Comparing Effectiveness, Efficiency, Ease of Use, Usability and User Experience When Using Tablets and Laptops

Werner Wetzlinger, Andreas Auinger, and Michael Dörflinger

University of Applied Sciences Upper Austria, Campus Steyr
{werner.wetzlinger,andreas.auinger}@fh-steyr.at,
michael.doerflinger@outlook.com

Abstract. Initially perceived as a consumer device, in recent years tablets have become more frequently used in business contexts where they often replace laptops as mobile computing devices. Since they follow different user interaction paradigms we conducted a study comparing effectiveness, efficiency, ease of use, usability and user experience when using tablets and laptops in typical private and business tasks. To measure these characteristics we used the task completion rate, the task completion time, the Single Ease Question (SEQ), the Software Usability Scale (SUS) and AttrakDiff. Results indicate that there is a difference between effectiveness, efficiency and the users' assessment of the devices. Users can carry out tasks more effectively and efficiently on laptops, but rate tablets higher in perceived usability and user experience, indicating that a pleasant and meaningful experience depends on more characteristics than work-related qualities such as effectiveness and efficiency.

Keywords: usability, user experience, satisfaction, ease of use, AttrakDiff.

1 Motivation

In recent years tablets have reached a considerable market penetration.[1, 2] In comparison to more established devices like laptops they rely on touch based controls, are smaller and consequently lead to different user interactions. These different interaction paradigms have also led to different usage scenarios. Laptops have always been used in work contexts and people are therefore used to carrying out work related tasks with them (e.g. writing emails or scheduling meetings). Tablets are primarily designed for the consumer market and are mostly used at home (e.g. as media consumption device or as companion to television viewing and other living-room activities) [3] However, studies show that the adoption rate is also surging in the business context. Especially IT and business professionals use tablets extensively and do so for private as well as work related tasks [4]. This is especially true for mobile workers (e.g in customer relations, sales and management) [2]. Furthermore there is a tendency to use private devices to access work related applications and services [4]. Yet despite the growing adoption rate in business and consumer areas, there is still

A. Marcus (Ed.): DUXU 2014, Part I, LNCS 8517, pp. 402–412, 2014.

little comparative research with focus on usability aspects. Ozok et al. [5] conducted a study in 2008 comparing tablets and laptop on a number of tasks, but the usability of modern tablets has improved heavily in the last years. We therefore conducted a pilot study to examine the differences between these two device types taking into consideration their different application areas. We also wanted to examine different evaluation concepts and metrics. Since laptops are often used in work- and task-related contexts, we considered usability as defined by the ISO as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction…"[6] as an appropriate characteristic for the comparison. Nonetheless, especially in the consumer market, this work and task-related perspective has been broadened by focusing on the whole user experience which incorporates further needs and emotions such as enjoyment, fun, excitement or appeal when using a device or system[7]. Since tablets are a typical consumer product we also considered user experience as an interesting metric.

We therefore decided to conduct a study comparing the two devices on usability and user experience metrics when performing typical tasks in a private and business context. To gain experience carrying out the experiment we started with a small scale pilot study with eight participants. The results will be used as lessons learned for a following study with more participants.

The remainder of the paper is structured as follows: In chapter 2 we describe the methodology of the study. Chapter 3 contains the results of the user tests and the limitations of the study. In chapter 4 we draw conclusions and point out future work.

2 Methodology

The evaluation compares tablets and laptops based on a number of metrics. Since we wanted to compare these devices based on user performance metrics derived from the field of usability [U] as well as perceived qualities by users, we defined effectiveness [U1], efficiency [U2], ease of use [U3], usability [U4] and user experience [UX] as relevant evaluation criteria. Furthermore, we adopted a 2-step approach by using metrics that were measured after every task and metrics that were measured for the devices after completing all tasks.

