E-Commerce Websites for Developing Countries – A Usability Evaluation Framework

Abstract

Purpose: The purpose of this research was to develop a methodological usability evaluation framework for e-commerce websites.

Design/methodology/approach: A multiple-case study (comparative design) was used, where three usability methods (user testing, heuristic evaluation and web analytics) were applied to three Jordanian e-commerce companies. The resulting framework took into account the advantages and disadvantages of the methods in identifying specific usability problems on the e-commerce websites.

Findings: A four-step framework was developed to facilitate the evaluation of e-commerce sites.

Research limitations: The framework was developed and tested using Jordanian users, experts and e-commerce sites. The study compared the ability of the methods to detect problems that were present, however, usability issues not present on any of the sites could not be considered when creating the framework.

Practical implications: The framework helps e-commerce retailers evaluate the usability of their websites and understand which usability method(s) best matches their need.

Originality/value: This research proposes a new approach for evaluating the usability of websites, specifically e-commerce sites. A particularly novel approach is the use of web analytics (Google Analytics software) as a component in the usability evaluation in conjunction with heuristics and user testing.

Keywords: framework, e-commerce websites, developing countries, user testing, heuristic evaluation, Google Analytics.

Paper Type: Research Paper**.**

1. Introduction

To be successful, websites need to have good usability. Usability is a measure of how easy the interface is to use (Najjar, 2005; Nielsen, 2003). Nielsen (2003), for example, stated that the first law of e-commerce is that, if users are unable to find a product, they cannot buy it. Consequently, unusable websites will lead shoppers to abandon them, resulting in a loss of sales (McKinney *et al*., 2002).

A variety of usability evaluation methods have been developed which could be employed to identify usability problems of e-commerce websites. These methods can be categorised in terms of how the usability problems are identified: for example by users, evaluators or tools.

User-based usability evaluation methods usually involve users being observed undertaking pre-defined tasks with the purpose of identifying usability problems (Brinck *et al*., 2001). User-based approaches have been frequently used to evaluate the usability of e-commerce websites (Agarwal and Venkatesh, 2002; McKinney *et al*., 2002). For example, McKinney *et al*. (2002) developed constructs and corresponding measurement scales with users for measuring web customer satisfaction and Tilson *et al*. (1998) asked sixteen users to complete tasks on four e-e-commerce sites and report what they liked and disliked. Freeman and Hyland (2003) also used a similar technique to evaluate and compare the usability of e-commerce sites. Research outcomes proved the usefulness of user-based methods in identifying major design problems which prevented users from interacting with the sites successfully.

Evaluator-based usability evaluation methods involve having a number of expert evaluators assess the user interface to judge whether it conforms to a set of usability principles (namely 'heuristics') (Nielsen and Mack, 1994). Agarwal and Venkatesh (2002) described a heuristic evaluation procedure for assessing a firm's website usability. In terms of e-commerce sites, Chen and Macredie (2005), for example, used expert evaluator methods to evaluate the usability of four online supermarkets. Again, research outcomes proved this method successful in identifying both usability problems and good design features on the sites.

Several studies have compared both user testing and heuristic evaluation methods. Barnard and Wesson (2004), for example, used both methods to investigate design issues and problems that were of high significance for e-commerce sites in South Africa from the perspective of both experts and users. Problems identified by both

users and experts were considered to be significant while those that were missed or uniquely identified by only one of those methods were ignored. Other comparative studies have shown that heuristic evaluation methods uniquely identify more usability problems than usability testing (Desurvire *et al*., 1991; Doubleday *et al*., 1997; Fu *et al*., 2002). Although these studies did not explain the distribution of usability problems identified by the methods in terms of whether they were major or minor problems, other studies have discussed this issue and offer various findings (Jeffries *et al*., 1991; Law and Hvannberg, 2002). Jeffries *et al*. (1991), for example, found that heuristic evaluation identified a larger number of both serious and minor problems in comparison to user testing. However, Gray and Salzman (1998) criticised the design of this study because the number of participants was too small for statistical analysis. By contrast, Law and Hvannberg (2002), found that while heuristic evaluation found a larger number of minor problems compared to user testing, the latter was better at uniquely identifying major problems.

Only a few studies, however, highlighted the types of specific usability problems identified by user testing and heuristic evaluation methods. One such study by Mariage and Vanderdonckt (2000) evaluated an electronic newspaper. They identified examples of usability problems that were identified by user testing and missed by heuristic evaluation (i.e. inappropriate choice of font size), and examples of problems that were identified by heuristic evaluation and confirmed by user testing (i.e. a home page layout that was regarded as being too long).

Tan *et al*. (2009), who compared user testing and heuristic evaluation by evaluating four commercial websites, also classified usability problems by their types. Both methods were found to be equally effective in identifying different usability problems related to five categories (navigation, information content, layout organisation and structure, usability and availability of tools, and common look and feel) but user testing did not identify problems relating to two issues (compatibility, and security and privacy issues).

Software-based usability evaluation methods use software tools to identify usability problems. Web analytics is an example of this approach and involves collecting, measuring, monitoring, analysing and reporting web usage data to understand visitors' experiences (Web Analytics Association, [n.d.]). There are two common approaches to web analytics. These are server-based log file analysis and client-based page-tagging (JavaScript tagging). Analysis of server-based log files was

the first approach used for web analytics. It involves the use of a server's log file to collect access and visit data. Kaushik (2007) indicated that while the log file technique was used widely as a data source for web analytics, the disadvantages of using this approach (i.e. the use of caching techniques, and the use of IP addresses to identify unique visitors) were noticed by both web analytics vendors and customers.

These challenges led to the emergence of page-tagging techniques as a new source for collecting data from websites. Page-tagging involves adding lines of script (JavaScript code) to the pages of a website to gather statistics from them. Page-tagging is typically more accurate than using web server log-files. Reasons for the improved accuracy of this method are that most page tags are based on cookies to determine the uniqueness of a visitor and not on the IP address, and this method is not influenced by caching techniques (Kaushik, 2007; Peterson, 2004). An example of a web analytic tool that uses the page-tagging approach, and which has had a major effect on the web analytics' industry, is Google Analytics (GA).

Web metrics give meaning to data collected by web analytics tools and can be placed into two categories: basic and advanced. Basic metrics are raw data which are usually expressed in raw numbers (i.e. visits, page views etc.). Advanced metrics are metrics which are expressed in ratios or percentages instead of raw numbers and are designed to simplify the presentation of web data, and to guide actions that optimise online business (Peterson, 2006). An example of one such metric is *bounce rate,* which represents the percentage of single page visits: i.e. visits where users left the site after visiting only one page (Peterson, 2006). The use of basic metrics to measure the traffic of websites has been criticised for several reasons, one of which relates to their simplicity in addressing only some aspects of web measurement (Inan, 2006; Phippen *et al*., 2004).

Most of the earlier studies that used web analytics to evaluate and improve the design and functionality of websites used log-file based web analytics and employed basic metrics included in the reports generated by the web log analyser (Jana and Chatterjee, 2004; Ong *et al*., 2003; Peacock, 2003; Xue, 2004; Yeadon, 2001). However, at least two studies have recognised the appearance of GA software, which uses the page-tagging approach, and have used this tool to evaluate and improve the design and content of websites (Fang, 2007; Prom, 2007). Both used the standard reports from GA (i.e. funnel navigation) without deriving specific metrics. These two studies suggested that the GA tool could be useful since GA's reports enable

problems to be identified quickly and help determine whether a site provides the necessary information to its visitors.

