Use of WebLabs as a platform for remote customer research

Dirk van de Mortel
h.m.j.v.d.mortel@tue.nl
Philips Research Laboratories Eindhoven
March - May 2008
Last saved @ 19-05-2008 10:18:00

Contents

1. Introduction

WebLabs is the project name for a new, web-based platform for gathering user feedback. During most, if not all phases in the (iterative) design cycle, representative users and/or clients can be involved and leave feedback/input as part of User-Centered Design (UCD) process. This feedback is valuable for an evaluator (e.g. HCI expert), who then compiles yielded information and implements severe parts, by rendering them in the development of the product or service. Executing usability tests by means of a User Evaluation Method (UEM) as part of UCD is a well-known method. Other usability methods are e.g. contextual interviews, focus groups, heuristic evaluation, interviews, personas, task analysis etc that all might be applicable as feedback system on the web.

However, until now little information is known on how laboratory testing compares to remote testing, concerning Philips Research applications (i.e. multimedia applications for consumer electronics). Important advantages of remote testing are A. Time and space independence (a-synchrone) B. Relatively low cost for a worldwide (and cultural diffused) audience/participants and C. Automation of results. Important disadvantages are I) No control after publication or invitation of users II) Applies on the imagination of participants when a situation is simulated/mimicked III) Requires special preparation (low fidelity prototypes become digital or hybrid: send by post and evaluated with Web applications).

1.1. Goal

By using a usability evaluation method for investigating whether results between traditional ways of conducting user evaluation experiments in the laboratory, compared to a remote setting (over the Internet and on a designated website). Our research question:

> *How do laboratory and remote testing compare in finding usability issues of Edit While Watching?*

Sub-ordinate to WebLabs is the application for evaluation, is the video editing application named Edit While Watching[*] (EWW). Since this application is unknown, still under development and a typical Philips Research application, feedback on the user-interface could provide valuable end-user insights.

Initially, EWW aims as television application and by its intuitive interface provoke multi-user or social use. For testing purposes to a global audience, it has been adapted to a personal computer (desktop) application to answer fundamental questions on functionality. In spite of shift this change in focus, which introduced new interaction problems which evidently generated new and obvious usability issues, which should relatively easy to find by potential users/participants.

1.2. Tools

By using frequently used and applied questionnaires in the HCI field, errors and biased results that might lead to confounding factors, attempted to exclude this way. Scoring between the two groups is compared by averaging each question, which might be split over several questions. Our

---

[*] Based on algorithms, this application makes a summarization from raw DV material and is personified during playback by applying reduced editing functions.

hypothesis is that we find no significant difference between these two groups throughout the four questionnaires.

2. Methods and Materials

2.1. Criteria participants

Recruiting of participants focussed on these aspects:
- Age between 18-60 years: avoiding legal issues <18.
- Affinity with computers: users have to know their way around with computer applications (and interaction styles and methods).
- None to a little experience in video editing and video shooting: advanced users have in-depth demands and requirements in which EWW does not intent to provide. EWW focuses on quick and easy use.
- English language at an acceptable level of understanding: the experiment is set-up and user interface are in English language.
- The amount of participants required for the experiment to become valid, is eight in each group.

2.2. Recruitment

Flyers were posted throughout the Philips High Tech Campus (HTC), digitally at the Research newsgroup, TUE/USI email group and by verbal advertisement in private circles. People, who responded on the advert to participate in the research, received an invitation email with detailed information regarding, along with the option to choose between online or coming over for the laboratory experiment (See Appendix 6.1). They were informed about testing a new video edit application and were requested to return a short *proficiency enquiry* by email on reply. This returned information was collected in a spreadsheet and the person placed in either or chosen experiment group. At this point they are participant (See Appendix 6.8).

After both conditions contained enough participants and experiment materials were prepared, participants were invited to start the experiment. Instructions and video material were sent out and/or appointments made. Remote participants were free to choose their starting time but were asked to finish before a deadline.

EWW evaluated in two conditions, invited:
1. 12 participants in a controlled laboratory experiment at HTC
2. 10 participants were invited for the experiment remotely

2.3. Procedure

Both groups used the same Web 2.0 interface (WebLabs environment[†]), guiding users through the experiments' procedure. Subsequent steps are:
- Home page with visual link of EWW
- Promotional video of EWW

[†] http://www.simplicitylabs.net/intern/ log-in: weblabs password: SIMPLICITY

Next follow successive experiment phases, advanced with *proceed buttons* at the bottom of every page:
- Email address submit; for tracking and identifying (returning) user
- Background; overview experiment steps and information (See Appendix 6.4)
- Consent form with checkmark to continue (See Appendix 6.5)
- Download and tasks (See Appendix 6.6):
    o Remote users need to download and install the EWW application. In addition to laboratory experiment, where EWW application is installed
    o Tasks are shown one-by-one, after pressing a continuation link ("Click here to continue with the next task") as manner to tag off
- Questionnaires (See Appendix 6.7):
  Validated, formal and/or often practiced questionnaires have been selected and combined into one:
    o CSUQ (Computer System Usability Questionnaire),
    o QUIS (Questionnaire for User Interface Satisfaction),
    o SUS (System Usability Scale) and
    o USE (Usability Evaluation)

In the lab are two laptops situated, either of them have (EWW) installed; one *without* analyzed video data and one *with* analyzed data. During lab experiments, this saves analysing time.

Remote group used the same website and received a DVD with same video material (sent out beforehand), as in the lab. They must install EWW and wait for the analysis of the video.

2.3 Logging

All participants were monitored using WebLabs on their email address and session id thus every step in the experiment flow is tracked in the database. Besides browser interaction, function calls from within the application logged into a database.

As extra observation during laboratory experiments, screens of the laptops were duplicated over the network and observed by the experimenter. This made an interview at the end easier specific and quick to troubleshoot.

