Association for Information Systems
AIS Electronic Library (AISeL)

CONF-IRM 2016 Proceedings

International Conference on Information Resources Management (CONF-IRM)

2016

# The value of comparative usability and UX evaluation for e-commerce organisations

Rendani Maarten Kruger
*University of Pretoria*, rendi619@gmail.com

Helene Gelderblom
*University of Pretoria*, helene.gelderblom@up.ac.za

Wynand Beukes
*University of Pretoria*, wynand.beukes@up.ac.za

# 71. The value of comparative usability and UX evaluation for e-commerce organisations

Rendani Maarten Kruger[1]
University of Pretoria
rendi619@gmail.com

Helene Gelderblom[1]
University of Pretoria
helene.gelderblom@up.ac.za

Wynand Beukes[1]
University of Pretoria
wynand.beukes@up.ac.za

## *Abstract*

The objective of this paper is to investigate the possible value of comparative user experience (UX) or usability evaluations for e-commerce organisations. Poor website usability has been identified as one of the main reasons why users abandon potential transactions. Appropriate evaluation of these sites is therefore essential. A problem with usability and UX evaluation is the lack of trust designers have in the evaluator's recommendations due to the subjective nature thereof. This paper investigates the possible enhancement of the objectivity of such evaluations through cross-company comparative evaluations, so that designers can assess their design success against that of direct competitors in the market. We conducted an empirical, comparative evaluation of three similar organisations' e-commerce websites using eye tracking as the primary data collection mechanism, and then demonstrated the potential value and usefulness of the outcomes.

## *Keywords*

Web Design, User Experience, Usability, Comparative Evaluation, Eye Tracking, Telecommunications Websites

## 1. Introduction

In an investigation into why a large telecommunications organisation failed to invest in appropriate usability and user experience evaluation of their website despite the website's obvious importance for successful business, Beukes (2015) established that there is a general disregard for traditional good web design and evaluation practice amongst designers in the organisation. Very little attention was given to proper user experience (UX) and usability evaluation. The broader problem that we investigate here is how such organisations can be convinced of the value that UX evaluation and how this could contribute to their business success.

---

[1] Department of Informatics

Flawed website design is still a major barrier for successful e-commerce (Sivaji et al. 2011). A survey by Sivaji et al. amongst 10,000 online shoppers revealed that 30% of the respondents regarded poor website organisation as the main reason for abandoning potential transactions.

The value of evaluation is often obscured by negative influences of doubt that are rooted in the potentially subjective nature of evaluator recommendations (Law, van Schaik & Roto 2012). Also, UX is a theoretical concept that is difficult to operationalise in a way that prescribes what exactly should be measured when evaluating it (Law et al. 2012). An evaluation method that relies clearly on factual information may thus be of great value. UX and usability evaluation through eye tracking is one approach that endeavours to provide objective data about user interaction behaviour that, if analysed accurately and appropriately, could provide convincing evidence and design recommendations (Djamasbi 2014). Designers can however still discard these results as inaccurate or prone to evaluator bias if they are strongly committed to their design ideas.

This paper investigates a method that uses factual user behaviour data that are collected from interaction with the website to be evaluated, as well as data from user interaction with closely competing websites that offer directly comparable services. In this study we evaluated the websites of three similar telecommunications organisations using eye tracking as the main data collection method. Our aim was to determine if comparison of the evaluation results can overcome some of the reservations about the evaluation process and outcomes.

Although we believe our findings apply to UX and usability evaluation in general, our emphasis in this empirical case study was on the usability aspects of web design rather than on the more affective UX aspects. We do, however, see these concepts as inseparable.

# 2. Background

## 2.1 The value of usability and user experience evaluation

The ISO defines usability as the extent to which users can use a product to achieve their goals effectively, efficiently and with satisfaction within a specific context of use (Travis 2013). User experience (UX) refers to the experience(s) that result from encountering technology systems (Roto et al. 2010). UX includes encounters with systems – not only active, personal use, but also in a more passive way, such as observing someone else using a system. UX is unique to an individual, it is influenced by prior experiences and expectations, and is rooted in a social and cultural context.