Despite the literature outlined above, there has been little research evaluating the usability of e-commerce websites employing user-based, evaluator-based and software-based (GA) usability evaluation methods together. Studies by Fang (2007) and Prom (2007) illustrated the potential usefulness of using GA to evaluate websites with the intention of improving their usability. However, further research is needed to explain how GA can be used to evaluate the usability of e-commerce websites by employing advanced web metrics. Similarly, additional research is needed to investigate and compare detailed types of specific usability problems that could be identified uniquely by user testing, heuristic evaluation methods and GA, and those problems that could be either missed or commonly identified by these methods when evaluating an e-commerce website.

The research described here aims to address this gap and presents a framework which involves user-, evaluator- and software-based methods. A combination of these different methods is proposed in an attempt to reduce the time, effort and money expended by e-commerce vendors when assessing the usability of their websites. It is based on e-commerce websites in Jordan, which like other developing countries, face additional challenges to those in the West in making their sites acceptable, usable and profitable in the absence of legislation and regulations (Obeidat, 2001).

The aim of the research described here was, therefore, to develop a methodological framework to investigate the usability of e-commerce websites in Jordan. The specific objectives for the research were:

- To use three different approaches (user testing, heuristic evaluation and a leading web analytics package (*Google Analytics*) to evaluate the usability of a selection of e-commerce websites by identifying the main usability problem areas on the sites from three different perspectives: evaluators, users and software tools.
- To determine which of these approaches were the most effective in evaluating each usability problem area.
- To create a framework to identify how to evaluate e-commerce sites in relation to specific areas.

The remaining sections of this paper describe the methods used, present the main results, illustrate the framework and finally, present the conclusions.

2. Methodology

2.1 Selection of usability evaluation methodologies

The selection of the three usability evaluation methods (user testing, heuristic evaluation and GA) was based on evidence that showed the methods complement each other in as much as they are able to identify usability problems from different perspectives (Desurvire *et al*., 1992; Fu *et al*., 2002; Jeffries *et al*., 1991; Kaushik, 2007; Law and Hvannberg, 2002; Nielsen, 2003; Peterson, 2004). Several researchers have suggested the need to use both heuristic and user testing in tandem in order to identify different kinds of usability problems (Desurvire *et al*., 1991; Fu *et al*., 2002; Jeffries *et al*., 1991; Law and Hvannberg, 2002; Nielsen, 2003). Regarding the use of the *Google Analytics* approach, other researchers have stressed the importance of employing other methods, such as usability methods, alongside the web analytics approach (Kaushik, 2007; Peterson, 2004). Web analytics packages monitor users' behaviour over a long time period relative to user testing and identify issues, often missed by user testing, that could help in identifying additional usability problems.

2.2 Case studies

The research was based on a multiple-case study (comparative design) where the three methods (user testing, heuristic evaluation and GA) were applied to three Jordanian e-commerce companies in order to develop the framework. To make the identification of usability faults as efficient and effective as possible it was necessary to ascertain the order of deploying the methods as well as determining which methods should be used for what purposes. Initially, twenty seven e-commerce companies in Jordan were identified from five electronic Jordanian and Arab directories and a Google search. These companies were contacted and three of them agreed to participate. Two of them sold women's clothes and the third sold hand-crafted gifts to both national and international customers. This study focused on investigating the usability of these websites from the point view of national (Jordanian) customers. The three methods were employed concurrently at each site to ensure that the results were not affected by any website changes made by that particular e-commerce vendor.

Initially, the required GA script was inserted on the companies' web sites to enable GA software to track the usage of the e-commerce sites. The usage of the websites was then monitored for three months.

In order to employ the user testing method, a task scenario was developed for each of the three websites. This involved specifying typical tasks for e-commerce websites, such as finding information and products, using the site's search engine, purchasing products, changing the content of the shopping cart, and changing the user profile. Each company was asked to explain the characteristics of their current and proposed user base and then twenty users were recruited with characteristics that matched those specified by the companies. Data were gathered from each user testing session using screen capture software (Camtasia), with three post-test questionnaires (one post-test questionnaire was given to each user after completing the tasks for each site to get user feedback). Observation of the users working through the tasks, in addition to taking comments from the users while interacting with each site, was also undertaken.

In addition, a set of comprehensive heuristics, specific to e-commerce websites, was devised based on a detailed review of the HCI literature. Specifically, the developed heuristics were based on: general texts on how to design usable websites (Nielsen, 2000; Sharp *et al*., 2007; Pearrow, 2000; Brinck *et al*., 2001); design criteria developed for evaluating e-commerce websites (Elliott *et al*., 2000; Davidson and Lambert, 2005; Oppenheim and Ward, 2006; Van der Merwe and Bekker, 2003; Hung and McQueen, 2004); research investigating the relative importance of e-commerce web design issues and features that affect purchasing habits (Barnard and Wesson, 2004; Claudio and Antonio, [n.d.]; Chen and Macredie, 2005; Freeman and Hyland, 2003; Oppenheim and Ward, 2006; Pearson *et al*., 2007; Tarafdar and Zhang, 2005; Tilson *et al*., 1998; White and Manning, 1998). The heuristics used in the research were organised into five major categories: architecture and navigation, content, accessibility and customer service, design, and the purchasing process. Five web experts, who had extensive design experience in e-commerce websites (more than ten years), evaluated the sites using the heuristic guidelines. This number was considered appropriate given that Nielsen and Mack (1994) and Pickard (2007) both recommend using between three to five evaluators. The heuristic evaluators were asked to indicate whether they felt any usability problems were likely to be minor or major. To ensure inter-rater reliability (i.e. the extent of agreement between the heuristic evaluators) Kappa statistics were calculated. Overall, the Kappa statistic for all the usability problems identified by the evaluators was 0.69, which, according to Altman (1991), indicates *good* agreement among the evaluators.

After all three methods had been deployed the data were analysed in three stages to determine which methods identified each usability problem area. The first stage involved analysing each usability method separately for each case and identifying the usability problems obtained from each method. The user testing method was analysed by examining: performance data; in-session observation notes; notes taken from reviewing the Camtasia sessions; users' comments noted during the test; and quantitative and qualitative data from the post-test questionnaires. The heuristic evaluation method was analysed by examining the heuristic evaluators' comments obtained during the fifteen sessions. The web usage of the three sites, tracked using GA, was measured using a trial matrix of 41 advanced web metrics divided into nine categories (Table 1).

The second stage of the analysis aimed to identify a list of common usability problems pinpointed by each method. This was achieved by performing a comparison of each usability evaluation method across the three cases. The third stage of the analysis was undertaken in order to generate a list of standardised usability problem themes and sub-themes to facilitate comparison among the methods. Problem themes and sub-themes were identified from the common usability problem areas which were generated by each method. Ten problem themes (usability problem areas) and 44 problem sub-themes (usability problem sub-areas) were finally identified from an analysis of the methods and the ability of each method to accurately identify each problem sub-theme was recorded (see Appendix).