- **Task oriented metrics:** These post-task metrics were measured for every task after the participant had completed the task.
  - **Effectiveness [U1]**: Effectiveness was measured using the task completion rate by observing how many attempts participants needed and rating the attempts using a 4-step scoring model. A task was considered *"easy"* (step 1) if it was completed on the first attempt and *"hard"* (step 2) if it was completed on the second attempt. The task was considered not successfully completed if the participant had to be assisted (step 3) or failed (step 4).
  - **Efficiency [U2]**: Efficiency was measured using the task completion time. We compared the user performances based on the mean, the maximum and minimum completion time.
  - **Ease of Use [U3]**: To measure task-performance satisfaction, we used the Single Ease Question (SEQ) *"Overall, how difficult or easy did you find this*

task?" The question was asked after the completion of every task and participants had to rate the difficulty on a 7-step Likert scale from *"very difficult"* to *"very easy"*. This question was chosen, because it performs as well as other, more complicated measures of task-difficulty like the Subjective Mental Effort Questionnaire (SMEQ) or the Usability Magnitude Estimation (UME) [8]. Owing to the low number of participants we used the mean instead of the median to compare the difficulty of the tasks.

- **Device oriented metrics:** These post-study metrics were measured for every device after the participant had completed all tasks on the specific device.
  - **Usability [U4]**: Overall perceived usability using a device was measured using the Software Usability Scale (SUS) questionnaire [9]. SUS also provides a reliable global measure of system satisfaction. The questionnaire consists of 10 items that are answered using a 5-step Likert scale reaching from *"strongly disagree"* to *"strongly agree"*. It was chosen, because it is a reliable and valid measure of perceived usability [10, 11] and is widely applied, which allows for a comparison with existing results and products. The questionnaire was answered once for every device after carrying out all tasks on the respective device, because it is designed as a post-study questionnaire.
  - **User Experience [UX]**: User experience was measured using *AttrakDiff* [12], which contains 28 bipolar items that are rated on a 7-step scale to determine *"pragmatic quality"*, *"hedonic quality (stimulation)"*, *"hedonic quality (identity)"* and *"attractiveness"* of the devices. AttrakDiff was chosen, because the reliability of its subscales measured using Cronbach's Alpha coefficient is higher (between .73 and .90) [12] than the reliability of the User Experience Questionnaire (UEQ) (between .65 and .89) [13]. The questionnaire was answered once for every device after answering the SUS questionnaire.

Since all these metrics are based on measurable user performances (effectiveness, efficiency) or the users' perceptions of the capabilities of the devices to achieve certain goals (ease of use, usability, user experience), we had to define a number of tasks to be carried out. To ensure an objective comparison we included the most frequently executed tasks from the private as well as the business context. We determined these tasks by comparing and analyzing existing studies. Since the results of all studies referenced to similar tasks we defined the following tasks for our pilot study:

- **Business context**
  - **Task 1: Writing an email**: Participants had to open an email client, create a new email, add a specific recipient, a subject and content, insert a certain picture and send the email.
  - **Task 2: Creating a calendar entry**: Participants had to open a calendar app, create a new appointment, choose a date, the time and the duration, define a location, save the appointment and reopen it to add a reminder, switch to month-view and move the appointment to another day.
- **Private context**
  - **Task 3: Browsing and filling out forms**: Participants had to open a web browser, enter a specific URL, fill out a form completely and send the form.

- **Task 4: Finding and gathering information**: Participants had to open a web browser, research the address of a certain hotel and find out the route to this location.

Participants had to carry out all four tasks using an iPad 2 running iOS 5.1.1 and an Acer Laptop (Aspire 5740) running Microsoft Windows 7. These devices were chosen because at the time of the study the iPad was the top-selling tablet with iOS being the leading tablet platform[1, 14] while Windows was the leading platform on desktops and laptops.[15] To ensure participants were experienced in using the devices they had to have used them in a private or business context at least occasionally in the previous three months and had to be familiar with the standard applications *"Mail"* (iOS), *"Calendar"* (iOS) and *"Outlook"* (Windows). As web browsers *"Google Chrome"* (laptop) and *"Safari"* (iPad) were used, but the particular functionality of the browsers was not crucial for task fulfillment.

Eight people (7 men and 1 woman between 22 and 37 years old) participated in the study. Four participants began carrying out the tasks by using the iPad, four used the laptop first.

