

# Does Textual Word-of-Mouth Affect Look and Feel?

**Milica Stojmenovic**

Carleton University  
Ottawa, Canada  
milica.stojmenovic@carleton.ca

**John Grundy**

Deakin University  
Melbourne, Australia  
j.grundy@deakin.edu.au

**Vivienne Farrell**

Swinburne University of Technology  
Melbourne, Australia  
vfarrell@swin.edu.au

**Robert Biddle**

Carleton University  
Ottawa, Canada  
robert.biddle@carleton.ca

**Leonard Hoon**

Swinburne University of Technology  
Melbourne, Australia  
lhoon@swin.edu.au

## ABSTRACT

In the field of HCI, website usability and visual appeal have been studied extensively. Participant experience with a website genre influences the use and perception of the website. Word-of-Mouth (WOM), such as user reviews, influences users in hotel, restaurant, movie, and many other e-commerce domains. Thus, a company's or product's reputation can alter a consumer's behaviour towards that product. Our work aimed to acquire an understanding of the effect of textual WOM on usability and visual appeal. This is a novel approach to the topic. This research was undertaken using an unfamiliar city council website to exclude the influence of one's own past experiences and to allow for greater control of the textual WOM. We found that visual appeal, objective and subjective usability were all influenced by text that established reputations.

## Author Keywords

Visual appeal, aesthetics, usability, websites, written/textual WOM, reputation.

## ACM Classification Keywords

Design, Human Factors, Performance

## INTRODUCTION

The Internet offers major opportunities for competitive advantage in ecommerce, and provides a replacement for paper-based documents and personal services to the public, for those who 'get it right'. Poor websites will easily lose the attention of the user when alternatives exist. In particular, visual appeal and usability have long been identified as effectors of trust, enjoyment, and finally intentions to use websites. They have hence often been the subject of research. Yet, what factors influence the two are not known. This paper examines the effect of nuanced

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

OzCHI '16, November 29 – December 2 2016, Launceston, TAS, Australia

Copyright © 2016 ACM 978-1-4503-3673-4/15/12... \$15.00

<http://dx.doi.org/xx.xxxx/xxxxxxx.xxxxxxx>.

textual word-of-mouth (WOM), in the form of scenarios, on usability and visual appeal, in a non-commerce, gender- and age-neutral website genre where participants do not have highly developed mental models.

The three main concepts in this paper are WOM, visual appeal, and usability. WOM relays information between people and tends to be about experiences (Smith, 1993). In its initial definition, WOM entailed face-to-face communication (Granovetter, 1973). Since then, WOM was expanded to include textual communication, such as user reviews (ex. Smith, Menon, & Sivakumar, 2005). In this paper, the influence of WOM was examined on both visual appeal and usability. Visual appeal is linked with the concepts of aesthetics and beauty (Zettl, 1999). Aesthetic appraisal is the cognitive judgment of an object's aesthetic appearance (Blijlevens, 2011). This definition is used in this paper for visual appeal.

The definition of usability used here is the most widely used definition provided by the International Standards Organization (ISO). According to ISO 9241/11 (1996), "usability is the extent to which a given product can be used by a specific group of users, to achieve specific goals with effectiveness, efficiency, and satisfaction in a specific context of use". Effectiveness relates to how well users can achieve specific goals. Efficiency is the time taken to complete a given task. Satisfaction is the user's experience of acceptability. The context involves a predefined group of users, in set environments, who perform specific tasks.

The paper is outlined as follows. The next section gives a brief summary of related research. After that, an outline of the preliminary studies is presented. The method and results are then provided. Lastly, key findings, conclusions, and future work are discussed.

## BACKGROUND: USABILITY AND VISUAL APPEAL

It has been proposed that "what is beautiful is usable" (Tractinsky, Katz, & Ikar, 2000). The relationship between visual appeal and usability exists as there is a 'halo effect' where people and things are judged based on their appearance. Characteristics are assumed to be based on such judgments. In HCI, the halo effect has been applied to interfaces because beauty is a trait that is seen first,

significantly determines a user's first impression, and thus influences subsequent perceptions of characteristics (Tracktinsky et al., 2000).

Tracktinsky and colleagues (2000) examined the effects of aesthetics on a user's perception of quality of interaction with an Automatic Teller Machine (ATM). They found that ATM machine interface aesthetics affected both pre-use and post-use perceptions of usability. Satisfaction with using the ATM was strongly correlated with perceived aesthetics suggesting that aesthetic interfaces influence perceptions of quality as well as of performance. Tracktinsky and colleagues (2000) concluded that there is a strong relationship between a user's initial aesthetic perception and the perceived usability of a system, and that this relationship endures even after interacting with the system. However, recent literature demonstrates mixed findings with respect to the relationship between usability and aesthetics. For example Sangwon and Koubek (2010) found that user preference was significantly affected by aesthetics but marginally affected by usability before use. However, after use, user preference was significantly influenced by both usability and aesthetics. On a search engine website, Katz (2010) demonstrated significant correlations between perceived aesthetics and perceived usability and usefulness before system use, but not after. Tuch and colleagues (2012) found that aesthetics did not affect perceived usability; rather, usability affected perceived aesthetics after use, in an online shopping environment. The lack of consistency in the findings appears to indicate that the relationship between usability and aesthetics needs to be explored further; it may vary depending on the situation and/or familiarity with the genre being tested, or on WOM.