Table 1: Trial matrix of web metrics

No.	Metrics Category	Metrics
1	General usability metrics	Average time on site, average page views per visit, percentage of long, medium and short time spent visits, percentage of high, medium and low click depth (page view) visits, bounce rate
2	Internal search metrics	Average searches per visit, percent visits using search, search results to site exits ratio
3	Top landing pages metrics	Bounce rate, entrance sources, entrance keywords
4	Top content pages (most viewed pages) metrics	Bounce rate, average time, percentage of site exits
5	Top exit pages metrics	Percentage of site exits
6	Finding customer support information metrics	Information find conversion rate[1], feedback form conversion rate.
7	Purchasing process metrics	Cart start rate[2], cart completion rate, checkout start rate[3], checkout completion rate, ratio of checkout starts to cart starts, funnel report[4].
8	Visitors' metrics	Ratio of new to returning visits, visitor engagement index[5], percentage of high, medium and low frequency visits, percentage of high, medium and low recency visits, language, operating systems, browsers, screen colours, screen resolutions, flash versions, Java support, connection speed.
9	Financial performance metrics	Average order value, order conversion rate, average revenue per visit, average visits to purchase, average days to purchase, average items per cart completed.

3. Results

3.1 Google Analytics results

The results obtained from the trial matrix of 41web metrics shown in Table 1 were investigated. The metrics could either be used individually, or in combination, to identify potential usability problems on an e-commerce website in relation to six areas: navigation, architecture, content/design, internal search, customer service, and the purchasing process. Figure 1 shows the suggested matrix and the combination of metrics that were found useful in each area. The matrix also includes specific metrics which were useful in indicating specific web pages such as top landing pages, top content pages, top exit pages and those in the purchasing process that had potential usability problems in their content or design.

[1] Information find conversion rate: Percentage of visits where visitors viewed customer support pages [34].
[2] Cart start rate metric: Percentage of visits that involve visitors adding at least one item to their shopping cart [34].
[3] Checkout start rate metric: Percentage of visits that involve visitors who clicked the checkout button [34].
[4] Funnel report: This involves an analysis of the navigational paths followed by visitors based on a number of identified steps.
[5] Visitor engagement index: The average sessions or visits per visitor [34].

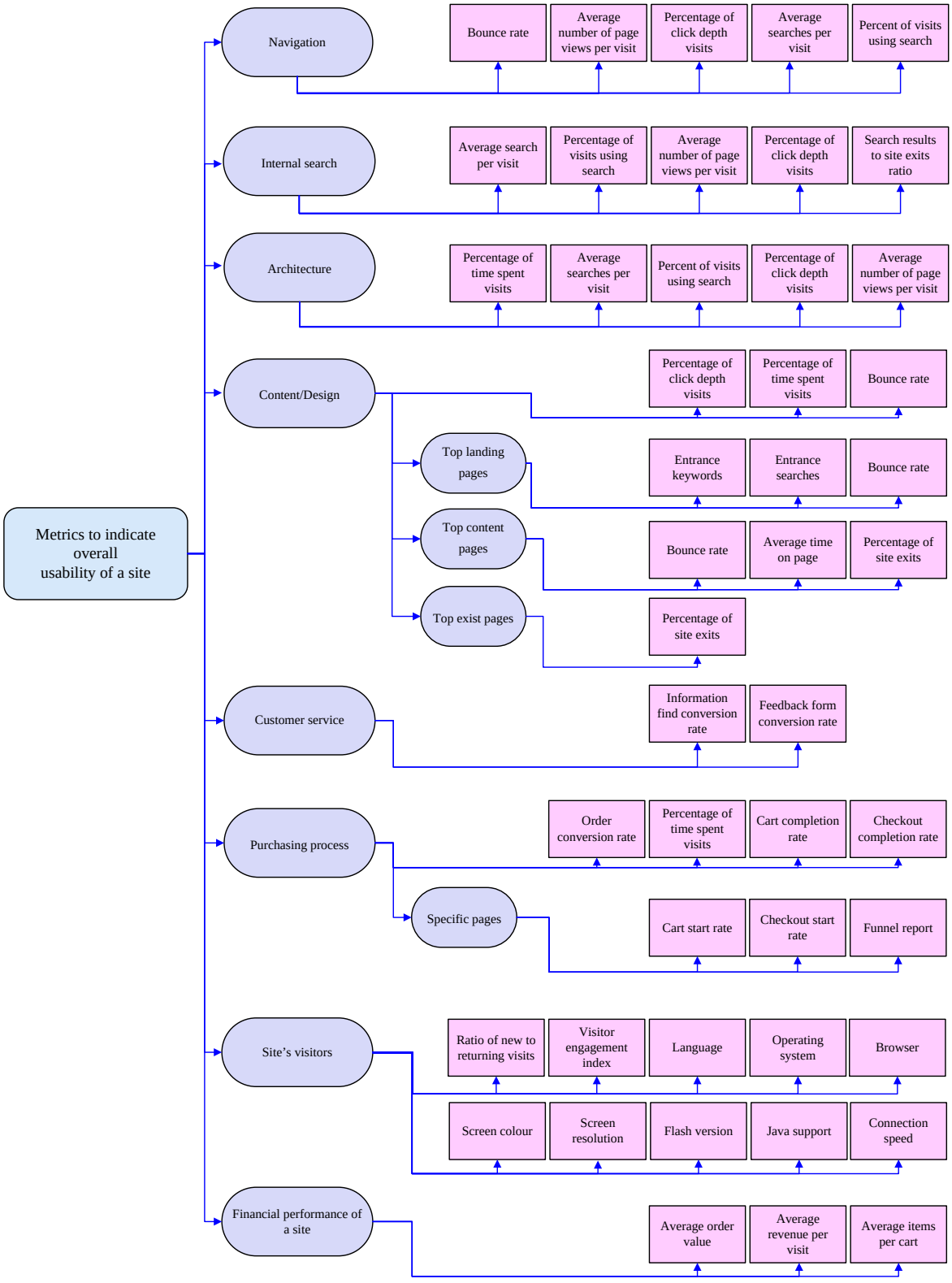

Figure 1: The suggested web matrix

Figure 1 also includes specific web metrics which helped to provide supplementary information about the site's visitors and its financial performance which could not be provided by the user testing and heuristic evaluation methods. These metrics enhanced the evaluator's understanding of the overall usability of a site.

To illustrate how the combination of metrics in Figure 1 provided potential indications of usability problems in the overall purchasing process for the three sites, as well as indications of potential problems with specific pages that make up the purchasing process, two examples are provided below:

- The combination of order conversion rate, time spent on site, cart completion and checkout rate metrics suggest that the three sites had usability problems in their overall purchasing process. Specifically, the low values of the order conversion rate metric of all sites indicated that few visits resulted in an order. The relatively low values for the percentage of long visits suggests that few visitors were engaged in purchasing activity on the three sites. The low cart completion rate and checkout completion rate metrics also suggest that the three sites had usability problems in their purchasing processes. These findings agreed with the user testing and heuristic evaluation methods where the users and the heuristic evaluators experienced problems with the purchasing process of all three sites.
- The low value of the *cart start rate* metric (which showed that few users added anything to the shopping cart) suggests that sites 1 and 2 had usability problems on their product pages. This was confirmed by the user testing and the heuristic evaluation methods. On site 1, for example, the performance data and observation, and the heuristic evaluation methods identified a navigation problem relating to misleading links. The qualitative data from the satisfaction questionnaire, together with the heuristic evaluation methods, also identified two content problems on sites 1 and 2 on the product pages; namely inaccurate information and missing product information.

The results, however, indicated some limitations of employing the metrics in the evaluation of the usability of e-commerce websites. These related to the fact that the web metrics indicated only a potential usability problem area which could relate to one or more specific problem sub-areas or sub-themes in this area. They could not provide in-depth detail about specific problems that might be present on a page. These

specific problem sub-themes were identified by the user testing and/or the heuristic methods.