Whereas traditional usability factors focus on performance and seamless interaction, UX factors are more concerned with affect, interpretation and meaning (Roto et al. 2010). Usability and UX are however very closely tied – if an online shopper cannot complete a transaction successfully because of usability issues, the UX will not be positive. This is confirmed by Rogers, Hutchinson and Fu (2010) who include ‘task successes’ as one of four crucial metrics in their UX evaluation framework.

It is difficult to quantify the value of UX and usability evaluation in comparison with its cost in terms of time and money, but researchers have for many years been trying to put a figure on the return on investment in proper user evaluation (Landauer, 1995; Heppner et al. 2005;

Weinschenk 2005). The results of launching flawed products often manifest in reduced usage which in turn can significantly impact on expected business. Studies have estimated that 80% of the total system maintenance costs incurred have been related to users having problems with the systems and not with a system's technical flaws; of this, 64% was directly related to system usability problems (Landauer 1995; Mentis & Gay 2003; Weinschenk 2005). Weinschenk quotes real world examples to explain the value of good usability and UX. One example shows that once a system is in development, it will cost 10 times more to fix the problem than when the same problem was solved during design. If the system has been released, this cost rises to 100 times more. In another example, a large computer company spent $20,700 on usability work to improve the sign-on procedure in a system used by several thousand people. The improvement in productivity was calculated to save the company $41,700 only on the first day the system was used. The benefit within the first year was $6,800,000.

According to Lee (2012), the user's virtual experience when using a company's technology will influence the strength of brand influence and the overall perception that the person has of that organisation. Clients can be attracted and retained by improving the performance of a website. E-commerce sites commonly drive away nearly half of repeat traffic by not making it easy for visitors to find the information they need (Weinschenk, 2005). Weinschenk reports on a study of an e-commerce site where first time users spent an average of $127 per purchase, while repeat users spent nearly twice that. Usable e-commerce sites build goodwill.

## 2.2 The purpose of comparative studies for organisations

Oxforddictionaries.com (2015) defines comparative as "measured or judged by estimating the similarity or dissimilarity between one thing and another". In reports on comparative usability studies available in the literature, different entities are used as the objects of comparison. Some studies compare different evaluation methods (Molich & Dumas 2008 and Ssemugabi & De Villiers 2007), some compare different implementations of the same system or prototype (Zhang & Moore 2014) and others compare different user groups using the same system (Arianezhad et al. 2013). Reports on studies that compare e-commerce applications of competing organisations are scarce. When initiating a business, a goal is to identify business potential in the current relevant markets. This requires a comparative study whereby the new business owners will compare the macro and micro environments of markets at the present time, to the potential macro and micro environments of the markets in the future (Pîndiche & Ionita 2013). A cross-company comparative study can reveal where the planned business stands in terms of its offering and how this overlaps with, or extends, the business of potential competitors (Czepiel & Kerin 2012).

From the above, when e-commerce is involved, comparison of competing organisations' online strategy and presence is essential (Czepiel & Kerin 2012; Weinschenk, 2005). The focus in our study is on how a cross-company comparative UX and usability evaluation of e-commerce websites can benefit a business. We define a cross-company comparative UX or usability study as the act of examining, in detail, the similarity or dissimilarity between the scientifically observed facts about user behaviour when using the technological interfaces of different organisations that have comparable purposes.

Any study that compares entities should follow some basic principles to ensure that the comparison is valid. Translating De Zepetnek's (1998) principles of a comparative literature study to a comparative UX or usability evaluation yields the following guidelines:

- Comparable entities should be included in the study; this means the entities must serve a similar purpose, operate within similar environments and be based on similar backgrounds.
- The tests for the various entities should be conducted in a similar way. This entails using the same tools, procedures and variables.
- The results from the tests must be comparable in nature.

Against the background described above, we can now formulate the research objective of the study reported in this paper and articulate it in terms of a research question.

### 2.3 Research objective and question

Based on the purpose of comparative studies for organisations and the potential value of usability and user experience evaluations, the objectives of this study were:

1. To demonstrate how a cross-company comparative UX or usability evaluation (as opposed to distinct evaluations) could be conducted.
2. To investigate the results of such a study to determine what value (if any) is added through the comparison.