3 Results

In the following sections we lay out the results of the user tests in respect to the 5 chosen metrics.

3.1 Effectiveness

Participants had problems completing task 1 on the tablet. One participant completed the task on the second attempt, 2 participants needed assistance, one participant failed completely. Nobody had problems using the laptop. All other task did not lead to considerable problems or differences (see Fig. 1).

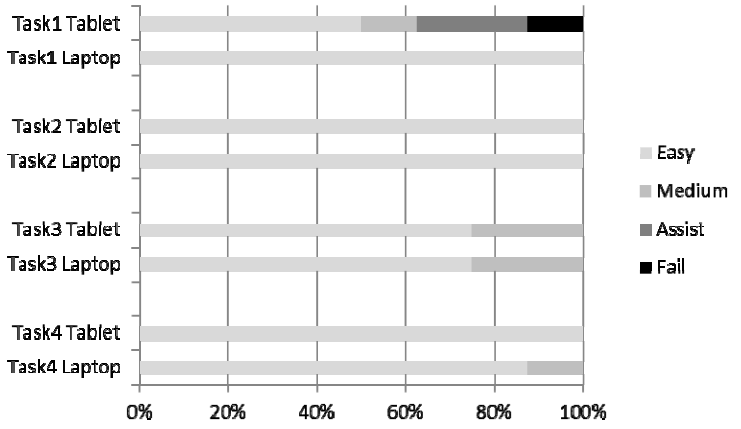

Fig. 1. Task Performance Rates results

3.2 Efficiency

Efficiency was measured by the task completion time. Fig. 2 illustrates the results using the mean time in seconds represented by the columns, as well as the minimum and maximum time needed using error bars. Participants who failed to complete a task were not considered in this analysis.

Comparing each task on the two devices results shows that all tasks were carried out faster on the laptop. In case of task 1 the difference in mean time was high (35.9%), because three participants had trouble completing the task. All other tasks were only slightly faster on the laptop (task 2 4.1%, task 3 16.2%, task 4 11.2%).

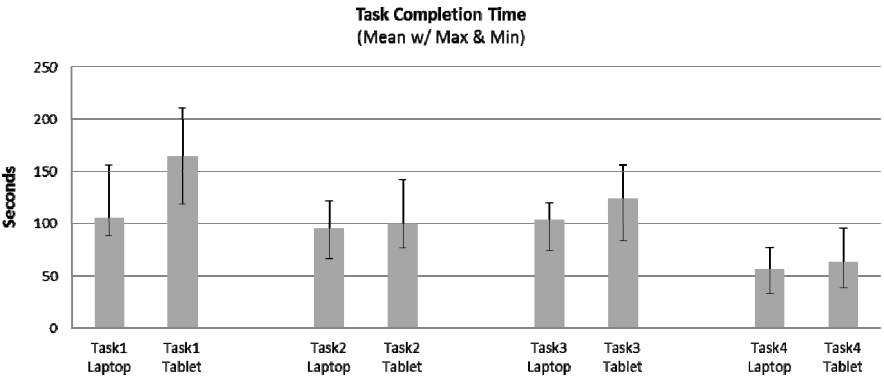

Fig. 2. Task Completion Time results

3.3 Ease of Use

Ease of Use was measured using the Singe Ease Question (SEQ). Answers were rated on a 7-step Likert scale from 1 (very easy) to 7 (very difficult). Table 1 summarizes the results using the mean value instead of the median because the number of participants was so low.

Results show a mixed picture. Writing emails is considered harder on the tablet, creating calendar entries is slightly easier on the tablet, browsing and filling out forms is slightly easier using the laptop, finding and gathering information is considered equally easy. These findings are in line with existing studies that find, that users tend to rate tasks more difficult if they take longer or do not succeed [16].

Table 1. Singe Ease Question results

	Mean	
	Laptop	**Tablet**
Task 1 (Writing Email)	1,13	3,25
Task 2 (creating a calendar entry)	2,00	1,63
Task 3 (browsing and filling out forms)	1,25	1,63
Task 4 (finding and gathering information)	1,25	1,25

3.4 Usability

Usability was measured using the Software Usability Scale (SUS). It consists of ten items that participants had to rate on a 5-step scale. The resulting SUS score was higher for the tablet (82.5) than the laptop (75) resulting in an overall higher perceived usability of the tablet (see Table 2 and Table 3).