3.2 User testing and heuristic evaluation methods

The heuristic evaluation was more effective than user testing in terms of identifying a larger proportion of problems (see Table 2), however, nearly all of these were rated by the evaluators as "minor". User testing found six more "major" problems than heuristic evaluation. An example of a "major" problem was when a user made an error and was unable to recover and complete the assigned task in the allotted time, whereas an example of a "minor" problem was when a user recognized they had made an error but then were able to recover from it.

Table 2: Distribution of usability problems uniquely identified by the two methods

	Heuristic testing	**User testing**
Minor	159	2
Major	13	19
Total	**172**	**21**

An analysis of the usability problems that were uniquely and commonly identified by the two methods based on the number of usability problems related to the ten problem themes is presented in Figure 2. It can be seen that the heuristic evaluation method was more effective in identifying a large number of problems compared to user testing with respect to all the problem themes, with the exception of one, the purchasing process, where user testing identified a larger number of problems.

Further specific details regarding the major and minor problems identified in the ten problem themes and 44 sub-themes are presented in the Appendix, highlighting the effectiveness of the user testing and heuristic evaluation methods in the case studies. The results show the number and severity level of each specific problem area identified by the user testing and heuristic evaluation methods with regard to the ten main usability problem areas and their corresponding subareas. The results also illustrate which method(s) are useful for identifying minor and major problems, those that might fail to identify specific types of problems, or those that will always fail to identify specific types of problems.

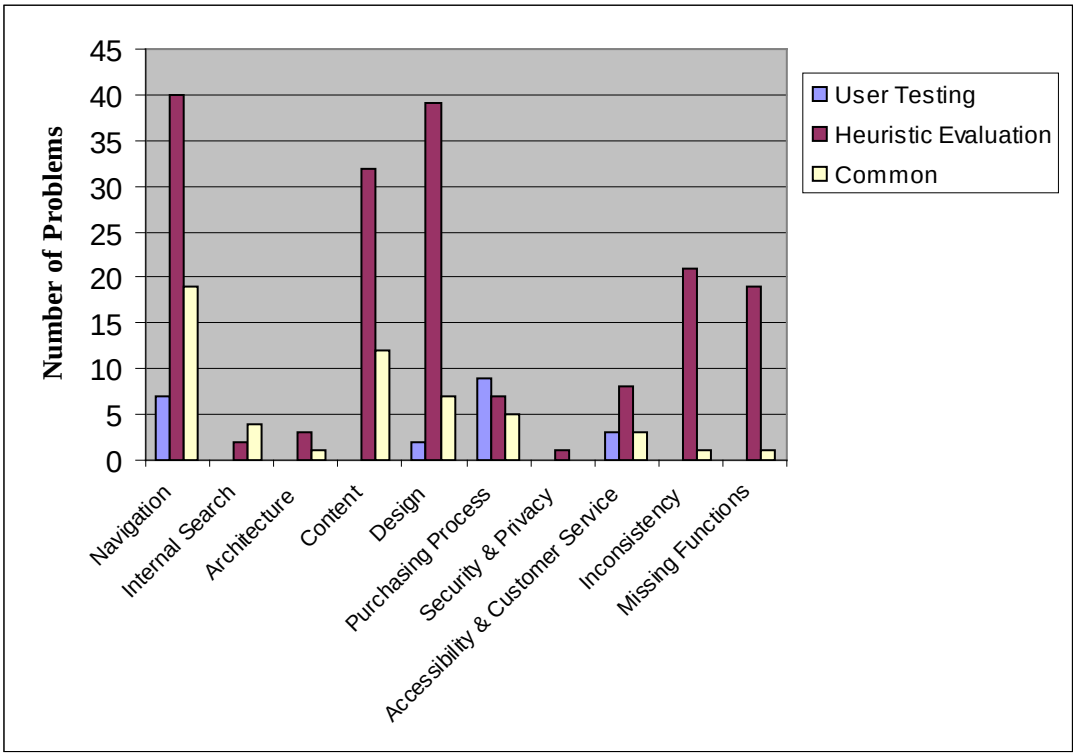

Figure 2: Distribution of usability problems identified by the two methods by number and types of problem

For example, regarding the navigation theme, it can be seen that the user testing method was more effective than the heuristic evaluation in uniquely identifying major problems related to two specific areas: misleading links (i.e. links with names that did not meet users' expectations as the name of the link did not match the content of its destination page) and links that were not obvious (i.e. links that were not situated in obvious locations on the sites). However, the heuristic evaluation was more effective than the user testing in uniquely identifying other major navigational problems (i.e. pages without a navigation menu) and minor problems related to four areas: misleading links, links that were not obvious, broken links and orphan pages (i.e. pages that did not have any links). One example from the case studies that relates to navigation problems was that on site 1, users did not expect the advanced search link to only allow searching the site by colour and price (Figure 3). This link therefore constitutes a problem as the link name ('Advanced Search') did not match the content of the destination page (Figure 3). Users expected this page to have search boxes with many options available to search the site. However, this page included only combo boxes that allowed users to only search the site on limited criteria. It did not allow

users to search by keywords or within a specific category. The heuristic evaluators, however, missed identifying this problem**.**

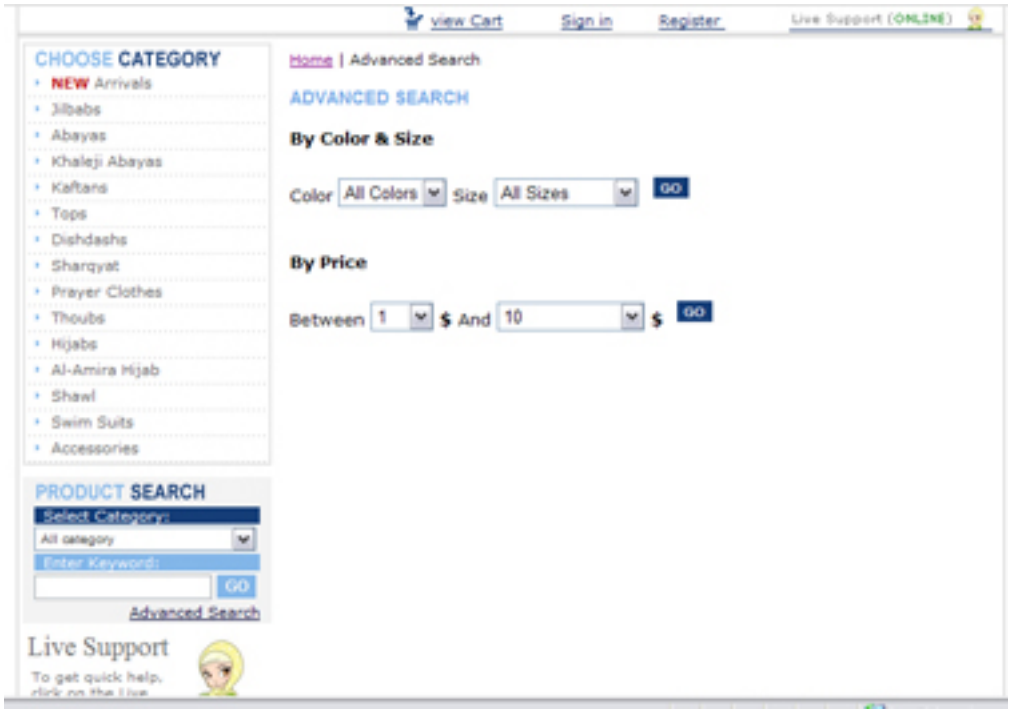

Figure 3: Advanced search link and advanced search page on Site 1

In addition to the different kinds of navigational issues that were identified by the two methods, there were many examples across the other usability themes that illustrated the different kinds of issues identified by these methods. These included:

- Internal search theme: Both heuristic evaluation and user testing identified similar major problems relating to inaccurate results and similar minor problems relating to limited search options, however heuristic evaluation identified additional minor problems such as an inappropriate position of the link to the search facility.
- Architecture theme: Both methods identified a major problem with one of the sites regarding overly complex categorization of the site's products, however heuristic evaluators identified more minor problems such as poorly ordered menu items.
- Content theme: The heuristic approach was generally more effective in identifying both major and minor problems such as irrelevant content, inaccurate or missing information and poor grammar.
- Design theme: User testing was more effective in identifying a major problem relating to inappropriate design, but heuristic testing was generally better at

identifying minor problems such as the poor quality of images, broken images and missing alternative text.