This was done by comparing the websites of three large organisations that conduct their business through their respective company websites, and showing how this can increase the perceived value of the evaluation results for the designers of the respective websites. First, we had to identify a suitable business sector, then choose at least three organisations whose websites provide similar services to customers with a similar profile. Next, we did separate evaluations focusing on one specific service offered by all three websites. Finally, we investigated whether a comparative analysis of the evaluation results provides more value than considering the three evaluations in isolation.

The question that we asked is: What value can e-commerce organisations derive from cross-company comparative UX and usability evaluation studies?

## 3. Methodology

This was an empirical evaluation study conducted in a usability laboratory. We chose a case study for our research design. We start this section by describing the case study and then the participants before explaining how data was collected and analysed.

### 3.1 The case study

We used a comparative case study (Yin 2003). Three prominent telecommunications organisations, all based in South Africa, with their respective client bases dispersed throughout Africa were selected for the study. We refer to them anonymously as TelecomA, TelecomB and TelecomC as their specific identities are irrelevant in the context of this research. Being a mobile service provider is one of the core business functions of all three organisations. Their business success relies greatly on their respective websites where their clients can perform similar tasks.

One of the core functions of a mobile service provider – mobile data top-up – was selected as the focus in this comparative evaluation.

Following exactly the same procedure in each sub-case, we conducted an evaluation of each website to determine how well they have designed their top-up functionality respectively, in terms of usability and UX.

## 3.2 Participants

For each organisation, fifteen to twenty users were recruited to participate in a user experience evaluation experiment. A combination of convenience and snowball sampling was used. The Tobii Studio software used to record their eye tracking data indicates the accuracy of the captured data and the data of participants with accuracy lower than 60% were not included in the data set. Table 1 describes the participant groups whose data were included. Two studies included 15 users and one 14. A samples size of 16±4 is accepted as adequate for usability studies (AlRoobaea & Mayhew 2014).

The participant profiles were similar. There was an even distribution between males and females, with slightly more females in each study. The age distribution and average ages were also comparable across companies. All participants use computers on a daily basis and none of them had exposure to the particular web site they were allocated to use during the evaluation (i.e. they had never been clients of that specific organisation).

| | TelecomA | TelecomB | TelecomC |
|---|---|---|---|
| Number of participants | 15 | 14 | 15 |
| Gender distribution | 8=F, 7=M | 9=F, 5=M | 9=F, 6=M |
| Age range | 14-60 | 14-42 | 20-47 |
| Average age | 33 | 26 | 30 |

**Table 1**: Demographic information about participants

## 3.3 Data collection

Demographic data about users were collected prior to the evaluation task using a simple questionnaire that was either self-administered or evaluator-administered depending on the participant's preference.

User experience and usability data were gathered through eye tracking and informal post-test interviews. The interviews were only conducted when the evaluators observed specific behaviours during the interaction that needed clarification. Eye tracking is a method to record people's eye movements while they are looking at a stimulus. In UX and usability evaluation, it provides an objective measure of the users' attention on interface elements throughout the interaction period (Duchowski 2007). Capturing users' gaze patterns (i.e. saccades and fixation points) provides accurate information on what the areas of focus were and which parts of the interface they ignored.

The experiments were conducted at the University of Pretoria. We used a Tobii T120 desktop eye tracker to record users' eye movements while they completed the same task on each of the three websites. The eye tracker was calibrated for each participant's eyes to ensure accurate data

recording. In this study a five-point calibration was used. Once calibration was completed the instructions were displayed to the participants, followed by the task they were required to complete. They performed a task known as "mobile data top-up" or "purchasing of a data bundle" after the following scenario was explained to them: "You just bought a new mobile SIM card. After arriving home you insert the SIM card into your phone and realise that you are not able to browse the Internet because you do not have any data available. On your computer, you open the TelecomX website to top-up your SIM card with mobile data". When they started the task we had already logged into the website so that they did not need to go through the login process. The task ended when they indicated that they have located the top-up function.

