	35%	53%	35%																				
card 12	29%	24%	29%	41%	24%	35%	47%	47%	47%	29%	47%																			
card 13	47%	76%	41%	65%	47%	41%	41%	24%	29%	59%	41%	29%																		
card 14	29%	29%	47%	53%	41%	29%	47%	35%	59%	47%	41%	47%	35%																	
card 15	53%	65%	41%	59%	53%	35%	35%	18%	24%	47%	53%	29%	71%	35%																
card 16	41%	59%	41%	71%	47%	35%	47%	47%	59%	65%	47%	41%	59%	59%	47%															
card 17	41%	47%	53%	41%	47%	41%	59%	47%	59%	65%	35%	41%	41%	65%	41%	59%														
card 18	47%	59%	47%	53%	65%	59%	59%	53%	59%	71%	41%	35%	59%	35%	59%	47%	59%													
card 19	71%	35%	65%	47%	59%	24%	29%	35%	41%	53%	35%	47%	35%	41%	41%	53%	47%	41%												
card 20	76%	47%	82%	41%	76%	24%	35%	24%	47%	71%	47%	18%	47%	47%	47%	53%	53%	47%	59%											
card 21	76%	53%	65%	41%	71%	24%	29%	35%	29%	65%	41%	29%	53%	18%	59%	47%	35%	53%	65%	65%										
card 22	47%	53%	47%	47%	53%	41%	41%	35%	41%	59%	35%	41%	53%	47%	53%	41%	53%	76%	41%	47%	41%									
card 23	47%	47%	53%	59%	71%	29%	53%	41%	59%	65%	53%	35%	47%	65%	41%	59%	53%	59%	47%	53%	47%	59%								
card 24	82%	53%	71%	47%	76%	24%	35%	29%	35%	71%	47%	29%	53%	35%	59%	53%	53%	59%	71%	82%	76%	59%	53%							
card 25	76%	35%	71%	29%	65%	12%	18%	24%	24%	47%	41%	29%	35%	24%	41%	29%	29%	35%	59%	71%	76%	35%	41%	71%						
card 26	29%	41%	29%	35%	35%	59%	47%	41%	47%	41%	35%	47%	35%	29%	35%	29%	53%	65%	29%	29%	24%	59%	29%	41%	24%					
card 27	53%	47%	53%	65%	59%	35%	65%	35%	53%	71%	47%	29%	53%	53%	41%	65%	59%	47%	53%	59%	47%	35%	59%	53%	35%	24%				
card 28	59%	29%	71%	12%	53%	18%	24%	29%	29%	59%	24%	41%	29%	35%	41%	29%	53%	41%	53%	59%	59%	41%	29%	59%	59%	29%	35%			
card 29	71%	53%	71%	41%	76%	35%	41%	29%	35%	71%	41%	35%	53%	29%	59%	47%	47%	59%	65%	59%	76%	47%	53%	71%	59%	35%	53%	59%		
card 30	53%	47%	59%	41%	47%	53%	47%	29%	47%	59%	18%	24%	47%	47%	47%	35%	65%	65%	35%	53%	35%	59%	41%	47%	35%	47%	47%	53%	41%	

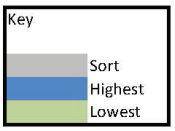
Appendix W Original criteria and categories for overall impression

Original Criteria	Original categories	Gist categories
Aesthetic impression	Boring	Not a good impression
	Interesting	Good impression
How aesthetically appealing the logo was	Not great	Not a good impression
	No logo or bad logo	Not a good impression
	The best	Good impression
	Good logo but not great	Average impression
How pretty the page is (level of effort put into the design)	Not a very good design	Not a good impression
	Close to terrible	Not a good impression
	Prettiest	Good impression
	Sleek and pretty	Good impression
	Pretty average for a reasonable website	Average impression
Impression of level of enjoyment from activity	Bad time	Not a good impression
	Great time	Good impression
	Average time	Average impression
Being aesthetically pleasing	Is not aesthetically pleasing	Not a good impression
	Is aesthetically pleasing and enjoyable to look at	Good impression
Immersive impression?(makes me feel like being in a balloon)	Is not immersive	Not a good impression
	It is immersive	Good impression
Ones I like or don't like	Ones I don't like	Not a good impression
	Ones I do like	Good impression
Does it make the activity look exciting?	It doesn't make it look exciting	Not a good impression
	It makes it look exciting	Good impression
Aesthetic opinion	Ones I don't like	Not a good impression
	Ones I like	Good impression
How air balloony they felt (do the aesthetics relate to the topic)	Not very balloony	Not a good impression
	Very balloony	Good impression
	Somewhat balloony	Average impression
Whether the pictures look exciting or not	Not very exciting	Not a good impression
	Exciting pictures	Good impression
First impression of site	Willing to stay on site	I would use the site
	Not willing to stay on site	I would not use the site

Appendix X Card sort criteria coded for importance

Positioning of criteria by importance

Participant	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11
P1											
P2											
P3											
P4											
P5											
P6											
P7											
P8											
P9											
P10											
P11											
P12											
P13											
P14											
P15											
P16											
P17											
P18											



Independent Judge 1

Participant	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11
P1											
P2											
P3											
P4											
P5				?	?						
P6											
P7											
P8											
P9											
P10											
P11											
P12											
P13											
P14											
P15											
P16											
P17											
P18											

Independent judge 2

Participant	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11
P1											
P2											
P3											
P4											
P5											
P6											
P7											
P8											
P9											
P10											
P11											
P12											
P13											
P14											
P15											
P16											
P17											
P18											