We tend to make assumptions that things like WOM will influence behaviour with websites because the halo effect exists, but evidence is sometimes lacking. Here, we examine the applicability of this phenomenon on websites. Textual WOM and its impact on visual appeal and usability is an unexplored area. The importance of examining products after prolonged use, when users are more experienced and have developed mental models, is often overlooked in studies, focusing instead on first impressions (Karapanos, 2013). An example of how an increased understanding of a website can influence ratings of visual appeal is in the health care information website genre, where expressive and classical aesthetics and credibility were rated significantly higher with repeated exposure (Setterstrom, 2010). It has also been found that increased exposure to a stimulus replaces the immediate emotional reaction with a more considered cognitive response (Ye & van Raaij, 1997). Therefore, familiarity with a particular genre may influence first impressions, as they create a threshold for what the individual's norm would be for the given genre. Thus it can influence aesthetics and subjective ratings of usability. Thus, we needed a genre that was unfamiliar to users, addressed in the preliminary studies.

## **Textual WOM**

Research on the impact of WOM on visual appeal and usability in websites is limited. The relevant literature found is summarized in this section.

Generally, having polarized descriptions of upcoming tasks can be considered biasing participants. Yet, this occurs in life: social media and user reviews tell us what products are good/bad (Smith et al., 2005). We speculate that positive or negative texts taint users and can alter their perception and interaction with websites.

Online marketplaces such as eBay incorporate both seller and buyer feedback into their business models. These reputations help both parties acquire trust in each other (e.g. Gefen et al., 2003). However, buyers cannot reliably trust or ascertain a seller's credibility and benevolence with just a numerical star rating. Instead, a much more reliable predictor is feedback left by previous buyers (Pavlou & Dimoka, 2006). One study found that people relied on peer and editorial reviews and recommendations more so than other means, such as paid ads, yet user reviews were seriously under-researched (Smith et al., 2005). Textual reviews can implicitly convey information about perceived quality, ease of use, and usefulness (Davis, 1989).

The overwhelming majority (97%) of users rely on the textual feedback left by previous buyers before proceeding to purchase something from an unknown seller (Pavlou & Dimoka, 2006). Textual feedback offers evidence of a seller's history which is used to predict the seller's future behaviour in transactions (Pavlou & Dimoka, 2006).

Hence, textual WOM can influence the actions of consumers. In this paper, it is investigated that positive or negative texts can alter the perception and use of websites (outside of the consumer domain). Thus, we examined if positive or negative WOM could impact users when examining a government website. To the best of our knowledge, no prior work has been done on this.

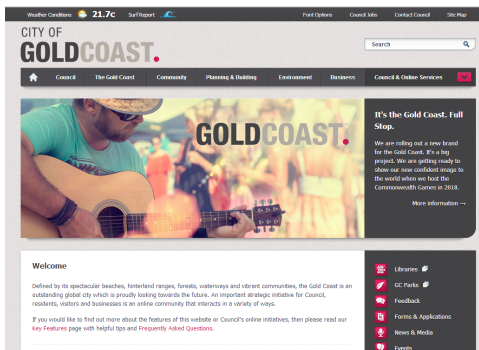
To examine the impact of textual WOM on usability and visual appeal, an unfamiliar city council website was used to exclude the influence of past experiences and to allow for greater control of reputation implementation. In this paper the term 'genre' means the category of website (e.g. bank, shop, etc.). The process of acquiring the website via the preliminary studies is described next.

## **Preliminary Studies**

Five preliminary studies were carried out. The purpose was twofold. First, it was to obtain a website genre that was unfamiliar to participants to enable the control of the textual WOM's impact. Second, it was to create a website dataset that statistically varied in visual appeal and usability. More details on all of the preliminary studies can be found in (Stojmenovic et al., 2014; Stojmenovic, 2016).

*Preliminary Study 1.* The first study was done to examine two different genres (tourism and city councils) in order to determine which genre was less familiar to participants. A

controlled laboratory experiment with 30 participants was done in which the collected data included visual appeal, usability, and familiarity questionnaires. Data was statistically analysed. The city council websites had more random results, indicating that they were less familiar. This was established by comparing results of the familiarity questionnaire (checklist of items that would appear on a city council or tourism website, and semantic differential scales for opinion of the genres) to actual website item counts and to their ratings of perceived usability and visual appeal of 52 websites. In addition to having inaccurate expectations of what items appear on city council websites, their answers indicated that city council websites would be cumbersome to use and uglier than the average website. Yet, their actual website ratings (upon viewing and rating several in each genre) showed that the highest rated in visual appeal was in fact a city council website. Thus, city council websites were chosen because they were deemed less familiar (so that reputations could be manipulated with greater ease) and the Gold Coast website was chosen because it was rated highest in visual appeal, seen in Fig. 1.