After the eye-tracking recording, data was exported with the Tobii Studio software. The data included static gaze plots indicating users' eye movements across selected pages of the respective websites (see Figure 1 for an example); time taken to complete the task, time and number of fixations until first click, time spent on the final screen until the top-up function was selected, and the number of pages visited during the task. These are standard metrics used in usability and UX evaluation.

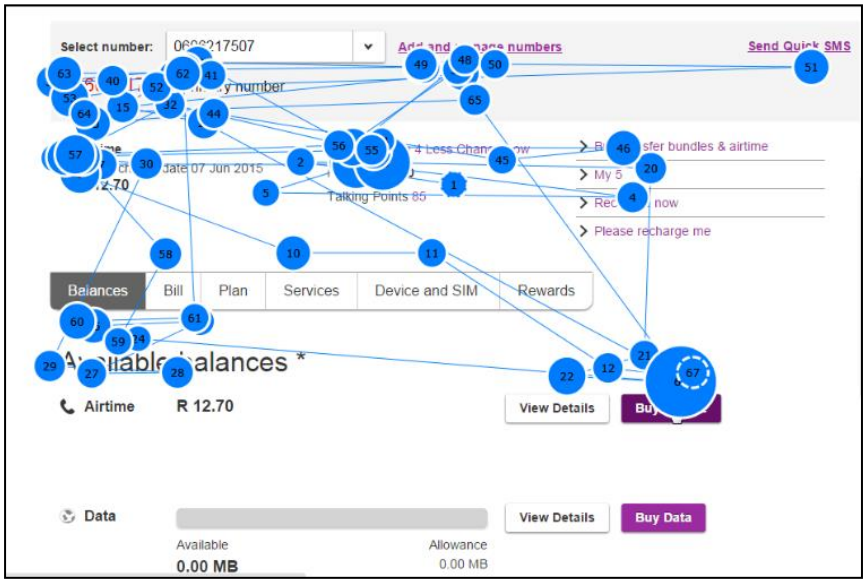


**Figure 1:** Example of a gaze plot

## 3.4 Data analysis

Eye tracking data were exported and analysed using the Tobii Studio software. This included quantitative metrics such as task completion time, number of fixations, time to first click, time spent on the home screen and time spent on the top-up screen. We transferred this data to an Excel spread sheet where descriptive statistics including minimum, maximum, average and median values were calculated for each set of metrics, for each of the three organisations. The results for the three organisations were then compared and the comparative data summarised in tables and graphs (see section 4 below).

# 4. Results of the comparative evaluation

The results are discussed with reference to time to complete the task, time and number of fixations until first click, time on destination screen until top-up is selected and the number of pages visited.

## 4.1 Total time to complete the task

Table 2 gives the time to complete the task on all three web sites in terms of the slowest user, the quickest user and the average and median times. Although the slowest respective users on the three web sites took equally long, there is a notable difference between the median and average completion times across the websites (see Figure 2). It took users of the TelecomC web site an average of 44 seconds to complete the task, while the average completion time for TelecomA was 3 minutes 14 seconds. This indicates a problem with TelecomA's site, especially if the competition allows users to complete the task in less than one minute.

| Partcipant | TelecomA | TelecomB | TelecomC |
|---|---|---|---|
| Slowest | 05:51 | 05:30 | 04:20 |
| Average | 03:14 | 01:35 | 00:44 |
| Median | 03:28 | 01:28 | 00:25 |
| Quickest | 00:50 | 00:21 | 00:14 |