Example database entrée:

3. Results

3.1. Turnout

Laboratory experiment:
- One person did not show up, nor cancelled the appointment
- In another occasion the participant stopped because of a frequently crashing application and consuming more than the designated time

Remote experiment:
- Ten people agreed upon experiment location, or did not mind either and were placed in the online experiment group
- At the end of the deadline, 6 people had started the experiment and 2 finished
- After the deadline was extended by two weeks, 2 more finished and 1 afterwards
- In total merely 5 people finished the experiment by submitting the final questionnaire, 1 did started but did not finish
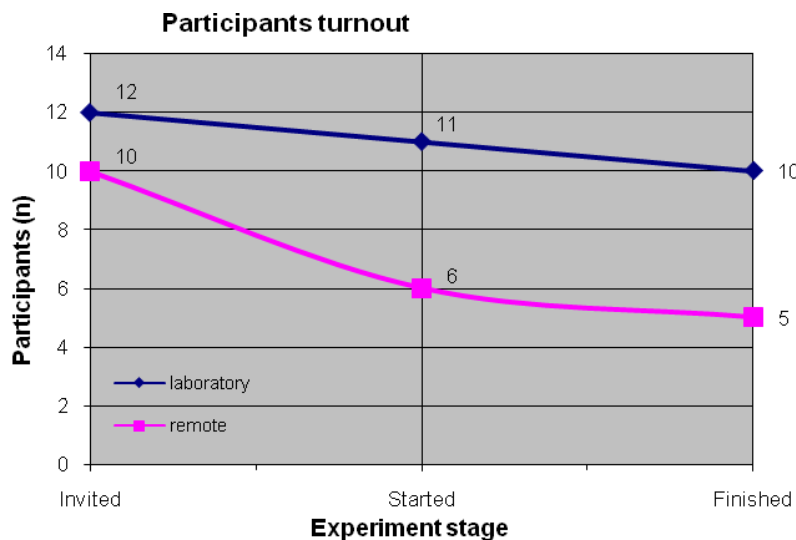
Table 1: Experiment attendance; turnout.

Finished	Laboratory	Remote
n (participants)	10	5
Average(age) (all)	32.6	30.8

Remarks in relation to turnout:
- EWW has oddities; stability is an issue and in certain circumstances

Questionnaires in general

Results of questions are clustered onto usability aspects where each question represents (as part of) a group. Scoring on each separately <u>does not</u> represent a stronger answer on a particular usability aspect; it visualizes cumulated questions on one specific aspect and show difference between both groups, to compare two test conditions. Table 1 shows the result for CSUQ regarding system usefulness involving questions 1 through 8 in both groups.

Table 1: Example calculation and comparison of an aspect of CSUQ.

CSUQ System Usefulness								
Question	1	2	3	4	5	6	7	8
Laboratory								
n (turnout)	10	10	10	10	10	10	10	10
Sum	33	33	30	31	35	39	31	32
CSUQ $_{SU\ lab}$ (average)								**3.3**
Remote								
n (turnout)	5	5	5	5	5	5	5	5
Sum	21	17	19	18	16	20	17	16
CSUQ $_{SU\ remote}$ (average)								**3.6**

SUS

The following calculation is made for every respondent and ranges from 0 through 100. Every question is rewarded with points:

- Uneven questions; - 1
- Even questions; - 5
    - Hence that scores have absolute numbers
- The sum is multiplied with 2.5

Table 2: SUS calculation of 1 participant.

Question	1	2	3	4	5	6	7	8	9	10
Score	1	1	5	1	4	1	5	1	3	1
Calculation	0	4	4	4	3	4	4	4	2	4

Sum calculation	33
	*2.5
Total	82.5

To compare both groups, each group is averaged.

CSUQ (Computer System Usability Questionnaire)