**Figure 1. The high visual appeal and usability website version (i.e. easy and pretty).**

*Preliminary Study 2.* Three researchers performed an expert based heuristic evaluation usability test to establish a usability level for the chosen website. The website was found to be easy to use, with no major issues.

*Preliminary Study 3.* This study was done to confirm that the website’s usability level was high with users. This was done via a controlled laboratory experiment with ten participants in a user-based usability test. Data collected included subjective visual appeal and perceived usability questionnaires, retrospective think aloud, and objective usability measures. The results confirmed the expert-based usability results; it was a usable website. Thus, the original Gold Coast city council website was deemed to be the high usability and high visual appeal website; henceforward referred to as the ‘good’ website version.

*Preliminary Study 4.* Phase four of the preliminary studies included the manipulation of the original good website to create a hard and ugly version. The good version was manipulated in five ways, two manipulations for usability and three for visual appeal. To lower usability, (1) the titles

in the top menu bar were changed. This was done by randomizing the title options, to alter the consistency and simplicity of the menu. With each click on the website, all of the titles in the menu bar would change to synonyms of the original title. For example, the “Council” tab would randomly change to: Board, Jury, Assembly, Committee, Congress, Politics, Law, or Government. To further lower the usability level, (2) item contrast in the dropdown hover menu was lowered by removing the contrast between the titles and the other menu options. This was done by un-bolding and un-underlining the titles, creating un-categorized lists under the menus.

The visual appeal was altered by (3) changing the main background colour from the original grey-beige to lilac and (4) changing the menu background from a grey-black to evergreen. The combination of lilac and evergreen was chosen since it was the colour scheme of the Toowoomba tourism website, found in the first preliminary study to be rated worst in visual appeal. The main text background was not changed to not alter the legibility. In addition, (5) the colours of all the images in the website were inverted to be negatives. Elements of classical aesthetics align themselves with CRAP (Contrast, Repetition, Alignment and Proximity) principles, so the structure of the layout was not changed in order to keep visual appeal and usability manipulations separate.

The hard to use (low usability) and ugly (low visual appeal) website version was tested with participants to ensure that the manipulations were sufficient and varied significantly from the original website. This was done via controlled laboratory experiment with ten participants. The collected data included visual appeal and perceived usability questionnaires, retrospective think aloud, and objective usability data. Statistical analysis was applied to the applicable data. The website manipulations showed that visual appeal was significantly worse but the results were not as clear for usability. To address these results, another preliminary study was added.

*Preliminary Study 5.* The purpose was to re-manipulate the usability and re-test it. Tuch and colleagues (2012) altered the usability of their website by changing the labels on three levels (the main menu and two submenus) but not on the final answer page. The same was done here, with the addition of randomizing the labels from the previous preliminary study so that the labels along the completion paths for each task changed with every click on the website. In addition, the location of each of the menu items that were needed in order to complete the tasks were randomly scattered in the menu system so that the menu had no categories. For example, the link to the “Pet Registration” was put under “Business” rather than the “Community” tab. Ten new participants usability tested it the same way as in Preliminary 4. The results showed that the low usability website was now significantly harder to use than the high usability website. The dataset was then ready for the main

study. This low visual appeal and low usability website is referred to as the ‘bad’ website from now on.

*Summary.* The five preliminary studies resulted in an empirically chosen and tested website data sample that statistically varied in usability and visual appeal. The data sample consisted of one website that was easy and pretty (i.e. the original, good website), manipulated into being hard and ugly (the bad website). According to participant responses, this website data sample was from a less familiar genre (i.e. city councils), which was necessary in order to control for the impact of the textual WOMs.

### Main Study Goal

The goal was to examine the influence of written WOM on visual appeal and usability of an unfamiliar website. To test this, the two website versions (good and bad) were used. Both websites were subjected to three WOM conditions: high usability and visual appeal (referred to as positive WOM), low in both (negative WOM), and neutral WOM (i.e. control condition). Thus, there were six conditions: good website with congruent WOM, good website with negative WOM, good website with neutral WOM, bad website with positive WOM, bad website with negative WOM, and bad website with neutral WOM.

The first hypothesis: *when textual WOM about visual appeal and usability is positive, then usability and visual appeal will be rated higher, both pre- and post-use.* The second hypothesis: *when textual WOM is negative, then usability and visual appeal will be lower pre- and post-use.*

### METHOD

#### Participants

A sample of 60 (39 males, 21 females; 48 aged 18-30 years, 12 aged 31+) university student volunteers participated, all with 20/20 or corrected to 20/20 vision, and screened for colour blindness. All participants were technology-savvy regular Internet users. Participants were randomly assigned and individually tested, approximately one hour per session, ten participants per condition, and each was given a \$20 gift voucher at the end of the session.