**Table 2:** Time to complete the task

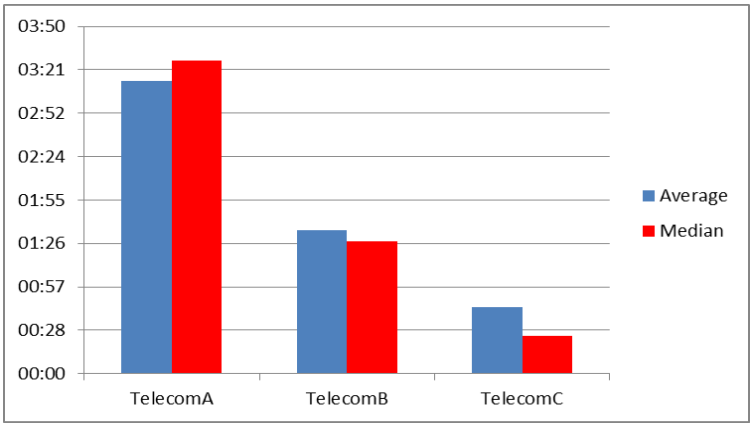


**Figure 2:** Average and Median time to complete the task

## 4.2 Time and number of fixations until first click

Table 3 gives the time it took users to make a decision to click on the home page, as well as the number of fixations they had on the screen up to the first click. Figure 3 compares the number of fixations. This data do not demonstrate significant differences in terms of the median and average times until first click, but there were a large number of fixations before first click (more than the competition) on TelecomB's website. An examination of the gaze plots of TelecomB's home page, confirmed that the users looked at a large range of elements before deciding on a route to take to complete the task. The number of elements on the TelecomB home screen may therefore be one of the reasons why there were so many fixations and so much time spent on the home screen before first click.

This data can however not be analysed in isolation because the first click may not have been on target, and this could lead users astray in the remainder of the task. In section 4.4 we look at the navigation paths which are more meaningful in this context.

| | TelecomA | TelecomB | TelecomC |
|---|---|---|---|
| | Time (Fixations) | Time (Fixations) | Time (Fixations) |
| Minimum | 00:00:06 (117) | 00:00:02 (7) | 00:00:02 (5) |
| Maximum | 00:00:33 (48) | 00:01:23 (244) | 00:00:20 (67) |
| Median | 00:00:10 (36) | 00:00:19 (57) | 00:00:11 (36) |
| Average | 00:00:14 (12) | 00:00:24 (71) | 00:00:11 (34) |

**Table 3:** Time (hour:min:sec) and fixations until first click

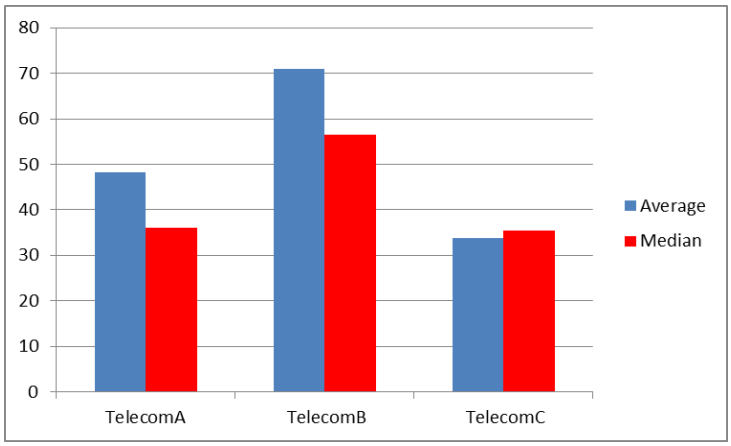


**Figure 3:** Fixations until first click

## 4.3 Time on target screen until top-up is selected

Here we compare how long it took users to locate the target on the final screen. As can be seen in Table 4, the slowest user on TelecomA's web site took just over three minutes, while the slowest users on TelecomB and TelecomC took 26 and 12 seconds respectively. This indicates a usability problem on TelecomA's website.

The quickest user on the TelecomA's top-up page took 26 seconds and the median time on that page was 41 seconds. This may set the designers at ease with regard to the slowest user, but comparing this time to the average and median times on the TelecomB and TelecomC web sites (see Figure 4) proves that there are probably usability problems on the TelecomA top-up page.

| Partcipant | TelecomA | TelecomB | TelecomC |
|---|---|---|---|
| Slowest | 03:06 | 00:26 | 00:12 |
| Median | 00:41 | 00:09 | 00:07 |
| Quickest | 00:26 | 00:04 | 00:04 |