QUIS
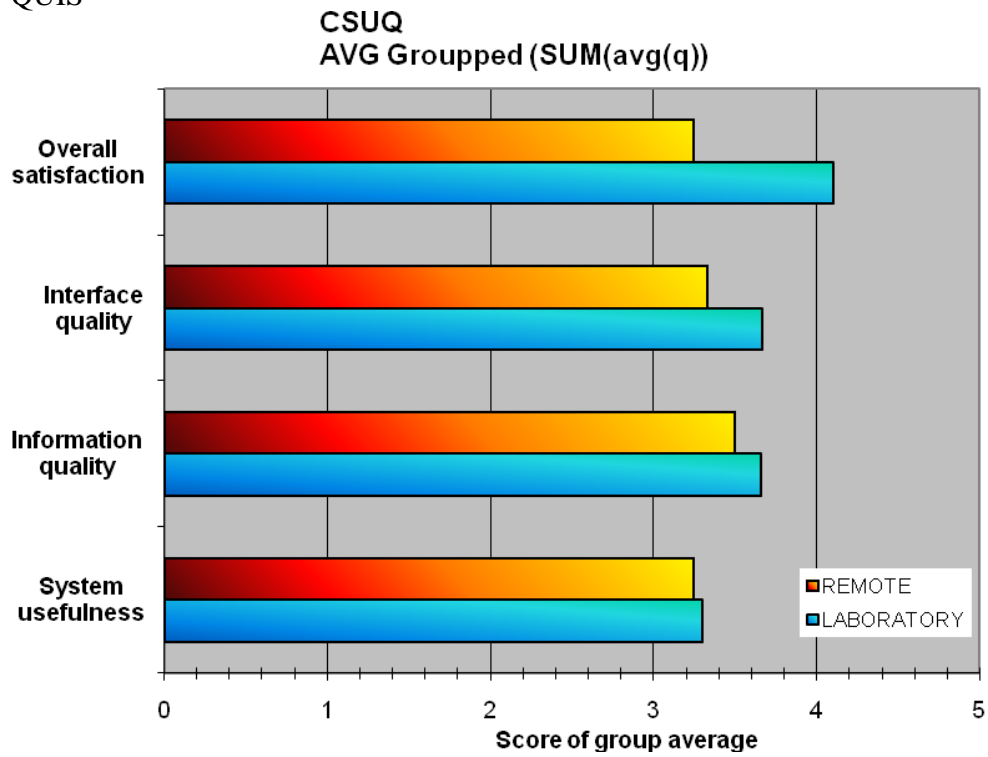
SUS – System Usability Scale

SUS

USE

4. Conclusion

Preliminary conclusions of a prior WebLabs experiment (April 2008) showed a divided success: at one hand a maximum expected users finished the experiment. Studies of experiments over the Internet show a turnout between 10 and 50%. However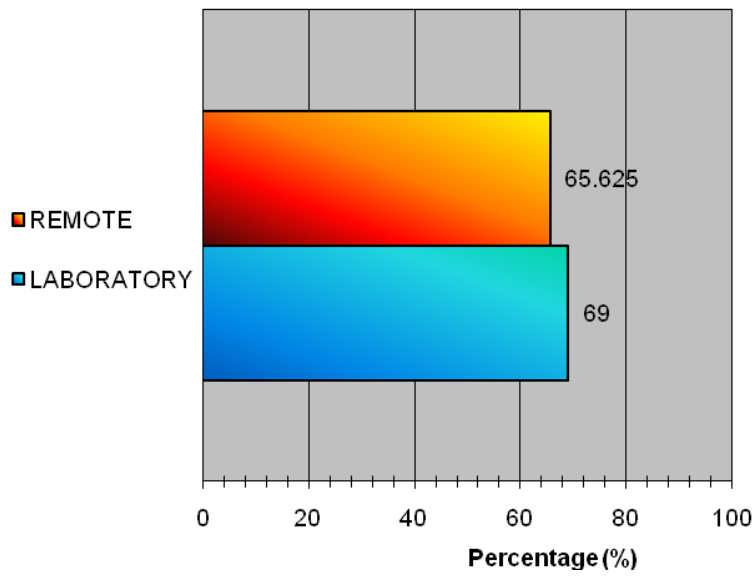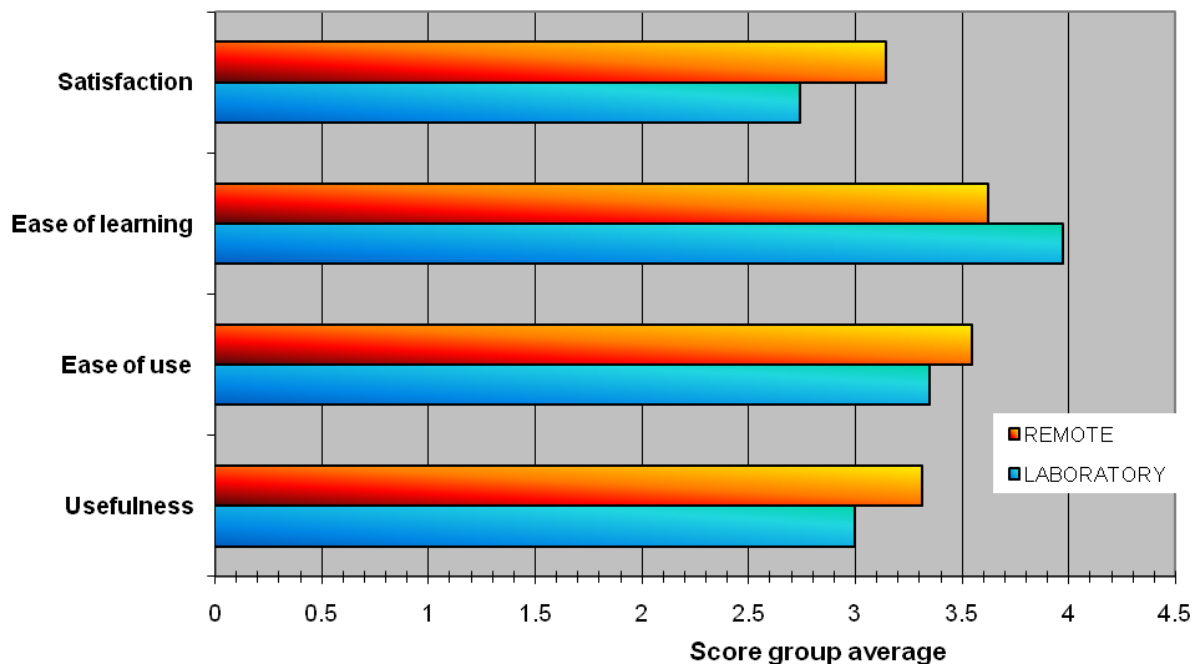, the previous WebLabs experiment based validity on the expectation that most participants would finish before a deadline, this was not the case. It had a turnout of 40% with n=4.

EWW did not prove to be stable at consumers; however after the experiment deadline the turnout was 20%. After encouragement this number doubled but did not stimulate all invited participants to become active.

Confounding factors

Aside previous experiment turnout, the sub-ordinate application or *stimulus* might be reconsidered. A *middleware* is suggested to blank out confounding factors; an interface on WebLabs that interacts with the application on the server. Or another suggestion might be a video instruction regarding a topic; were people of different intelligence performing tasks is shown. Afterwards participants rate the application.

5. Discussion

6. Appendices

6.1 Interest participation

6.2 Invitation remote experiment

6.3 Instructions remote experiment

Steps in workflow at WebLabs website:

6.4 Experiment background

6.5 Consent form

6.6 Download and tasks

6.7 Questionnaires

- *CSUQ - Computer System Usability Questionnaire*
- *QUIS - Questionnaire for User Interface Satisfaction*
- *SUS- system usability scale*
- *USE – usability evaluation*

6.8 Participants

6.1. Interest participation

Dear <name>,

First of all, thank you for your response!
We are still looking for participants for the experiment.