Table 2. SUS-Score of the laptop

	System Usability Scale	**Strongly Disagree**				**Strongly Agree**	
	75% SUS Score:	**1**	**2**	**3**	**4**	**5**	**Mean**
I1	I think I would like to use this software product frequently.	0	0	0	2	6	4,750
I2	I found the product unnecessarily complex.	3	2	2	1	0	2,125
I3	I thought the product was easy to use.	0	1	1	3	3	4,000
I4	I think I would need Tech Support to be able to use this product.	6	1	0	0	1	1,625
I5	I found the various functions in this product were well integrated.	1	1	2	2	2	3,375
I6	I thought there was too much inconsistency in this product.	3	3	2	0	0	1,875
I7	I imagine that most people would learn to use this product very quickly.	0	1	4	2	1	3,375
I8	I found the product very cumbersome to use.	4	3	1	0	0	1,625
I9	I felt very confident using this product.	1	0	1	2	4	4,000
I10	I need to learn a lot about this product before I could effectively use it.	4	1	1	1	1	2,250

Table 3. SUS-Score of the tablet

	System Usability Scale	**Strongly Disagree**				**Strongly Agree**	
	82,5% SUS Score:	**1**	**2**	**3**	**4**	**5**	**Mean**
I1	I think I would like to use this software product frequently.	0	0	1	0	7	4,750
I2	I found the product unnecessarily complex.	2	6	0	0	0	1,750
I3	I thought the product was easy to use.	0	2	0	1	5	4,125
I4	I think I would need Tech Support to be able to use this product.	7	0	0	0	1	1,500
I5	I found the various functions in this product were well integrated.	0	2	1	4	1	3,500
I6	I thought there was too much inconsistency in this product.	6	2	0	0	0	1,250
I7	I imagine that most people would learn to use this product very quickly.	1	0	2	1	4	3,875
I8	I found the product very cumbersome to use.	5	2	1	0	0	1,500
I9	I felt very confident using this product.	1	0	1	2	4	4,000
I10	I need to learn a lot about this product before I could effectively use it.	6	2	0	0	0	1,250

Furthermore the tablet was also considered to be less complex (I2), easier to learn (I7) and more consistent (I6).

Results from I3 (I thought the product was easy to use) seem to contradict the results from the Singe Ease Question (SEQ) in section 3.3. Table 1 shows that using the mean of all answers to the SEQ the laptop is perceived as easier. When asked in I3 of the SUS about ease of use, participants rate the tablet as easier to use. This is true not only for the mean value, but also the median.

3.5 User Experience

To measure user experience, we used *AttrakDiff [12],* which evaluates devices based on the four dimensions *"pragmatic quality"* (PQ), *"hedonic quality - identity"* (HQ-I), *"hedonic quality - stimulation"* (HQ-S) and *"attractiveness"* (ATT).

- Pragmatic quality (PQ): A quality of a product that describes whether its functions are appropriate to achieve certain goals (e.g. practical)
- Hedonic quality - identity" (HQ-I): A quality of a product that describes whether its attributes can satisfy the human need to be perceived by others in a certain way (e.g. presentable or stylish)
- Hedonic quality - stimulation (HQ-S): A quality of a product that describes whether it can fulfill the human need to improve personal skills and knowledge (e.g. creative).
- Attractiveness" (ATT): A general positive or negative assessment of the appeal of a product.

Within these four dimensions the devices had to be rated based on 7 word pairs using a 7-step scale. Fig. 3 illustrates the results for every word pair using the mean rating of all participants.