**Table 4:** Time on last screen until top-up

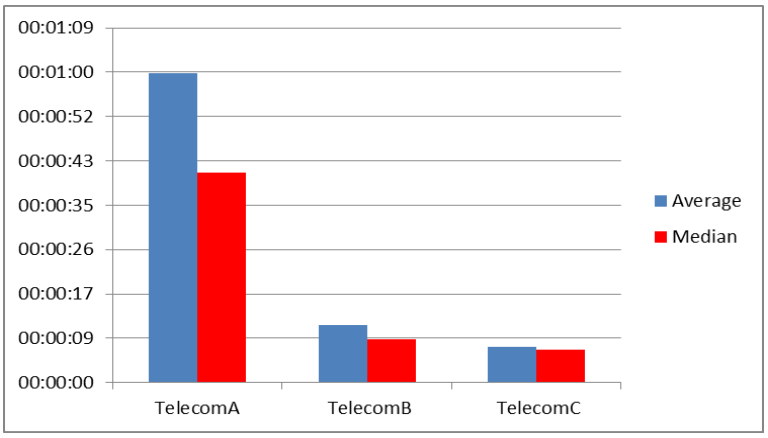


**Figure 4:** Time on last screen until top-up

The average number of pages visited by the TelecomA users is 8.87, while the averages for TelecomB and TelecomC are 3.36 and 2.57 respectively. This may be an indication of a problem with navigation on the TelecomA web site.

There is a correlation between the number of pages visited and the time spent completing the tasks. The more pages visited, the more the time spent completing the task. Assuming that time can be seen as an indication of the amount of effort it takes to complete a task then the average user put in the biggest effort to complete the same task on TelecomA's website. A supporting indication of this is the route taken to complete the task and the number of pages revisited.

| Partcipant | TelecomA | TelecomB | TelecomC |
|---|---|---|---|
| Most | 16 (1) | 10 (1) | 10 (1) |
| Average | 8.87 | 3.36 | 2.57 |
| Optimum | 4 (3) | 2 (5) | 2 (13) |

**Table 5:** Number of pages visited during the task (the number of users involved appear in brackets)

The results of the evaluation provide clear evidence that the TelecomA web site may have severe usability problems with their data top-up functionality on the website when compared to the TelecomB, and especially, the TelecomC websites.

# 5. Value added by the comparative evaluation

When considering the results of the evaluation of each of the individual websites in isolation, an experienced UX and usability evaluator would be able to identify problem areas in the design. For example, the number of screens that users visited on the TelecomA website to reach the top-up screen clearly indicates navigational problems, even without comparing it to the other websites. Also, the time taken to complete the task on the TelecomA website is an indication that there are usability issues. It would however be much easier for the evaluator to convince the TelecomA web designers that there are serious problems when the comparative data is presented together with the data about their own website. When they have the factual evidence that users of competing organisations find it easier to perform a function, they would be less likely to discard the evaluation results based on suspected evaluator bias or personal opinion.

The added value of the comparative results is further illustrated when considering user performance on the TelecomB website. Comparing their results with that of TelecomA will boost the organisation's confidence in the success of their web design. However, when compared to the results of TelecomC, the results appear less positive for TelecomB.

For TelecomC the outcomes were very positive. Their obvious superior results compared to the other two websites provide objective evidence that they have made design decisions that support usability – in particular with regard to the top-up functionality.

We recommend that comparative UX and usability studies include at least three competing organisations. Although we only tested one function – mobile data top-up – in our comparative evaluation and still obtained very useful comparative data, we would recommend that more than one function is tested for a more complete comparison.

# 6. Conclusions

This paper reported on an empirical study to demonstrate the value of a comparative usability study through a case study. The case involved three telecommunications organisations with similar lines of business and the same functionality was tested on all three websites with users with comparable demographic profiles.

Using eye-tracking technology, similar user behaviour data was collected for each website, including time spent to complete the task, time until first click, the number of fixations until first click and the number of pages visited. The data were then summarised to facilitate comparison across the organisations and analysed.

TelecomA's results revealed clear usability and navigational issues, especially in comparison to the results of TelecomB and TelecomC. TelecomB users had more fixations on parts of the website indicating that there are more screen elements than on the other two sites. Although TelecomB fared well when compared to TelecomA, its results show that TelecomC supports the top-up functionality much more effectively.