#### Materials

An informed consent, project information form, and demographic questionnaire were administered to inform participants of the study and to determine the participants’

background information. The System Usability Scale (SUS; Brooke, 1986) and the Visual Aesthetics of Websites Inventory – Short version (VisAWI-S; Moshagen & Thielsch, 2012) were used before and after use. The number of clicks per task was also counted, to see if objective usability (i.e. performance) was impacted by textual WOM.

The slideshow (the second step in Figure 2) three screenshots were taken of the website: (1) home page, (2) a main menu page, and (3) one from deeper in the menu. The website version in the slideshow corresponded to the participant’s condition.

Three textual WOMs were prepared, a paragraph long each. They set the context for the website use that occurred during the session. For example, the positive WOM was:

“Welcome to Gold Coast, Australia’s greatest travel destination! Your boss was delighted with your work and decided to promote you to senior manager of the company in Gold Coast. You are bound to love it there and the job’s pay is great. Before you start packing and head off, you’re going to check the city’s city council website out, to get some information which will help you get ready for the move. Recent surveys have found that the website is as beautiful as the gorgeous city. People are finding it incredibly easy to use, and they all recommended it to their family and friends. The developers created a masterpiece and the website won an award for best city council website in Australia in 2013.”

Ten information-retrieval tasks were then completed by each participant to examine the impact of textual WOM on website use and post-use. An example of a task was: “Who is eligible for free vaccinations in the Gold Coast?”

#### Design

This study adopted a two-by-three (two websites, three textual WOMs) between-group design. The website was shown in two parts, the first was the slideshow needed for pre-use data, and the second was the functioning website needed for use and post-use data.



Figure 2. Study flow

## Procedure

Once a participant was briefed on the purpose and procedure, they signed the consent and filled the demographics form. Following that, participants were given the written WOM paragraph according to the condition they were in, as seen in Figure 2. They were then given instructions, viewed the pre-use slideshow of the website and rated it on usability and visual appeal. Then participants were given ten information retrieval tasks to complete using the website. Post-use, participants filled out the visual appeal and usability scales again. The researcher then asked them for feedback before giving them a gift card and thanked them for their help.

## RESULTS

The results are first presented using beanplots to enable a visual comparison of the single variable data, i.e. the participant ratings between each of the websites. The statistical assumptions were tested to determine if the data were normally distributed and had homogenous variance. This was done to ensure that the appropriate statistics were applied. The SUS and VisAWI-S results are presented last.

### Preliminary Beanplot Results

Using the results from the SUS and VisAWI-S, beanplots were created to gain a general understanding of the data. Beanplots are a more advanced form of box plots, where the distributions are shown on both sides of the middle bar (Kampstra, 2008). They give the population spread which allows for more accurate conclusions to be drawn. They can also visually present more complicated results.

Pre-use and post-use visual appeal ratings are in Figures 3 and 4 respectively, and pre-use and post-use perceived usability in Figures 5 and 6, respectively. The SUS and VisAWI-S scores are presented on the vertical axis. In all of the figures, the grey beanplots on the top relate to the good website and the white ones on the bottom relate to the bad website. In Figures 3, 4, 5, and 6, the first columns on the left represents neutral WOM, the middle columns are where the subjects were given positive WOM, and the ones on the far right are where the participants were given negative WOM. The thicker lines indicate each condition's mean. The dotted lines are the website's overall mean, across the three conditions.

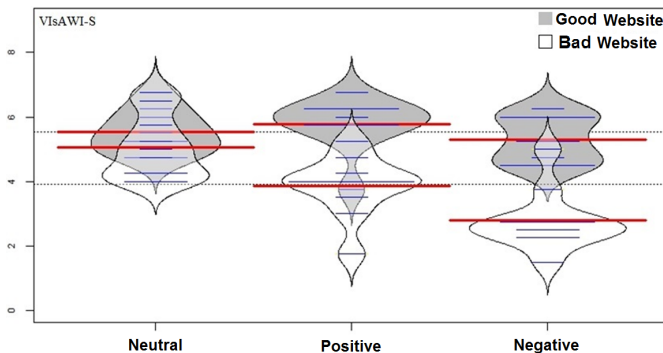


Figure 3. Beanplot of the pre-use visual appeal results.

For the good website, there are slight differences between the three WOM conditions, evidenced by the differences in the means (the thicker lines each bean has intersecting it) between the beans in Figure 3. The second column is the highest, followed closely by neutral WOM, in the first column, and then the lowest is the third column which is the negative WOM condition.