- Purchasing theme: User testing was more effective than heuristic evaluation in identifying major problems such as missing information relating to the purchase process (such as the content of the shopping cart) and missing indicators for required fields; however heuristic testing uniquely identified other problems such as difficulty with log-on procedures, overly long registration pages, and missing confirmation information as a result of an action.
- Security theme: This is an area where heuristic evaluation performed better than user testing; user testing did not identify any problems in this regard whereas the heuristic evaluators identified an issue with one site relating to lack of privacy/security statements and policies.
- Accessibility theme: Although user testing uniquely identified major problems relating to difficulties in finding help information, heuristic testing seemed better at identifying minor problems such as currency support, lack of feedback options and ease of accessing the site via search engines.
- Inconsistency theme: The heuristic evaluation was more effective than user testing in identifying a large number of inconsistencies such as inconsistent colour, typefaces, terminology etc.
- Capability theme: The heuristic evaluation was more effective in identifying a large number of minor problems such as lack of delivery options.

In summary, the appendix shows that although heuristic evaluation identified more problems in total, many of these were minor problems. Both methods identified a reasonable number of major problems, with 13 of these being uniquely identified by heuristic evaluation and 19 being uniquely identified by user testing. It seems clear that for a complete and thorough evaluation of an e-commerce website both of these evaluation methods need to be considered.

3.3 Costs of employing the three methods

The cost of employing the three methods (heuristic evaluation, user testing and the GA tool) was estimated in terms of the time spent designing and analysing each of these methods. The approximate time taken to design and analyse the heuristic evaluation, user testing and GA methods was 247 hours, 326 hours and 360 hours respectively (see Table 3). Identifying and combining suitable web metrics for use in the study took a long time, 232 of the total 360 hours spent on GA. However, if these were to be used again, then the time required for future GA tracking and data analysis would be was considerably less (approximately 120 hours). Compared to other research, the amount of time spent on heuristic evaluation in this research was considerably more, possibly due to the fact more time was spent setting up and designing the evaluation and analyzing the data. Additional time was also spent recruiting the specialists, which proved difficult in a developing country.

Table 3: Cost of employing usability evaluation methods

Study	Time Spent on User Testing	Time Spent on Heuristic Evaluation
Jeffries *et al*. (1999)	**199 hours** Time spent on analysis. Six subjects participated.	**35 hours** Time spent learning the method and becoming familiar with the interface under investigation (15 hours) and on analysis (20 hours). Four usability specialists conducted this method.
Law and Hvannberg (2002)	**200 hours** Time spent on the design and application of this method. Ten subjects participated.	**9 hours** Time spent on the design and conduction of this method by two evaluators
Doubleday *et al.* (1997)	**125 hours** Time included 25 hours conducting 20 user sessions, 25 hours of evaluator support time, 75 hours of statistical analysis	**33.5 hours** Time included 12.5 hours of evaluators' time, 21 hours transcription of the experts' comments and analysis
This Research	**326 hours** Time included 136 hours setup and design, 20 hours collecting data from 20 user sessions, 170 hours analysis	**247 hours** Time included 128 hours setup and designing, 15 hours collecting data from five web experts, 104 hours analysis

4. An evaluation framework

A framework for evaluating e-commerce websites was developed based on the results and, in particular, a detailed analysis of the advantages and disadvantages of three methods (user testing, heuristic evaluation and GA software), in terms of the specific areas of usability problems that they could or could not identify on the test websites (see Figure 4).

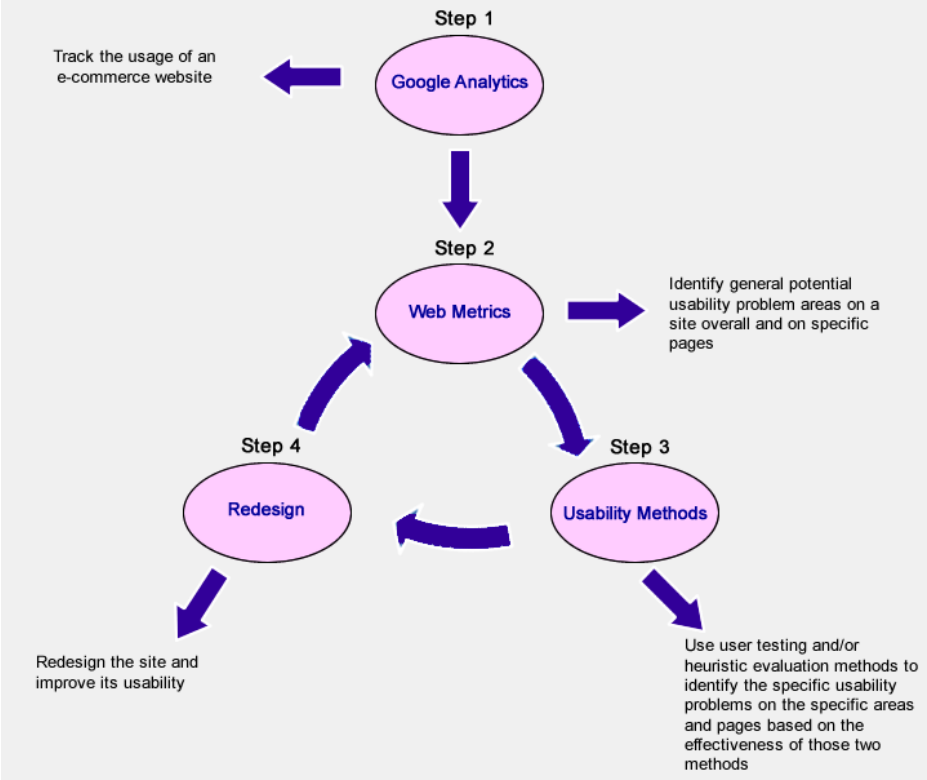

Figure 4: A framework to evaluate the usability of an e-commerce website

Specifically, the framework capitalises on the advantages of GA software by using the recommended web metrics, (Figure 1), to identify the areas of an e-commerce site that appear to have usability problems. Then, and because of the limitations of these web metrics, the framework identifies specific areas of focus, enabling user testing and/or heuristic evaluation to provide more precise details regarding the specific usability problem areas on a site. The use of GA at an initial stage in the framework enables evaluators to identify the specific usability areas that are most problematical; emphasis can then be placed on investigating these specific areas which, in turn, may reduce the time taken to undertake the user testing and heuristic evaluation procedures. Table 4 provides a summary of the four steps of the suggested framework.