As said, we want to test new video editing software.
You can use your own video material (format is raw avi); take about 15-30 minutes or –in case you don't have- we provide you video footage.
The experiment takes place at the High Tech Campus, or at home (on your own). Both will take approximately 1 hour.
But first of all, we need your background before we place you into a group.

Can you please, fill out the questions and send it back to me:

- Age:
- Highest finished education:
- Occupation:
- English level:
- Hours spent daily (on average) with computers:
- How often did you use video editing software last year:
- Are you willing to visit us at the campus?
- Do you have own video material (raw avi)?
- (If you don't have video material) Postal address:

If we invite you for the experiment to come to the campus, which days and possible hours suit you best? (Please provide several time slots)

Thanks for your interest in participation!

Best regards,

6.2. Invitation remote experiment

Dear <name>,

We selected you to participate in the online experiment. A mailing is sent out which contains instructions and a DVD. It is possible to use your own DV AVI video however it's recommended to use the supplied video.

You can start the experiment at anytime in your home environment. Try to finish it in one session despite the possibility taking it up again.

Success and best regards,

6.3. Instructions remote experiment

Dear participant,

You have been selected to carry out the video-editing experiment on the Internet. This means you will execute the experiment **at home** (preferably <u>not in your office</u> at work but in your home situation) at any time you want. Therefore you receive the DVD, included with this post, with the contents of an "Indian Wedding" (DV AVI) video material. It is of course possible to use your own DV AVI material instead. Please finish the experiment in one go, once you started the tasks, if possible.

Requirements

- An active internet connection during the experiment for download of the application, instructions and input and during running the Edit-While-Watching application

- 5 Gigabyte of free space on your hard disk

Instructions

- Copy the contents of the DVD (Indian_Wedding.avi) to your computer's local disk

- Surf to http://www.simplicitylabs.net/intern/

- Login with name *weblabs* and password *SIMPLICITY (caps sensitive)*

- Click on Edit-While-Watching image and continue by following the instructions on the website

It is desirable to finish the experiment before March 31, 2008 as of then the results will be compiled and analyzed. In case of difficulties, or when you deviate from the instructions, please report them to me: dirk.van.de.mortel@philips.com.

Thank you for your cooperation and good luck,

Kind regards,

P.S.:
When you close the application, it asks you to give feedback. This is not related to the experiment but of course you can fill it out. The questionnaire used for the experiment can be found at:
http://www.simplicitylabs.net/intern/

6.4. Experiment background

Welcome to Edit-While-Watching experiment

This website provides you the opportunity to get involved into this online experiment to improve the quality of our application. The experiment takes roughly 60 minutes to complete.
When you are prompted to enter an email address, enter *your_full_name@exp.nl* e.g. bartjanssen@exp.nl

The website will guide you through the steps needed, from begin to end, to complete the experiment. Respectively, these steps are:

- Consent form: Common with experiments, that you agree upon participating.

- Download & tasks:
    - Download: If you do the test at home, you first need to download the application (and maybe you need to download extra Windows components, such as .NET Framework 3.0 and Windows Media Player 10+).

      Test if the application is working (you need to be connected to the internet) and you might want to play around a bit to get acquainted. Once you are ready, you can start working on the provided tasks.

    - Tasks: There are six tasks in total, read each task carefully before executing. A task list will be displayed, one by one, upon completion of the previous one. You need to switch between the website/tasks and the application to execute them. Click the link to continue, after you successfully finished a task.

      For testers at home, use video material from the DVD (which you might want to copy first to your local hard drive).

      For testers at Philips, use the video material in C:\My Video.

- Questionnaire: Continue with the questions regarding the application, once you completed all tasks. You need approximately 10 minutes to fill out the questionnaire.

  Remote participants only: It is advised that you finish the experiment in one go, preferably without interruptions. However, you can always stop the experiment at any time and return to finish it later by providing your email address as identification.

Remember, we are testing our software, not you. You can stop the experiment at any time.

6.5. Consent form

Consent form

Your personal information and your input will be kept strictly confidential and used for research purposes only. Personal information will not be used in reports and/or publications nor spread without your explicit permission.

By checking the checkbox at the bottom of the page, continuing with the experiment, you agree upon the following:

- o I am freely volunteering to participate in this experiment.

- o I have been told that I will be asked to complete tasks before and a questionnaire after, using of the application. I understand that the questionnaire will ask about my own opinion about using the application.

- o I am aware that I have the right at any time to withdraw consent and discontinue participation at any time without being penalized in any way.

- o While running the application, usage information is gathered of the application (Edit-While-Watching) and sent to a dedicated server. The application monitors mouse clicks only from within the application itself. None of your photo or video material itself is monitored, nor is any activity outside of the application.

- o As participant you fulfill tasks and questions sincerely and do your best.

- o Information of the experiment with the application presented will only be used by you without exploiting it; commercially and/or distribute it or its information in any way then it is intended for.

Thanks for taking time to do this experiment.

☐ Yes, I have read, understand and accept these conditions

6.6. Download and tasks

Download and tasks

Download Edit-While-Watching

This is the application you are about the test. Click the download button below and choose RUN to download and install. The Edit-While-Watching application supports only one video format (DV AVI) at this moment, which is video directly imported from your video camera. More information will be given in the instructions provided in the application itself.

[Download Edit-While-Watching button]

You might be prompted to update Microsoft Windows packages prior to the installation. If you have already installed a previous version of Edit-While-Watching, please uninstall this previous version first. If you have difficulties downloading or running the application please provide us feedback.