This finding suggests that participants perceived the same website differently based on their experimental condition. However, the difference between the means does not appear to be large. There also appears to be a bimodal distribution of data in the pre-use visual appeal ratings of the negative WOM (rightmost, top, gray bean in Figure 3). This suggests that about half of the participants agreed with the textual WOM and rated visual appeal lower.

The bad website depicts a slightly different story. With the bad website, the control was considered to be the 'prettiest', with positive WOM in the middle, and negative WOM being the lowest. Specifically, participants in the negative WOM condition perceived visual appeal to be on average one point lower (uglier) than participants in the positive WOM condition, and two points lower (uglier) than in the control condition. The positive WOM condition was most likely rated lower than the control condition because participants experienced some disappointment with the ugly website, but still rated it better than participants in the negative WOM condition.

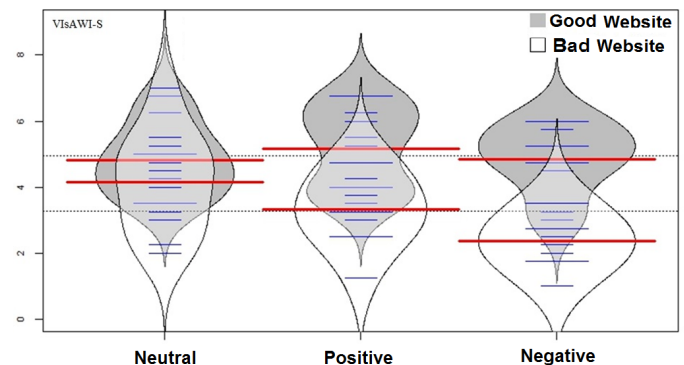


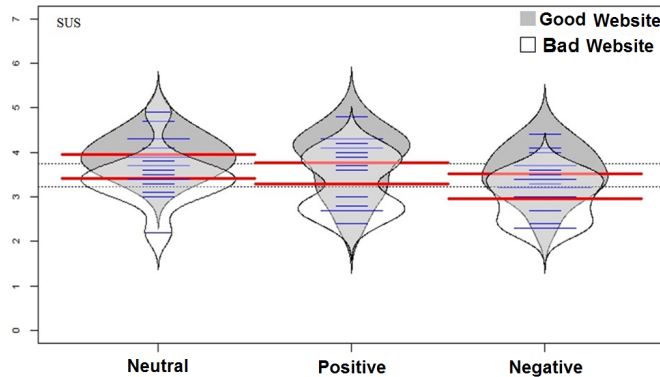
Figure 4. Beanplot of the post-use visual appeal results.

Across the good website conditions, there are slight changes between pre-use (Figure 3) and post-use (Figure 4) and between conditions; altogether dropping roughly half a point in visual appeal after use. A point refers to the 1-7 point VisAWI-S scale used for measuring visual appeal. Participants in the good website with positive WOM (middle upper bean in Figure 4) appear to be polarized, either rating post-use visual appeal higher or lower than the group's mean. The means from the neutral and negative WOM conditions appear to be similar yet the distributions are heavier in the opposite directions. Specifically, some of the negative WOM participants rated it higher than the control group. One reason this may have occurred is that some of the participants in the negative WOM group may

have thought that it was not as bad as the WOM said it would be, slightly raising their ratings.

Post-use visual appeal means in the bad websites are also similar to pre-use, but are more normally distributed. In addition, the control condition for the bad website has a larger spread suggesting that one or two participants thought it was a lot more visually appealing and one or two thought that it was a lot less visually appealing pre-use, but the majority was still located around the mean. Participants with negative WOM rated visual appeal lower than participants in the other two conditions. Again, the positive WOM condition was most likely rated lower than the control condition because participants may have experienced disappointment with the ugly website, yet still rated it as prettier than participants in the negative WOM condition.

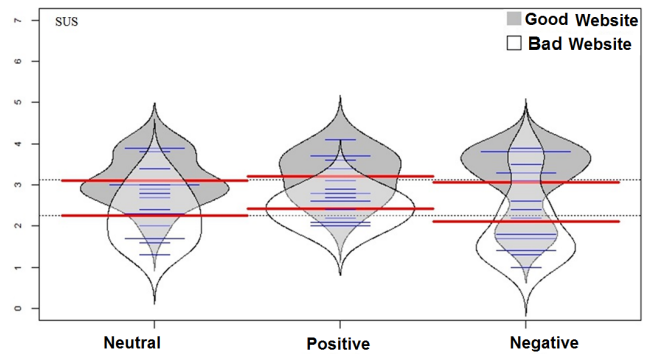
Altogether, evidence exists to support the first hypothesis that visual appeal differs between conditions in the good website, both pre- and post-use.