Table 4: Summary of the steps of the suggested framework

Step	Objective	Task	Expected Outcomes
1	To configure an e-commerce website and GA software to make them ready so that GA software could track the usage of the website.	1. Identify the key business processes in an e-commerce website and the required pages users are expected to go through while completing the processes. 2. Configure GA software by adding the identified key business processes. 3. Insert GA code in the pages of the e-commerce site to be tracked by GA software.	GA software will start to collect data regarding the usage of the e-commerce website.
2	To identify general potential usability problem areas on an e-commerce website overall, and to highlight specific pages on the site that have potential usability problems.	1. Use the suggested matrix of web metrics (Figure 1) to measure the site's usage. 2. Identify the metrics with values that may indicate problems (i.e. low value for *average number of page views per visits*). 3. Use Figure 1 to identify the problem areas on the site and/or on its specific pages. For example: If the site has low values for *average number of page views per visits* and *percentage of high or medium click depth visits* metrics together with high values for *bounce rate, average searches per visits* and *percentage of visits using search* metrics, then this indicates a navigational problem in the site.	• The identification of potential usability problem areas on a site overall. • The identification of specific pages on the site that appear to have potential usability problems. These pages will include pages encountered by visitors while completing the identified key business processes (i.e. those identified in Step 1). Entry pages, most viewed pages and exit pages that have potential usability problems will also be identified. • The description of the site's visitors and its financial performance.
3	To identify the detailed specific usability problems on the specific areas and pages on the e-commerce website (resulting from Step 2).	1. Use the Appendix, which summarises the effectiveness of user testing and heuristic evaluation methods with regard to their ability to identify specific problems on a site, to decide which method(s) to employ. For instance, if Step 2 suggests a navigational problem, then the evaluator should make a judgment on whether this may be related to misleading or broken links; if it is related to misleading links then the guidance indicates that this should be investigated by user testing but if it relates to broken links then this should be investigated by heuristic evaluation. If both misleading and broken links are indicated then the guidance suggests that these should be investigated	• The identification of specific usability problems on the site overall and on the specific pages on the site.

		using both methods.	
4	To redesign the e-commerce website and improve the usability problems identified in Step 3	1. For each usability problem identified in Step 3, suggest a recommendation on how to correct the problem. 2. Implement the suggested recommendations and re-design the website taking into consideration visitors' characteristic identified in Step 2. 3. Move back to Step 2 to track and measure the usage of the re-designed website	• A new design of the e-commerce website with improved usability.

5. Conclusion

A framework was developed to evaluate the usability of e-commerce websites which combines the use of GA software and the strategic use of user testing and heuristic evaluation methods. It is based on the comparison of the benefits and drawbacks of these methods in terms of the specific areas of usability problems that they could or could not identify on these types of websites.

The framework involves GA software as a preliminary step to provide a quick, easy and cheap indication of general potential usability problem areas on an e-commerce website and its specific pages. Then the framework enables evaluators to choose other method (s) to provide in-depth detail about specific problems on the site. Using the methods strategically could help to reduce both time and evaluation costs.

The suggested framework has managerial and academic implications. Regarding the managerial implication: E-commerce companies need to evaluate and improve their e-commerce websites in a way that will improve their success. The suggested framework is particularly useful for managers of e-commerce companies who might be interested in identifying usability problems on their sites and improving the design to meet users' needs. The framework, which explicitly clarifies the effectiveness of three usability evaluation methods, highlights the usefulness of the methods. It therefore helps e-commerce retailers to determine the usability method that best matches their need. It is expected that the framework will aid e-commerce companies in taking appropriate decisions regarding which usability method to apply and how to apply it in order to improve part or the overall usability of their websites, which could help increase their profitability.

Regarding the academic implications: This paper presents an evaluation of three e-commerce sites in Jordan as the basis for proposing a new approach for evaluating the usability of websites, specifically e-commerce sites. A particularly novel approach is the use of web analytics (Google Analytics software) as a component in the usability evaluation in conjunction with heuristics and user testing. This research has provided a detailed account of the use and evaluation of usability techniques for e-commence and a reflective account of the merits of individual approaches.

A limitation of the framework is that it was developed and tested using Jordanian users, experts and e-commerce sites. While it may have general applicability to e-commerce sites in developed countries, it has yet to be tested. It could be, for example, that users and/or heuristic evaluators in more developed countries would be able to identify different types of problems based on their greater experience. The suggested framework, therefore, has a particular value if applied to e-commerce sites in developing countries like Jordan and was an attempt to confront the challenging environment of e-commerce in such countries.

Despite the fact that multiple sites were used in determining the framework, and previous literature and studies were used extensively to determine the heuristic guidelines, user testing and web metrics, there is a potential second limitation of the framework. This is related to the ability of the methods to detect major issues that were not present on any of the three websites. The study compared the ability of the methods to detect problems that were present. However, this study does not consider usability issues not present on any of the examined sites. An extension of the study would be to set up a website with a set of known usability issues and apply the three methods.

The framework also offers a base for future research. Future research is needed to evaluate the applicability and usefulness of the framework in e-commerce companies in more developed countries. In particular, the extent to which the application of a framework which uses the three methods strategically rather than individually is able to reduce the time required to evaluate e-commerce websites should be investigated. Further research is also necessary to ensure that the component parts of the framework identify the expected specific usability problem areas when applied to more sophisticated e-commerce websites.

References

1. Agarwal, R. and Venkatesh, V. (2002), "Assessing a Firm's Web Presence: A Heuristic Evaluation Procedure for the Measurement of Usability", *Information Systems Research*, 13 (2), pp. 168–186.
2. Altman, DG. (1991), *Practical statistics for medical research*. London: Chapman and Hall.
3. Barnard, L. and Wesson J. (2004), "A trust model for e-commerce in South Africa", *In the Proceedings of SAICSIT 2004*, pp. 23-32.
4. Brinck, T. Gergle, D. and Wood S.D. (2001), *Usability for the web: designing websites that work*. Morgan Kaufmann Publishers.
5. Chen, S.Y. and Macredie, R.D. (2005), "An assessment of usability of electronic shopping: a heuristic evaluation", *International Journal of Information Management*, 25, pp. 516-532.
6. Claudio O. & Antonio, J., "A model for evaluating B2C e-commerce websites: Application in the CD e-retailing industry in Brazil". <http://www.oppapers.com/topics/website-evaluation/0>, [accessed 19.02.07].
7. Davidson R. & Lambert S. (2005), "Australian winery B2C websites: are they delivering the full bottle*?*". Flinders University, School of commerce research paper series, South Australia.
8. Desurvire, H., Lawrence, D., and Atwood, M. (1991), "Empiricism versus judgment: comparing user interface evaluation methods on a new telephone-based interface", *SIGCHI Bulletin*, 23 (4), pp. 58-59.
9. Doubleday, M. Ryan, M. Springett, A., and Sutcliffe (1997), "A comparison of usability techniques for evaluating design", *Proceedings of the 2nd conference on Designing interactive systems*, Amsterdam, The Netherlands, pp.101 - 110.
10. Elliott, S., Morup-Peterson A. & Bjorn-Andersen N. (2000), "Towards a framework for evaluation of commercial websites", *13th International Bled Electronic Commerce Conference*, Bled, Slovenia, June 19-21.
11. Fang, W. (2007), "Using Google Analytics for improving library website content and design: A case study", *Journal of Library Philosophy and Practice*, pp. 1-17.
12. Freeman, M.B. and Hyland, P. (2003), "Australian online supermarket usability", *Technical Report*, Decision Systems Lab, University of Wollongong.
13. Fu, L., Salvendy, G. and Turley, L. (2002), "Effectiveness of user testing and heuristic evaluation as a function of performance classification", *Behaviour & Information Technology*, 21(2), pp. 137-143.
14. Gray, W. and Salzman, C. (1998), "Damaged merchandise? a review of experiments that compare usability evaluation methods", *Human-Computer Interaction*, 13, pp. 203-261.
15. Hung, W. & McQueen, R.J. (2004), "Developing an evaluation instrument for e-commerce websites from 1st buyer's view point", *Electronic Journal of Information Systems Evaluation*, 7(1), pp. 31-42.
16. Inan, H. (2006), *Measuring the success of your website: a customer-centric approach to website management (Software Book)*, Hurol Inan.
17. Jana, S. and Chatterjee, S. (2004), "Quantifying website visits using web statistics: an extended cybermetrics study", *Online Information Review*, 28(3), pp. 91-199.
18. Jeffries, R., Miller, J.R., Wharton, C., and Uyeda, K.M. (1991), "User interface evaluation in the real world: a comparison of four techniques", *Proceedings ACM CHI'91 Conference*, New Orleans, LA, April 28-May 2, pp.119-124.
19. Kaushik (2007), *Web analytics, an hour a day*, Wiley Publishing, Inc.
20. Law, L. and Hvannberg, E. (2002), "Complementarity and convergence of heuristic evaluation and usability test: a case study of Universal Brokerage