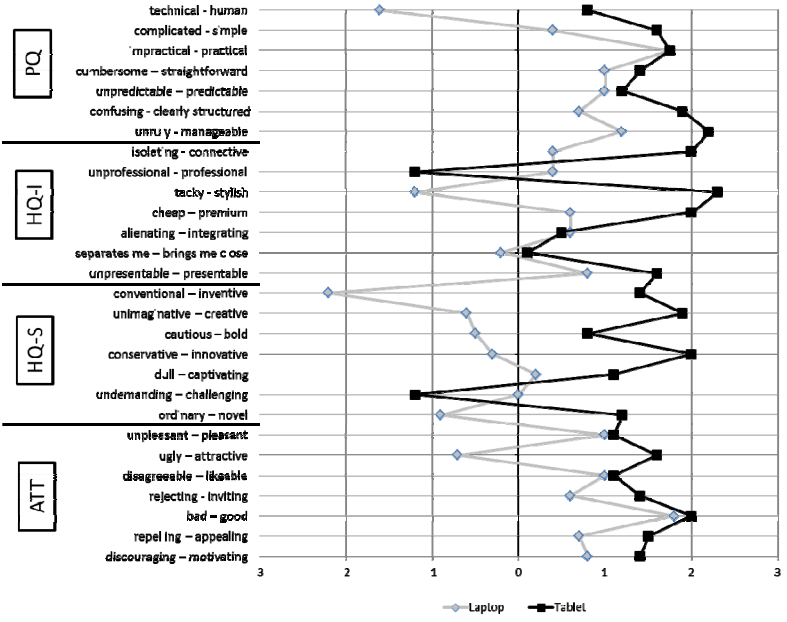

Fig. 3. Mean ratings of word pairs

The results show that the tablet scores higher in nearly all word pairs. Consequently, using the mean value of the including word pairs, the tablet scores higher in all four dimensions of AttrakDiff (see Fig. 4).

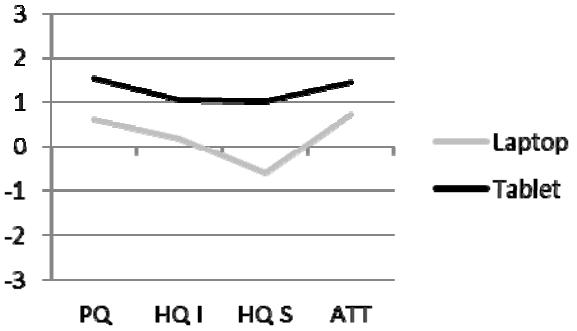

Fig. 4. Mean values of all four AttrakDiff dimensions

Furthermore, there is an influence between these dimensions. Pragmatic quality and hedonic quality are independent and both influence attractiveness. Therefore AttrakDiff compares the devices based on these two dimensions and positions them in a two-dimensional portfolio (see Fig. 5), which consists of nine areas. Products that have a high pragmatic as well as hedonic quality are desired. If the pragmatic quality is average or low products are considered to be too self-oriented. If the hedonic quality is average or low the product is too focused on task fulfillment.

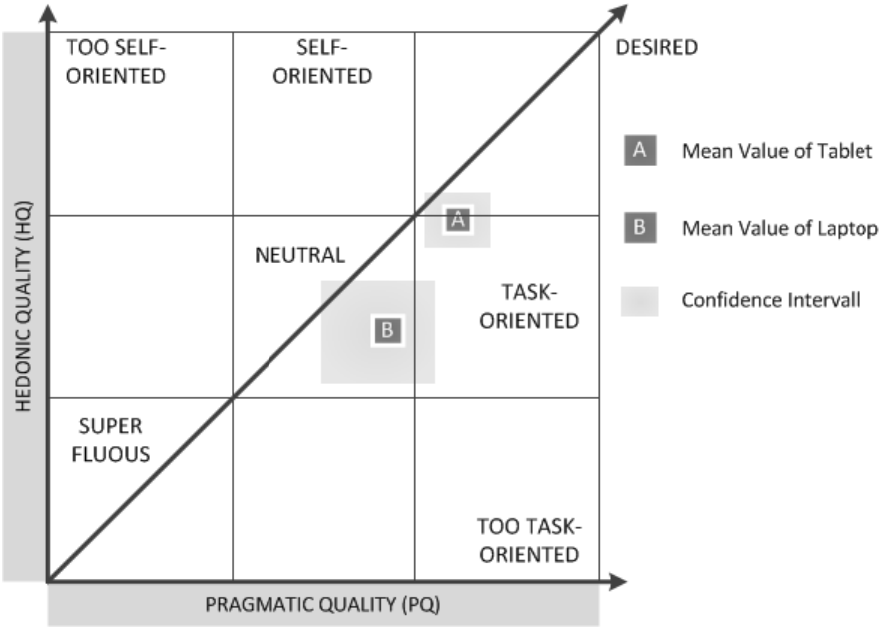

Fig. 5. Pragmatic and hedonic quality of tablets (A) and laptops (B)