**Figure 5. Beanplot of the pre-use perceived usability results.**

There seems to be a small yet consistent difference between the groups across all conditions for pre-use usability, as seen in Figure 5. The neutral WOM condition seems to be rated as the easiest to use, followed by the positive WOM, and then by negative WOM which is rated as the hardest, across both website versions. All three conditions for the bad website appear to have a bimodal distribution, suggesting that some participants either had hope that the website would be easy to use or that they actually thought it would be bad. Given that the layout of both websites was not changed (so that visual appeal would not be altered as well), the similarity in usability ratings across conditions appears appropriate pre-use.

After use, ratings of usability dropped, especially for the bad website, as seen in Figures 5 and 6. Post-use, in both the good and the bad websites, the positive textual WOM condition was rated as the easiest, as seen in Figure 6. There also appears to be a difference between the control and negative WOM conditions, where the negative textual WOM condition was rated as lower, in both websites.



**Figure 6. Beanplot of the post-use perceived usability results.**

*Beanplot summary.* The main results are summarized in Table 1, where the results we expected are grayed out. The results that are not grayed out partially agree with what we expected would happen (positive WOM conditions were rated higher than negative WOM, for identical websites). Overall, pre-use visual appeal across the good website was similarly rated, whereas across the bad website, the neutral WOM was rated as prettiest, with the negative WOM rated the lowest. The same trend occurs with post-use visual appeal. Pre-use perceived usability also rated the control condition as the highest, followed by positive and negative. Post-use, the trend in perceived usability changes slightly, with positive WOM conditions being rated as the easiest and the negative rated as the hardest. Therefore, evidence appears to support the two hypotheses. Both visual appeal and perceived usability seemed to differ between conditions in the two websites, both pre-use and post-use. This suggests that written WOMs do affect the perception and ratings of visual appeal and usability.

Measure	Good Website	Bad Website
Subj. Usab Pre-use	-neutral WOM rated highest, positive WOM rated slightly lower, negative WOM lowest	-neutral WOM rated highest, positive WOM rated slightly lower, negative WOM lowest
Subj. Usab Post-use	-scores drop after use -positive WOM rated highest and negative WOM rated lowest	-scores drop after use -positive WOM rated highest and negative WOM rated lowest
VisAppl Pre-use	-positive WOM rated highest and negative WOM rated lowest	-neutral WOM was highest, negative WOM rated lowest
VisAppl Post-use	-scores drop after use -positive WOM rated highest and negative WOM rated lowest	-scores drop after use -neutral WOM rated highest, negative WOM rated lowest

**Table 1. Beanplot result summary of Subjective Usability (Sub. Usab), Objective Usability (Obj. Usab), and Visual Appeal (VisAppl).**

*Assumptions Testing*

The assumptions for normality and homogeneity of variance were checked for each variable across all conditions and were not always met. In addition, sample size per condition was relatively small (n=10), which meant that ANOVAs could not be applied to the data. Therefore,

non-parametric statistics were used. Thus, Kruskal-Wallis and Wilcoxon Mann-Whitney tests for pairwise comparisons were applied. Only the adjusted significance values are reported.

*Statistical Hypothesis Testing*

The result summary can be seen in Table 2 (where  $\alpha = .05$ ). Pre-use ( $p < .01$ ) and post-use ( $p < .05$ ) visual appeal ratings differed significantly within the bad website (main effects). Paired comparisons showed that negative WOM and control conditions in visual appeal differed both pre-use ( $p < .001$ ) and post-use ( $p < .05$ ). The differences found in the good website were not found to be significant.

For objective usability, the average number of clicks per task significantly differed ( $p < .05$ ) within the good website. Pairwise comparisons showed that the number of clicks were different ( $p < .05$ ) between the good website with positive WOM (where the average number of clicks was  $\bar{x} = 3.98$ ) and same website with negative WOM (where the average number of clicks was  $\bar{x} = 3.07$ ). This suggests that participants interacted more with the website that had the positive WOM. This is contrary to what was expected, suggesting that while participants had interacted less with the website when they read it was going to be hard to use, as if they were discouraged from using it. However, less interaction with the website would have resulted less completed tasks. Yet, completion rates were not found to differ. Thus, it is likely that users completed the tasks more efficiently with negative WOM as it may have encouraged users to be more careful.

Measure	Good Website	Bad Website
Sub.Usab	none	none
VisAppl Pre-use	none	negative and neutral WOM differed ( $p < .001$ )
VisAppl Post-use	none	negative and neutral WOM differed ( $p < .05$ )
Obj. Usab	clicks/task differed between positive and negative WOM ( $p < .05$ )	none

**Table 2. Statistical result summary of Subjective Usability (Sub. Usab), Objective Usability (Obj. Usab) and Visual Appeal (VisAppl).**

*Correlations*

Correlations were examined to illustrate any trends that may be occurring. In the conditions where the WOM is congruent with the website’s actual levels, the hypothesis was that correlations would be positive and strong between visual appeal and perceived usability pre-use and post-use. The hypothesis for incongruent situations was that the correlations would be less evident, because the textual WOM would influence ratings, changing any otherwise existing patterns.