Platform", *ACM International Conference Proceeding Series, 3,* Aarhus, Denmark, pp. 71-80.
21. Mariage, C. and Vanderdonckt, J. (2000), "A comparative usability study of electronic newspapers", *Proceedings of Int. Workshop on Tools for Working with Guidelines TFWWG'2000*.
22. McKinney, V., Yoon, K., and Zahedi, F. (2002), "The Measurement of Web-Customer Satis*faction: An Expectation and* Disconfirmation Approach", *Information Systems Research*, 13(2), pp. 296-315.
23. Nielsen, J. (2003), "Usability 101: introduction to usability", Useit.com. <useit.com/alertbox/20030825.html>
24. Nielsen, J. and Mack, R.L (1994), *Usability inspection methods* (Eds.), John Wiley & Sons, New York.
25. Nielsen, J. (2000), *Designing web usability: the practice of sim*plicity. New Riders Publishing.
26. Nielsen, J. and Norman, D. (2000), "Web-Site usability: usability on the web isn't a luxury", *Information Week*, http://www.informationweek.com/773/web.htm
27. Obeidat, M. (2001), "Consumer protection and electronic commerce in Jordan (an exploratory study)", *In Proceedings of the Public Voice in Emerging Market Economies Conference*, Dubai, UAE.
28. *Oppen*heim C. & Ward L. (2006), "Evaluation of websites for B2C e-commerce", *Aslib Proceedings: New Information Perspectives*, 58(3).
29. Ong, K., Kingham, B., Sotiridy, K., Kaufman, D., Polkowski, M., Schofield, J. (2003), "Web presence of an integrated delivery system at year one: lessons learned", *International Journal of Medical Informatics*, 70, pp. 1-18.
30. Peacock, D. (2003), "Statistics, structures & satisfied customers: using web log data to improve site performance", *in the Proceedings of Museums and the Web 2002*, Boston.
31. Pearrow, M. (2000), *Website usability handbook*. Charles River Media.
32. Pearson, J. M., Pearson, A. & Green, D. (2007), "Determining the importance of key criteria in web usability", *Management Research News*. 30(11), pp. 816-828.
33. Peterson, E. (2004), *Web analytics demystified*, Celilo Group Media and CafePress,.
34. Peterson, E.T. (2006), *The big book of key performance indicators*, First Edition, , <http://www.webanalyticsdemystified.com>
35. Phippen, L. Sheppard, S. Furnell (2004), "A practical evaluation of web analytics", *Internet Research*, 14(4), pp. 284-293.
36. Pickard, A. (2007), *Research methods in information*. Facet. London.
37. Prom (2007), "Understanding on-line archival use through web analytics", ICA-SUV Seminar, Dundee, Scotland,. http://www.library.uiuc.edu/archives/workpap/PromSUV2007.pdf
38. Sharp, H., Rogers, Y. & Preece, J. (2007), *Interaction Design: Beyond Human-Computer Interaction*. Wiley; Second Edition.
39. Tan, W., Liu, D., Bishu, R. (2009), "Web evaluation: heuristic evaluation vs. user testing", *International Journal of Industrial Ergonomics*, 39, pp. 621-627.
40. Tarafdar, M. & Zhang, J. (2005), "Analyzing the Influence of Website Design Parameters on Website Usability", *Information Resources Management Journal*. 18(4), pp. 62 – 80.
41. Tilson, R., Dong, J., Martin, S., Kieke, E. (1998), "Factors and principles affecting the usability of four e-commerce sites", *4th Conference on Human Factors and the Web (CHFW)*, AT&TLabs, USA.

42. Van der Merwe, R. & Bekker, J. (2003), “A framework and methodology for evaluating e-commerce websites”, *Internet Research*: *Electronic Networking Applications and Policy*, 13(5), pp. 330-341.
43. Web Analytics Association, [n.d.]. http://www.webanalyticsassociation.org
44. Xue, S. (2004), “Web usage statistics and website evaluation”, *Online Information Review*, 28(3), pp. 180-190.
45. White, G.K. & Manning B.J. (1998), “Commercial WWW site appeal: how does it affect online food and drink consumers’ purchasing behaviour?”, *Internet Research: Electronic Networking Application and Policy*, 8(1), pp. 32-38.
46. Yeadon, J. (2001), “Website statistics”, *Vine*, 31(3), pp. 55-60.

Appendix: Summary of the Specific Problem Areas and Sub-areas Identified by the User Testing and Heuristic Evaluation Methods and their Severity Level