[Click here to continue with the tasks button]

Tasks

1.) Import a new video

2.) Edit the summary

> 2.1.) Remove scenes you don't like

> 2.2.) Restart playing in Fine Mode

> 2.3.) Add and remove scenes you like to make your summary complete

> 2.4.) Save the updated summary

3.) Create a new short summary video

> 3.1.) Remove everything in the summary

> 3.2.) Fill the summary with your favorite parts from the clip

> 3.3.) Save the summary

4.) Create a new summary

5.) Delete a summary

6.) Export the summary you like best Click here to continue with the next task

6.7. Questionnaires

CSUQ - Computer System Usability Questionnaire

Strongly disagree – strongly agree: 7 point scale, with a NA option in the end.

System usefulness:
1. Overall, I am satisfied with how easy it is to use this system.
2. It was simple to use this system.
3. I can effectively complete my work using this system.
4. I am able to complete my work quickly using this system.
5. I am able to efficiently complete my work using this system.
6. I feel comfortable using this system.
7. It was easy to learn to use this system.
8. I believe I became productive quickly using this system.

Information quality:
9. The system gives error messages that clearly tell me how to fix problems.
10. Whenever I make a mistake using the system, I recover easily and quickly.
11. The information (such as online help, on-page messages, and other documentation) provided with this system is clear.
12. It is easy to find the information I need.
13. The information provided by the system is easy to understand.
14. The information is effective in helping me complete the tasks and scenarios.
15. The organization of the information on the system pages is clear.

Interface quality:
16. The interface of this system is pleasant.
17. I like using the interface of this system.
18. The system has all the functions and capabilities I expect it to have.

Overall satisfaction:
19. Overall, I am satisfied with this system.

QUIS - Questionnaire for User Interface Satisfaction

Strongly disagree – strongly agree: 9 point scale, with a NA option in the end.

Overall Reaction to the application
1. terrible – wonderful
2. difficult – easy
3. frustrating – satisfying
4. dull – stimulating
5. rigid – flexible

Application
6. Reading characters on the page   hard – easy
7. Organization of information   confusing – very clear
8. Sequence of pages   confusing – very clear

Terminology and Application information
9. User of term throughout application   inconsistent – consistent
10. Terminology is intuitive   never – always
11. Position of messages on screen   inconsistent – consistent
12. Prompts for input   confusing – clear
13. Application informs about its progress   never – always
14. Error messages   unhelpful – helpful

Learning
15. Learning to use the application   difficult – easy
16. Exploring new features by trial and error   difficult – easy
17. Performing tasks is straightforward   never – always
18. Help messages on the screen   unhelpful – helpful
19. Supplemental reference materials   confusing – clear

Application Capabilities
20. Application speed   too slow – fast enough
21. Application reliability   unreliable – reliable
22. Sound associated with this application   detracts value – adds value
23. Correcting your mistakes   difficult – easy
24. Designed for all levels of users   never – always

SUS- system usability scale

Strongly disagree (1) – strongly agree (5); 5 point scale

1. I think I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think I would need Tech Support to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I need to learn a lot about this system before I could effectively use it.

USE – usability evaluation

Strongly disagree (1) – strongly agree (5); 5 point scale

Usefulness
1. It helps me be more effective.
2. It helps me be more productive.
3. It is useful.
4. It gives me more control over the activities in my life.
5. It makes the things I want to accomplish easier to get done.
6. It saves me time when I use it.
7. *It meets my needs.*
8. It does everything I would expect it to do.

Ease of Use
9. It is easy to use.
10. It is simple to use.
11. It is user friendly.
12. It requires the fewest steps possible to accomplish what I want to do with it.
13. *It is flexible.*
14. *Using it is effortless.*
15. *I can use it without written instructions.*
16. *I don't notice any inconsistencies as I use it.*
17. *Both occasional and regular users would like it.*
18. *I can recover from mistakes quickly and easily.*
19. *I can use it successfully every time.*

Ease of Learning
20. I learned to use it quickly.
21. I easily remember how to use it.
22. It is easy to learn to use it.
23. *I quickly became skilful with it.*

Satisfaction
24. I am satisfied with it.
25. I would recommend it to a friend.
26. It is fun to use.
27. It works the way I want it to work.
28. It is wonderful.
29. I feel I need to have it.
30. It is pleasant to use.

6.8 Participants WebLabs experiment

	name	age	education	occupation	English	Daily com-puter usage	video editing (last year)	Interested for Lab visit	Own video
L	john	33	Ir	scientist	fluent	8	12	yes	no
L	jane	26	MSc	student	good	7	some	maybe	no
R	john	26	MSc	student	good	8	1	no	yes
R	jane	25	Ir	student	good	7	none	rather not	no
L	jane			Student	good	8			
R	john	28	bachelor	Student	good	8	5to10	yes	no
L	john	36	bachelor	designer	avg	8	some	yes	no
L	john	38	hbo	scientist	good	4	none	yes	no
R	john	45	Ir	fysicus	good	5	once	yes	no
L	john	29	MA	student	fluent	9	once	yes	could
R	jane	28	Ir	student	good	7	none	rather not	no
L	jane	22	MSc	physics engineer	fluent	4	3or4	yes	could
L	john	36	Msc	software engineer		8	0	yes	yes
L	john	57	HBO	electronic engineer	good	4	regularly	yes	yes
R	john	55	mavo	illuminator/webdesigner/internet adviseur/cultureel coördinator	good	3	a lot	yes	yes
R	john	44	hts	ICT	good	6	6	possible	no
L	john	25	MSc	ICT	good	8	once	yes	yes
R	john	34	Msc	designer	good	9	none	no	no
R	john	29	Ir	software design	good	8	0	yes	no
R	john	35	Ir	Elect. HW designer	good	8	4	yes	could
L	john	24	Student	(bachelor) student	avg	10	10	no	yes
L	jane	28	mbo	edelsmit	good	3	0	yes	no

L lab experiment   R remote experiment

Data analysis

WebLabs: Usability factors of a software application, comparing remote testing and laboratory testing.