*Correlations in the Control Conditions*

In the good website control condition, there were two significant correlations. Pre-use and post-use usability

( $r = .780, p < .01$ ), and pre-use and post-use visual appeal ( $r = .635, p < .05$ ) were highly and positively correlated.

In the bad website’s control condition, there were three significant correlations. Pre-use usability was highly and significantly correlated both with pre-use ( $r = .716, p < .05$ ) and post-use ( $r = .924, p < .01$ ) visual appeal. Pre-use and post-use visual appeal were also highly and significantly correlated with each other ( $r = .665, p < .05$ ).

*Correlations when Textual WOM and Website Levels are Congruent*

In the good website with positive WOM, there were four significant correlations, pre-use and post-use perceived usability were highly and positively correlated ( $r = .686, p < .01$ ). Pre-use perceived usability was also correlated highly and positively with both pre-use ( $r = .658, p < .05$ ) and post-use visual appeal ( $r = .782, p < .01$ ). Post-use perceived usability was correlated with post-use visual appeal ( $r = .715, p < .05$ ). In the bad website with negative WOM, there was one significant correlation: pre-use and post-use perceived usability ( $r = .797, p < .01$ ) were highly and positively correlated.

In the conditions where the reputation of visual appeal and usability were congruent with the website’s actual levels, the hypothesis was that correlations would be positive and strong between visual appeal and perceived usability pre-use and post-use. This was indeed the case in the good website. However, when both usability and visual appeal were worse in bad website, the hypothesis was partially supported, as it seems that participants attributed the poor website to usability and disagreed on visual appeal.

*Correlations when Textual WOM and Website Levels are Incongruent*

Nothing was correlated in the good website with negative textual WOM. Nothing was correlated when the bad website was given positive WOM. The discussion section comments more on these findings.

*Correlations summary.* The good website with positive WOM had the most variables correlated, having usability and visual appeal correlated pre-use and post-use, along with pre-use and post-use usability correlated. The bad website with negative WOM had only one significant correlation, where pre-use and post-use usability were correlated. The good website control condition showed no relationship between visual appeal and usability, only pre-use and post-use for visual appeal, and pre-use and post-use usability (i.e. each post-use rating was related to the pre-use but not with the other variable). The bad website control condition had significant correlations between pre-use visual appeal and usability, and with pre-use and post-use visual appeal.

Interestingly, both the good website with positive WOM and the bad website control condition showed that pre-use usability was correlated with post-use visual appeal. This suggests that the impression of pre-use usability influenced

their post-use ratings of visual appeal, but no causality can be established with correlations. No correlations were found in conditions where the textual WOM was incongruent with the website's actual visual appeal and usability levels. This was expected since WOM was meant to influence the relationship.

While causality cannot be inferred from correlations, there does seem to be some evidence that textual WOM does impact visual appeal and usability. Mainly, the control conditions appear to behave the same way as many studies in the literature. The conditions in which textual WOM and website visual appeal and usability levels are congruent also obtain similar results to the literature. However, when the reputation and website levels are incongruent, correlations disappear.

## DISCUSSION

### Results Summary

The beanplots showed slight variations in the means across the bad website where the control condition was rated highest pre-use and post-use visual appeal and pre-use perceived usability, with the negative WOM condition rated the lowest. Post-use, the trend in perceived usability changed slightly, with the positive WOM condition rated as easiest and the negative WOM condition rated as hardest. Therefore, a small trend did emerge, especially post-use, supporting the research hypotheses. The first hypothesis stated that: when the WOM of visual appeal and usability are positive, then usability and visual appeal will be rated higher, both pre-use and post-use. The second research hypothesis stated that: when textual WOMs are set to be negative, then usability and visual appeal will be lower pre-use and post-use.

In the bad website with congruently negative WOM, pre-use and post-use visual appeal ratings were rated as lower than the control use and post-use visual appeal ratings, respectively. This would suggest that participants in the bad website with negative WOM perceived the website to be uglier than participants in the control group, irrespective of use. This was also observed in the beanplots. Therefore, evidence exists for the second hypothesis. Written WOM impacts visual appeal since it was rated lower when WOM was negative, especially between the control group and the negative WOM group, which rated the website as a lot uglier. This difference lasted from pre-use to post-use, suggesting that *the impact of the WOM was strong enough to influence ratings after having been exposed to the website for roughly an hour*. However, this difference was not found in the good website, neither was it found between the positive and negative conditions of the bad website, nor with perceived usability.

Yet, the number of clicks per task varied significantly between the positive and negative WOM conditions in the good website. This suggests that *participants used the website less when they were told that it was hard to use,*

*seemingly uninterested or slightly deterred from using it*. Therefore, evidence suggests that textual WOM does impact both the perceptions and actions of people using the websites.

### Limitations

While the non-parametric statistics applied to the data in this study are appropriate, there is one limiting factor to the analysis. A relatively small sample size (ten participants per condition) may have not been sufficient to capture all differences that a larger sample would have shown.