Products are represented in this portfolio using the mean value and a confidence interval. The confidence interval is calculated using t-tests and a significance level of 0.05. A small confidence interval indicates that participants agree on the rating of the product. Consequently, the rating is also more likely to really apply to the product.

Results for the tablet show that it has a high pragmatic quality and is considered to be *"task-oriented"*. The hedonic quality of the tablet shows a mixed picture. It is also high, but the confidence interval is not completely in the high area ("desired"). The laptop is considered to be a more neutral product. The confidence interval reaches the top area in pragmatic quality, but the hedonic quality is average.

These differences between pragmatic and hedonic qualities of the two products are statistically significant. Furthermore, the confidence interval of the tablet is smaller for both qualities. Thus, participants agree more on the qualities of the tablet, whereby the assessment of the tablet applies with a higher probability to the actual product.

3.6 Limitations

Since this was only a pilot study to evaluate the chosen metrics and methods, the sample size serves as a limitation. Eight users are not a sufficiently high number to ensure valid results for such quantitative metrics.

The results for the metric *"effectiveness"* are very similar for all tasks. With the exception of task 1 on the tablet every task was carried out with very few problems. This is an indicator that the tasks may have been too easy.

Consequently, the results for the metric *"efficiency"* also just show a high variance for this one task. If the tasks had been more difficult, the assumption is that the variances would have been higher. Nevertheless these results are in accordance with other studies that show that typing performance is lower on virtual keyboards [17, 18]. Therefore it is likely that if the tasks had required more user inputs the results for effectiveness and efficiency would have been worse on the tablet. Consequently, these aspects should be considered when designing the follow up study in order to be able to reach a fair balance between the tasks and devices.

Finally, results do depend not only on the device but also on the platform and the software applications used. We tried to minimize this effect by utilizing the most used platforms and applications for both devices and only accepting participants that were regular users of both devices. To minimize this effect even further, different device types with different platforms and applications could be used. The familiarity with the device and platform could also be increased when users bring their own device to the experiment.

4 Conclusions

Results indicate that there is a difference between actually measureable characteristics such as effectiveness and efficiency and the users' assessment of the devices. Tests show that users can carry out tasks more effectively and efficiently on laptops. They also rate them as easier on laptops when asked directly after the task using the Single

Ease Question. However participants rated tablets higher in usability when asked after performing all tasks.

Furthermore, despite the problems in carrying out tasks the user experience in the sense of pragmatic quality, hedonic quality (stimulation and identity) and attractiveness (AttrakDiff) is rated higher for tablets, indicating that a pleasant and meaningful experience depends on more characteristics than work-related qualities like effectiveness and efficiency.

Another interesting aspect is the difference between post-task and the post-study metrics about difficulty. Answering the Singe Ease Question (Overall, how difficult or easy did you find this task?) after every task participants rated the laptop as easier to use (using the mean of all answers). However, in the post-study System Usability Score (SUS) questionnaire in item 3 (I thought the product was easy to use) participants rated the tablet as easier. Since this difference is small and there are a number of limitations to the study, this aspect should be examined in another study with a higher number of participants.

Since in this pilot study there seems to be a discrepancy between effectiveness, efficiency and ease of use after carrying out the tasks and the results of the post-study System Usability Scale, we want to analyze this aspect further. In a follow-up study we want to eliminate the limitations of this pilot study and investigate whether and how a higher attractiveness of a device influences peoples' ratings of task-related usability aspects in post-task and post-study evaluations.