Dirk van de Mortel
Philips Research Laboratories Eindhoven
June 2008 Last saved @ 25-06-2008 10:51:00

Introduction

The EWW comparison experiment yielded four questionnaires and logging data. Interviews and observation were not noted according a scheme and served merely as discussion for the interview afterwards.

Goal

By using a usability evaluation method for investigating whether results between traditional ways of conducting user evaluation experiments in the laboratory, compared to a remote setting (over the Internet and on a designated website). Our research question:

> *How do laboratory and remote testing compare in finding usability issues of Edit While Watching?*

Questionnaires

Answers are checked for significance between two experiment conditions. Wilcoxon-Mann-Whitney U (two-tailed) test (in SPSS v15.0) suits best for this data (ordinal, comparison of two independent between groups with different group size) and is used to calculate A. Questionnaire groups and B. Individual questions. (Original questionnaires are added at the end of the document)

Table: Legend for statistic tables.

<= .200	<=.05 significant	>= 0.950 Strong insignificant

Results

Questions which were rated with NA (not applicable) were given the number "0". These questions have been removed in the calculation. Most remarkable results:

CSUQ
Significance ($p < 0.050$): <mark>none</mark>

Significance ($p < 0.200$):
9 (.091): The system gives error messages that clearly tell me how to fix problems
Here remote rate higher then laboratory.

10 (.054): Whenever I make a mistake using the system, I recover easily and quickly

Insignificance ($p > 0.950$):
6 (1.000): I feel comfortable using this system
12 (.950): It is easy to find the information I need
17 (.950): I like using the interface of this system
19 / Sum D "Overall satisfaction" (.951): Overall, I am satisfied with this system

QUIS
Significance ($p < 0.050$):
<mark>20 (.041): Application speed   too slow – fast enough</mark>

Significance ($p < 0.200$):
3 (.154): Overall Reaction to the application; frustrating – satisfying
10 (.172): Terminology is intuitive   never – always
15 (.132): Learning to use the application   difficult – easy

Groups: Sum of I (.196) Application capabilities

Insignificance ($p > 0.950$):
7 (.950): Help messages on the screen   unhelpful – helpful

USE
Significance ($p < 0.050$):
<mark>8 (.036): It does everything I would expect it to do</mark>

Significance ($p < 0.200$):
3 (.193): It is useful
12 (.152): It requires the fewest steps possible to accomplish what I want to do with it
13 (.086): It is flexible
22 (.106): It is easy to learn to use it
27 (.188): It works the way I want it to work

Insignificance ($p > 0.950$):
29 (.949): I feel I need to have it

Conclusion / discussion

There is no significant data found in any of the four questionnaires in their respective usability groups where the null hypothesis is rejected. Overall usability scales are insignificant; difference between laboratory and remote experimenting **cannot** be found for the four questionnaires.

Overall laboratory users tend to rate slightly better, as well on significant individual usability items with one exception CSUQ question 9.

In both conditions participants expressed their feelings on the huge amount of questions. Remote experiment shows "questionnaire fatigue" by having more of the same repetitive ratings.

Option for choosing NA in any off the questionnaires is wrong; it is not common to place NA.

* Interpretation towards EWW
Apply individual usability items ?

* Qualitative data (metrics, data logging)

CSUQ – Computer System Usability Questionnaire

A, Q1-8: System usefulness
B, Q9-15: Information quality
C, Q16-18: Interface quality
D, Q19: Overall satisfaction

Test Statistics(b)

	csuq1	csuq2	csuq3	csuq4	csuq5	csuq6	csuq7	csuq8
Mann-Whitney U	16.500	23.500	18.500	19.000	21.000	25.000	22.500	22.500
Wilcoxon W	71.500	78.500	63.500	64.000	36.000	40.000	77.500	77.500
Z	-1.054	-.188	-.540	-.480	-.209	.000	-.313	-.312
Asymp. Sig. (2-tailed)	.292	.851	.589	.631	.834	1.000	.754	.755
Exact Sig. [2*(1-tailed Sig.)]	.310(a)	.859(a)	.606(a)	.699(a)	.898(a)	1.000(a)	.768(a)	.768(a)

a Not corrected for ties.
b Grouping Variable: testtype

	csuq9	csuq10	csuq11	csuq12	csuq13	csuq14	csuq15
Mann-Whitney U	5.500	5.000	17.500	24.500	22.500	21.500	17.500
Wilcoxon W	33.500	15.000	32.500	79.500	77.500	76.500	72.500
Z	-1.689	-1.930	-.677	-.062	-.320	-.439	-.958
Asymp. Sig. (2-tailed)	.091	.054	.498	.950	.749	.661	.338
Exact Sig. [2*(1-tailed Sig.)]	.109(a)	.073(a)	.518(a)	.953(a)	.768(a)	.679(a)	.371(a)

	csuq16	csuq17	csuq18
Mann-Whitney U	21.500	24.500	12.000
Wilcoxon W	76.500	79.500	22.000
Z	-.448	-.063	-1.183
Asymp. Sig. (2-tailed)	.654	.950	.237
Exact Sig. [2*(1-tailed Sig.)]	.679(a)	.953(a)	.304(a)

	csuq19
Mann-Whitney U	24.500
Wilcoxon W	39.500
Z	-.062
Asymp. Sig. (2-tailed)	.951
Exact Sig. [2*(1-tailed Sig.)]	.953(a)

	csuqSumA	csuqSumB	csuqSumC	csuqSumD
Mann-Whitney U	19.500	24.000	22.000	24.500
Wilcoxon W	74.500	79.000	37.000	39.500
Z	-.675	-.123	-.370	-.062
Asymp. Sig. (2-tailed)	.500	.902	.712	.951
Exact Sig. [2*(1-tailed Sig.)]	.513(a)	.953(a)	.768(a)	.953(a)