Usability Problem Area	Usability Problem Sub-Area	User Testing		Heuristic Evaluation	
		Minor Problems	Major Problems	Minor Problems	Major Problems
Navigation Problems	Misleading links	√ **$(0^1)(1^2)$**	√√ **$(5^3)(0^4)$**	√√ **$(14^5)(1^6)$**	**$(0^7)(0^8)$**
	Links were not obvious	**$(0^1)(0^2)$**	√√ **$(2^3)(2^4)$**	√√ **$(13^5)(0^6)$**	√ **$(1^7)(2^8)$**
	Broken links	√ **$(0^1)(3^2)$**	**$(0^3)(0^4)$**	√√ **$(3^5)(3^6)$**	**$(0^7)(0^8)$**
	Weak navigation support	**$(0^1)(0^2)$**	√ **$(0^3)(1^4)$**	**$(0^5)(0^6)$**	√√ **$(2^7)(1^8)$**
	Orphan pages	√ **$(0^1)(1^2)$**	**$(0^3)(0^4)$**	√√ **$(7^5)(1^6)$**	**$(0^7)(0^8)$**
Internal Search Problems	Inaccurate results	**$(0^1)(0^2)$**	√√ **$(0^3)(2^4)$**	√√ **$(1^5)(0^6)$**	√√ **$(0^7)(2^8)$**
	Limited options	√√ **$(0^1)(2^2)$**	**$(0^3)(0^4)$**	√√ **$(0^5)(2^6)$**	**$(0^7)(0^8)$**
	Poor visibility of search position	**$(0^1)(0^2)$**	**$(0^3)(0^4)$**	√√ **$(1^5)(0^6)$**	**$(0^7)(0^8)$**
Architecture Problems	Poor structure	**$(0^1)(0^2)$**	√√ **$(0^3)(1^4)$**	**$(0^5)(0^6)$**	√√ **$(0^7)(1^8)$**
	Illogical order of menu items	**$(0^1)(0^2)$**	**$(0^3)(0^4)$**	√√ **$(2^5)(0^6)$**	**$(0^7)(0^8)$**
	Illogical categorisation of menu items	**$(0^1)(0^2)$**	**$(0^3)(0^4)$**	√√ **$(1^5)(0^6)$**	**$(0^7)(0^8)$**
Content Problems	Irrelevant content	√ **$(0^1)(4^2)$**	√ **$(0^3)(2^4)$**	√√ **$(16^5)(4^6)$**	√√ **$(1^7)(2^8)$**
	Inaccurate information	√√ **$(0^1)(2^2)$**	**$(0^3)(0^4)$**	√√ **$(0^5)(2^6)$**	√√ **$(1^7)(0^8)$**
	Grammatical accuracy problems	**$(0^1)(0^2)$**	**$(0^3)(0^4)$**	√√ **$(2^5)(0^6)$**	**$(0^7)(0^8)$**
	Missing information about the company	**$(0^1)(0^2)$**	**$(0^3)(0^4)$**	√√ **$(2^5)(0^6)$**	**$(0^7)(0^8)$**
	Missing information about the products	√ **$(0^1)(3^2)$**	**$(0^3)(0^4)$**	√√ **$(10^5)(3^6)$**	**$(0^7)(0^8)$**
Design Problems	Misleading images	√ **$(0^1)(1^2)$**	**$(0^3)(0^4)$**	√√ **$(5^5)(1^6)$**	**$(0^7)(0^8)$**
	Inappropriate page design	√ **$(0^1)(2^2)$**	√√ **$(2^3)(1^4)$**	√√ **$(9^5)(2^6)$**	√ **$(0^7)(1^8)$**
	Unaesthetic design	**$(0^1)(0^2)$**	**$(0^3)(0^4)$**	√√ **$(3^5)(0^6)$**	**$(0^7)(0^8)$**
	Inappropriate quality of images	**$(0^1)(0^2)$**	**$(0^3)(0^4)$**	√√ **$(1^5)(0^6)$**	**$(0^7)(0^8)$**

	Missing alternative texts	$(0^1)(0^2)$	$(0^3)(0^4)$	√√ $(4^5)(0^6)$	$(0^7)(0^8)$
	Broken images	$(0^1)(0^2)$	$(0^3)(0^4)$	√√ $(10^5)(0^6)$	$(0^7)(0^8)$
	Inappropriate choice of fonts and colours	√ $(0^1)(1^2)$	$(0^3)(0^4)$	√√ $(4^5)(1^6)$	$(0^7)(0^8)$
	Inappropriate page titles	$(0^1)(0^2)$	$(0^3)(0^4)$	√√ $(3^5)(0^6)$	$(0^7)(0^8)$
Purchasing Process Problems	Difficulty in knowing what was required for some fields	√√ $(1^1)(1^2)$	$(0^3)(0^4)$	√ $(0^5)(1^6)$	$(0^7)(0^8)$
	Difficulty in distinguishing between required and non-required fields	$(0^1)(0^2)$	√√ $(3^3)(0^4)$	$(0^5)(0^6)$	$(0^7)(0^8)$
	Difficulty in knowing what links needed to be clicked	$(0^1)(0^2)$	√√ $(3^3)(0^4)$	$(0^5)(0^6)$	$(0^7)(0^8)$
	Long ordering process	√√ $(0^1)(1^2)$	$(0^3)(0^4)$	√√ $(0^5)(1^6)$	$(0^7)(0^8)$
	Session problem	$(0^1)(0^2)$	√√ $(0^3)(1^4)$	$(0^5)(0^6)$	√√ $(0^7)(1^8)$
	Not easy to log on to the site	$(0^1)(0^2)$	$(0^3)(0^4)$	$(0^5)(0^6)$	√√ $(1^7)(0^8)$
	Lack of confirmation if users deleted an item from their shopping cart	$(0^1)(0^2)$	$(0^3)(0^4)$	$(0^5)(0^6)$	√√ $(3^7)(0^8)$
	Long registration page	$(0^1)(0^2)$	$(0^3)(0^4)$	$(0^5)(0^6)$	√√ $(1^7)(0^8)$
	Compulsory registration	$(0^1)(0^2)$	$(0^3)(0^4)$	$(0^5)(0^6)$	√√ $(2^7)(0^8)$
	Illogical required fields	√√ $(0^1)(2^2)$	$(0^3)(0^4)$	√√ $(0^5)(2^6)$	$(0^7)(0^8)$
	Expected information not displayed after adding products to cart	√√ $(1^1)(0^2)$	√√ $(1^3)(0^4)$	$(0^5)(0^6)$	$(0^7)(0^8)$
Security and Privacy Problems	Lack of confidence in security and privacy	$(0^1)(0^2)$	$(0^3)(0^4)$	$(0^5)(0^6)$	√√ $(1^7)(0^8)$
Accessibility and Customer Service Problems	Not easy to find help/customer support information	$(0^1)(0^2)$	√√ $(3^3)(0^4)$	√√ $(1^5)(0^6)$	$(0^7)(0^8)$
	Not supporting more than one language	√√ $(0^1)(2^2)$	$(0^3)(0^4)$	√√ $(0^5)(2^6)$	$(0^7)(0^8)$
	Not supporting more than one currency	$(0^1)(0^2)$	$(0^3)(0^4)$	√√ $(2^5)(0^6)$	$(0^7)(0^8)$
	Inappropriate information provided within a help section/customer service	√ $(0^1)(1^2)$	$(0^3)(0^4)$	√√ $(1^5)(1^6)$	$(0^7)(0^8)$
	Not supporting the sending of comments from customers	$(0^1)(0^2)$	$(0^3)(0^4)$	√√ $(2^5)(0^6)$	$(0^7)(0^8)$

	Not easy to find and access the site from search engines	**$(0^1)(0^2)$**	**$(0^3)(0^4)$**	√√ **$(2^5)(0^6)$**	**$(0^7)(0^8)$**
Inconsistency Problems	Inconsistent page layout or style/colours/ terminology/content	√ **$(0^1)(1^2)$**	**$(0^3)(0^4)$**	√√ **$(21^5)(1^6)$**	**$(0^7)(0^8)$**
Missing capabilities	Missing functions/information	√ **$(0^1)(1^2)$**	**$(0^3)(0^4)$**	√√ **$(19^5)(1^6)$**	**$(0^7)(0^8)$**

√√: **Good identification of the specific problem area**
√: **Missed identification of some of the specific problem areas**
Blank: **Could not identify the specific problem area**

1: Number of minor usability problems uniquely identified by the user testing method

2: Number of minor usability problems commonly identified by the user testing and heuristic evaluation methods

3: Number of major usability problems uniquely identified by the user testing method

4: Number of major usability problems commonly identified by the user testing and heuristic evaluation methods

5: Number of minor usability problems uniquely identified by the heuristic evaluation method

6: Number of minor usability problems commonly identified by the user testing and heuristic evaluation methods

7: Number of major usability problems uniquely identified by the heuristic evaluation method

8: Number of major usability problems commonly identified by the user testing and heuristic evaluation methods