Another limitation in this study was the assumption that textual WOM can influence people in a matter of seconds, while being in an unfamiliar physical environment. Unfamiliarity of the location and experimenter could have also influenced trustworthiness of the WOM, lowering its value. Furthermore, the textual WOM outwardly stated what they were intended to believe. It may have been more effective to give the sentiment in the textual WOM less overtly.

### Future Studies

Future studies should examine the influence of verbal WOM. This can be done by adding a confederate acting like a participant to verbally reinforce the textual WOM by telling participant their positive or negative opinion after having completing the study themselves.

## CONCLUSIONS

Our work aimed to acquire an understanding of the effect of textual WOM on usability and visual appeal. Written WOM did affect visual appeal and usability in most circumstances. Thus, there is support for the hypotheses. These results further the knowledge relating to the factors that affect the acceptance of website usage in particular with consideration of visual appeal and usability. The findings also show that 'what is beautiful is usable' may not be the full story – it is much more complex than that.

The importance of having textual ads and marketing that create positive reputations before the consumer entered the website, thereby improving the visual appeal and the usability experience of the consumer has been demonstrated in this paper. The authors recognize that the actual usability and visual appeal levels of a website are not the only important factors for participants. There are other significant contributing factors, in particular the site's reputation. To gain a deeper understanding of the effect reputations and WOM have on the visual appeal and usability, future studies should examine the impact of verbal WOM within this context.

## ACKNOWLEDGEMENTS

Thanks to Gitte Lindgaard and Chris Pilgrim for help with previous related work.

## REFERENCES

- Blijlevens, J. Typically the Best? Perceived Typicality and Aesthetic Appraisal of Product Appearances. PhD thesis, (2011), Delft University of Technology, The Netherlands.



- Brooke, J. *System Usability Scale (SUS)*. © Digital Equipment Corporation, (1986), UK.
- Davis, F. D. Perceived usefulness, perceived ease of use, and user acceptance of information technology, *MIS Quarterly*, (1989), 13, 3, 319–340.
- Gefen, D., Karahanna, E., and Straub, D.. Trust and TAM in online shopping: An integrated model. *MIS Quart.*, (2003), 27, 1, 51-90.
- ISO 9241/11. International Organization for Standardization. (1996). Retrieved June 2012.
- Kampstra, P. Beanplot: A Boxplot Alternative for Visual Comparison of Distributions. *Journal of Statistical Software*, (2008), 28, 1, 1-9.
- Katz, A. Aesthetics, Usefulness and Performance in User-Search-Engine. *Interaction Journal of Applied Quantitative Methods*, (2010), 5, 3, 424-445.
- Karapanos, E. User experience over time. In *Modeling Users' Experiences with Interactive Systems*, Springer Berlin Heidelberg, (2013), 57-83.
- Moshagen, M. and Thielsch, M. A short version of the visual aesthetics of websites inventory. *Behaviour and Information Technology*, (2012), 1-7.
- Pavlou, P.A. and Dimoka, A. (2006). The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation. *Information Systems Research*, 17(4), 392-414.
- Sangwon L. and Koubek, R.J. Understanding user preferences based on usability and aesthetics before and after actual use, *Interacting with Computers*, (2010), 22, 6, 530-543.
- Setterstrom, S. Assessing credibility and aesthetic perception across different exposure times on a health care information website. Graduate thesis. (2010), Iowa State University.
- Smith, D., Menon, S., and Sivakumar, K. Online peer and editorial recommendations, trust, and choice in virtual markets. *Journal of interacting marketing*, (2005), 19, 3, 15-37.
- Stojmenovic, M., Lindgaard, G., & Pilgrim, C. (2014). Perceived and Objective Usability and Visual Appeal in a Website Domain with a Less Developed Mental Model, *The 26th ACM Australian Computer-Human Interaction Conference OZCHI'14*, December 2-5, 2014, Sydney, NSW, Australia, pp. 316-323.
- Stojmenovic, M. (2016). Your Reputation Precedes You: The influence of expectations on usability and visual appeal in a web environment. *PhD Thesis*, Swinburne University of Technology.
- Tractinsky, N., Katz, A.S., and Ikar, D. What is beautiful is usable, *Interacting with Computers*, (2000), 13, 2, 127-145.
- Tuch, A.N. Roth, S.P. Hornbæk, K. Opwis, K. and Bargas-Avila, J.A. Is beautiful really usable? Toward understanding the relation between usability, aesthetics, and affect in HCI, *Computers in Human Behavior*, (2012), 28, 5, 1596-1607.
- Ye G. and van Raaij W. F. What inhibits the mere-exposure effect: Recollection or familiarity? *Journal of Economic Psychology*, (1997), 18, 629-648.
- Zettl, H. *Sight, Sound, Motion: Applied media aesthetics*. (1999), Belmont, CA: Wadsworth.