QUIS – Questionnaire for User Interface Satisfaction

E, Q1-5: Overall reaction to the application
F, Q6-8: Application
G, Q9-14: Terminology and application information
H, Q15-19: Learning
I, Q20-24: Application capabilities

Test Statistics(b)

	quis1	quis2	quis3	quis4	quis5
Mann-Whitney U	17.000	21.000	13.500	23.000	24.000
Wilcoxon W	32.000	36.000	28.500	78.000	79.000
Z	-1.003	-.502	-1.426	-.249	-.126
Asymp. Sig. (2-tailed)	.316	.616	.154	.803	.900
Exact Sig. [2*(1-tailed Sig.)]	.371(a)	.679(a)	.165(a)	.859(a)	.953(a)

a Not corrected for ties.
b Grouping Variable: testtype

	quis6	quis7	quis8
Mann-Whitney U	16.500	24.500	18.000
Wilcoxon W	26.500	39.500	73.000
Z	-.240	-.062	-.289
Asymp. Sig. (2-tailed)	.810	.950	.772
Exact Sig. [2*(1-tailed Sig.)]	.825(a)	.953(a)	.839(a)

	quis9	quis10	quis11	quis12	quis13	quis14
Mann-Whitney U	13.000	14.000	22.000	11.000	17.000	10.000
Wilcoxon W	19.000	29.000	37.000	56.000	72.000	25.000
Z	-.094	-1.364	-.379	-1.143	-1.001	-.542
Asymp. Sig. (2-tailed)	.925	.172	.705	.253	.317	.588
Exact Sig. [2*(1-tailed Sig.)]	1.000(a)	.206(a)	.768(a)	.330(a)	.371(a)	.690(a)

	quis15	quis16	quis17	quis18	quis19
Mann-Whitney U	13.000	16.000	21.000	20.000	7.500
Wilcoxon W	28.000	31.000	36.000	35.000	22.500
Z	-1.506	-.889	-.496	-.342	-1.064
Asymp. Sig. (2-tailed)	.132	.374	.620	.733	.287
Exact Sig. [2*(1-tailed Sig.)]	.165(a)	.438(a)	.679(a)	.797(a)	.310(a)

	quis20	quis21	quis22	quis23	quis24
Mann-Whitney U	8.500	15.500	9.000	16.000	22.500
Wilcoxon W	23.500	30.500	54.000	71.000	37.500
Z	-2.047	-1.175	-.869	-.578	-.310
Asymp. Sig. (2-tailed)	.041	.240	.385	.563	.756
Exact Sig. [2*(1-tailed Sig.)]	.040(a)	.254(a)	.482(a)	.635(a)	.768(a)

Test Statistics(b)

	quisSumE	quisSumF	quisSumG	quisSumH	quisSumI
Mann-Whitney U	15.000	22.500	21.500	23.500	14.500
Wilcoxon W	30.000	37.500	36.500	78.500	29.500
Z	-1.229	-.308	-.430	-.184	-1.293
Asymp. Sig. (2-tailed)	.219	.758	.667	.854	.196
Exact Sig. [2*(1-tailed Sig.)]	.254(a)	.768(a)	.679(a)	.859(a)	.206(a)

a Not corrected for ties.
b Grouping Variable: testtype

USE – Usability evaluation

J, Q1-8: Usefulness
K, Q9-19: Ease of use
L, Q20-23: Ease of learning
M, Q24-30: Satisfaction

Test Statistics(b)

	use1	use2	use3	use4	use5	use6	use7	use8
Mann-Whitney U	24.000	23.000	15.000	23.000	22.000	24.000	24.500	8.500
Wilcoxon W	39.000	38.000	30.000	78.000	37.000	39.000	39.500	63.500
Z	-.134	-.267	-1.303	-.256	-.416	-.126	-.064	-2.099
Asymp. Sig. (2-tailed)	.893	.789	.193	.798	.677	.899	.949	.036
Exact Sig. [2*(1-tailed Sig.)]	.953(a)	.859(a)	.254(a)	.859(a)	.768(a)	.953(a)	.953(a)	.040(a)

	use9	use10	use11	use12	use13	use14	use15	use16	use17	use18	use19
Mann-Whitney U	20.500	23.000	21.500	14.000	11.500	17.000	16.000	23.500	23.000	23.500	18.000
Wilcoxon W	35.500	38.000	36.500	69.000	66.500	32.000	31.000	38.500	78.000	78.500	33.000
Z	-.589	-.259	-.443	-1.433	-1.719	-1.024	-1.136	-.188	-.256	-.190	-.881
Asymp. Sig. (2-tailed)	.556	.796	.658	.152	.086	.306	.256	.851	.798	.849	.378
Exact Sig. [2*(1-tailed Sig.)]	.594(a)	.859(a)	.679(a)	.206(a)	.099(a)	.371(a)	.310(a)	.859(a)	.859(a)	.859(a)	.440(a)

	use20	use21	use22	use23
Mann-Whitney U	15.500	19.000	12.500	20.000
Wilcoxon W	30.500	34.000	27.500	35.000
Z	-1.210	-.777	-1.619	-.638
Asymp. Sig. (2-tailed)	.226	.437	.106	.524
Exact Sig. [2*(1-tailed Sig.)]	.254(a)	.513(a)	.129(a)	.594(a)

	use24	use25	use26	use27	use28	use29	use30
Mann-Whitney U	20.000	17.000	23.000	13.000	21.500	24.500	18.000
Wilcoxon W	75.000	72.000	38.000	68.000	76.500	79.500	33.000
Z	-.627	-1.009	-.255	-1.562	-.449	-.064	-.909
Asymp. Sig. (2-tailed)	.531	.313	.799	.118	.653	.949	.363
Exact Sig. [2*(1-tailed Sig.)]	.594(a)	.371(a)	.859(a)	.165(a)	.679(a)	.953(a)	.440(a)

	useSumJ	useSumK	useSumL	useSumM
Mann-Whitney U	22.500	22.000	17.000	22.500
Wilcoxon W	37.500	37.000	32.000	77.500
Z	-.310	-.369	-.990	-.307
Asymp. Sig. (2-tailed)	.757	.712	.322	.759
Exact Sig. [2*(1-tailed Sig.)]	.768(a)	.768(a)	.371(a)	.768(a)