Although we did not report on the specific usability problems identified in the poorly designed websites, the next step from an organisation's point of view would be to identify and address the specific design issues. The added value of comparing UX and usability evaluation results across organisations has been successfully illustrated. The results from studies like these could be further used to set up benchmarks and improve the general ease of use of e-commerce websites.

## References

AlRoobaea, R. and P.J. Mayhew (2014) "How many participants are really enough for usability studies?", In *Science and Information Conference (SAI)*, IEEE 2014, pp. 48-56.

Beukes, W. (2015) *Why Website designers disregard design guidelines that support usability*. M.IT Dissertation, Dept of Informatics, University of Pretoria, South Africa.

Czepiel, J. A., and R. A. Kerin (2012) *Competitor analysis*, Handbook of Marketing Strategy, vol. 41.

De Zepetnek, S. T. (1998) Comparative literature: Theory, method, application. Rodopi.

Djamasbi, S. (2014) "Eye tracking and web experience", *AIS Transactions on Human-Computer Interaction*, 6(2), pp. 16–31.

Heppner, C.C., Kates, J., Lynch, J.S., and R.R. Moritz (2005) "At Sprint, Understanding the Language of Business Gives Usability a Positive Net Present Value", in *Bias*, R.G and Mayhew, D.J., *Cost-Justifying Usability, An Update for an Internet Age*, 2nd edition, pp. 601–611.

Landauer, T. K. (1995) *The Trouble with Computers*. MIT Press.

Law, E. L. C., van Schaik, P. and V. Roto (2012) "To Measure or not to Measure UX: An Interview Study", *I-UxSED*, pp. 58-63.

Lee, K. (2012) "Consumer processing of virtual experience in e-commerce: A test of an integrated framework", *Computers in Human Behavior*, (28)6, pp. 2134-2142.

Mentis, H. and G. Gay. (2003) "User recalled occurrences of usability errors: implications on the user experience", *Extended Abstracts of the 2003 Conference on Human Factors in Computing Systems (CHI)*. ACM Press, pp. 736–737.

Molich, R. and J.S. Dumas (2008) "Comparative usability evaluation (CUE-4)", *Behaviour & Information Technology*, (27)3, pp. 263-281.

Oxforddictionaries.com (2015) *Oxford Dictionaries - Dictionary, Thesaurus, & Grammar*. [online] Available at: http://www.oxforddictionaries.com/ [Accessed 30 Oct. 2015].

Pîndiche, E. and R. Ionita (2013) "The Influence of Micro and Macro Environment Components on Trade Companies in Romania", *Journal of Knowledge Management, Economics and Information Technology*, (1)1.

Rodden, K., Hutchinson, H. and X. Fu (2010) "Measuring the User Experience on a Large Scale: User-Centered Metrics for Web Applications", *Proceedings of CHI 2010*, pp. 2395-2395.

Roto, L, Law, E., Vermeeren, A., and J. Hoornhout (2010) "User experience white paper. Bringing clarity to the concept of user experience", Result from Dagstuhl Seminar on Demarcating User Experience, September 15-18, 2010.

Sivaji, A., Downe, A.G., Mazlan, M.F., Soo, S.T. and A. Abdullah (2011). "Importance of incorporating fundamental usability with social and trust elements for e-commerce website", *International Conference on Business, Engineering and Industrial Applications (ICBEIA)*, IEEE, pp. 221-226.

Ssemugabi, S. and R. De Villiers (2007) "A comparative study of two usability evaluation methods using a web-based e-learning application", *Proceedings of the 2007 annual research conference of the South African institute of computer scientists and information technologists SAICSIT*, ACM, pp. 132-142.

Travis, D. (2013) *ISO 9241 for Beginners*. Ninth. London: Userfocus Ltd.

Weinschenk, S. (2005) *Usability a business case.* White Paper, Human Factors International.

Yin, R.K. (2003) *Case study research: Design and methods*. SAGE.

Zhang, S. and A.B. Moore (2014) "The Usability of Online Geographic Virtual Reality for Urban Planning", In *Innovations in 3D Geo-Information Sciences*, Springer International Publishing, pp. 225-242.