References

1. Huberty, K., Holt, A., Moore, J., Devgan, S., Meunier, F., Gelblum, E., Devitt, S., Lu, J., Chen, G., Shih, S., Kim, S., Yoshikawa, K.: Tablet Landscape Evolution: Window(s) of Opportunity. Morgan Stanley Research (2012)
2. iPass: The iPass Global Mobile Workforce Report, http://www.ipass.com/wp-content/uploads/2013/03/ipass_mobile-workforce-report_q1_2013.pdf
3. Mueller, H., Gove, J.L., Webb, J.S.: Understanding Tablet Use: A Multi-Method Exploration. In: Proceedings of the 14th Conference on Human-Computer Interaction with Mobile Devices and Services (Mobile HCI 2012), pp. 1–10. ACM (2012)
4. International Data Group: iPad for Business. Survey (2012), http://www.idgconnectmarketers.com/wp-content/uploads/2012/01/iPad-Survey.pdf
5. Ozok, A.A., Benson, D., Chakraborty, J., Norcio, A.F.: A Comparative Study Between Tablet and Laptop PCs: User Satisfaction and Preferences. International Journal of Human-Computer Interaction 24, 329–352 (2008)
6. ISO: Ergonomic requirements for office work with visual display terminals - Part 11: Guidance on usability (1988)
7. Hassenzahl, M., Tractinsky, N.: User experience - a research agenda. Behaviour& Information Technology 25, 91–97 (2006)

8. Sauro, J., Dumas, J.S.: Comparison of three one-question, post-task usability questionnaires. In: Olsen Jr., D.R., Arthur, R.B., Hinckley, K., Morris, M.R., Hudson, S.E., Greenberg, S. (eds.) Proceedings of the 27th International Conference on Human Factors in Computing Systems, pp. 1599–1608. ACM, New York (2009)
9. Brooke, J.: SUS - A quick and dirty usability scale. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, A.L. (eds.) Usability Evaluation in Industry, pp. 189–194. Taylor and Francis, London (1996)
10. Bangor, A., Kortum, P.T., Miller, J.T.: An Empirical Evaluation of the System Usability Scale. International Journal of Human-Computer Interaction 24, 574–594 (2008)
11. Lewis, J.R., Sauro, J.: The Factor Structure of the System Usability Scale. In: Kurosu, M., et al. (eds.) Human Centered Design, HCII 2009. LNCS, vol. 5619, pp. 94–103. Springer, Heidelberg (2009)
12. Hassenzahl, M., Burmester, M., Koller, F.: AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In: Szwillus, G., Ziegler, J. (eds.) Mensch & Computer 2003, vol. 57, pp. 187–196. Vieweg+Teubner Verlag, Wiesbaden (2003)
13. Laugwitz, B., Held, T., Schrepp, M.: Construction and Evaluation of a User Experience Questionnaire. In: Holzinger, A. (ed.) USAB 2008. LNCS, vol. 5298, pp. 63–76. Springer, Heidelberg (2008)
14. Good Technology: Good Technology Device Activations Report Q4 (2012), `http://media.www1.good.com/documents/Good-Q4-Device-Activations.pdf`
15. Net Applications, `http://www.netmarketshare.com/`
16. Sauro, J., Lewis, J.: Correlations among Prototypical Usability Metrics: Evidence for the Construct of Usability. In: Olsen Jr., D.R., Arthur, R.B., Hinckley, K., Morris, M.R., Hudson, S.E., Greenberg, S. (eds.) Proceedings of the 27th International Conference on Human Factors in Computing Systems, pp. 1609–1618. ACM, New York (2009)
17. Kim, J.H., Aulck, L., Bartha, M.C., Harper, C.A., Johnson, P.W.: Are there differences in muscle activity, subjective discomfort, and typing performance between virtual and conventional keyboards? In: Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting, pp. 1104–1108. SAGE, Boston (2012)
18. Lee, S., Zhai, S.: The performance of touch screen soft buttons. In: Olsen, D.R. (ed.) Proceedings of the 27th International Conference on Human Factors in Computing Systems, CHI, pp. 309–318. ACM, New York (2009)