SUS – System Usability Scale

SU scores range from 0 to 100.

"SUS yields a single number representing a composite measure of the overall usability of the system being studied. Note that scores for individual items are not meaningful on their own." (J. Brooke, 1996, SUS-A quick and dirty usability scale)

Average experiment;
- Remote: 60.5
- Laboratory: 69

	SUS
Mann-Whitney U	19.500
Wilcoxon W	34.500
Z	-.676
Asymp. Sig. (2-tailed)	.499
Exact Sig. [2*(1-tailed Sig.)]	.513(a)

CSUQ - Computer System Usability Questionnaire

Strongly disagree (1) – strongly agree (7): 7 point scale, with a NA option (0) in the end.

.500 System usefulness:
.292   1. Overall, I am satisfied with how easy it is to use this system.
.851   2. It was simple to use this system.
.589   3. I can effectively complete my work using this system.
.631   4. I am able to complete my work quickly using this system.
.834   5. I am able to efficiently complete my work using this system.
1.000 6. I feel comfortable using this system.
.754   7. It was easy to learn to use this system.
.755   8. I believe I became productive quickly using this system.

.902 Information quality:
.091   9. The system gives error messages that clearly tell me how to fix problems.
.054   10. Whenever I make a mistake using the system, I recover easily and quickly.
.498   11. The information (such as online help, on-page messages, and other documentation) provided with this system is clear.
.950   12. It is easy to find the information I need.
.749   13. The information provided by the system is easy to understand.
.661   14. The information is effective in helping me complete the tasks and scenarios.
.338   15. The organization of the information on the system pages is clear.

.712 Interface quality:
.654   16. The interface of this system is pleasant.
.950   17. I like using the interface of this system.
.237   18. The system has all the functions and capabilities I expect it to have.

.951 Overall satisfaction:
.951   19. Overall, I am satisfied with this system.

QUIS - Questionnaire for User Interface Satisfaction

Strongly disagree (1) – strongly agree (9): 9 point scale, with a NA option (0) in the end.

.219 Overall Reaction to the application
.316   1.   terrible – wonderful
.616   2.   difficult – easy
.154   3.   frustrating – satisfying
.803   4.   dull – stimulating
.900   5.   rigid – flexible

.758 Application
.810   6.   Reading characters on the page   hard – easy
.950   7.   Organization of information   confusing – very clear
.772   8.   Sequence of pages   confusing – very clear

.667 Terminology and Application information
.925   9.   User of term throughout application   inconsistent – consistent
.172   10.   Terminology is intuitive   never – always
.705   11.   Position of messages on screen   inconsistent – consistent
.253   12.   Prompts for input   confusing – clear
.317   13.   Application informs about its progress never – always
.588   14.   Error messages   unhelpful – helpful

.854 Learning
.132   15.   Learning to use the application   difficult – easy
.374   16.   Exploring new features by trial and error   difficult – easy
.620   17.   Performing tasks is straightforward   never – always
.733   18.   Help messages on the screen   unhelpful – helpful
.287   19.   Supplemental reference materials   confusing – clear

.196 Application Capabilities
.041   20.   Application speed   too slow – fast enough
.240   21.   Application reliability   unreliable – reliable
.385   22.   Sound associated with this application detracts value – adds value
.563   23.   Correcting your mistakes   difficult – easy
.756   24.   Designed for all levels of users   never – always

SUS- system usability scale

Strongly disagree (1) – strongly agree (5); 5 point scale

1. I think I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think I would need Tech Support to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I need to learn a lot about this system before I could effectively use it.

USE – usability evaluation

Strongly disagree (1) – strongly agree (5); 5 point scale

.757 Usefulness
.893   1. It helps me be more effective.
.789   2. It helps me be more productive.
.193   3. It is useful.
.798   4. It gives me more control over the activities in my life.
.677   5. It makes the things I want to accomplish easier to get done.
.899   6. It saves me time when I use it.
.949   7. It meets my needs.
.036   8. It does everything I would expect it to do.

.712 Ease of Use
.556   9.It is easy to use.
.796   10. It is simple to use.
.658   11. It is user friendly.
.152   12. It requires the fewest steps possible to accomplish what I want to do with it.
.086   13. It is flexible.
.306   14. Using it is effortless.
.256   15. I can use it without written instructions.
.851   16. I don't notice any inconsistencies as I use it.
.798   17. Both occasional and regular users would like it.
.849   18. I can recover from mistakes quickly and easily.
.378   19. I can use it successfully every time.

.322 Ease of Learning
.226   20. I learned to use it quickly.
.437   21. I easily remember how to use it.
.106   22. It is easy to learn to use it.
.524   23. I quickly became skilful with it.

.759 Satisfaction
.531   24. I am satisfied with it.
.313   25. I would recommend it to a friend.
.799   26. It is fun to use.
.118   27. It works the way I want it to work.
.653   28. It is wonderful.
.949   29. I feel I need to have it.
.363   30. It is pleasant to use.