The webmail experience

Gmail versus Share

Beau Verdiesen - 0813648

Supervised by: Dr. Louis Vuurpijl

Contents

1. Introduction

E-mail has revolutionized the way modern society works. Nearly all businesses use e-mail in one form or another. Besides the commercial aspect e-mail is also being used for personal reasons by nearly every person that has a computer with an internet connection. One might assume that interfaces for such a widely used communication system would have been fully developed over the years with little room for improvement. However, some of the most widely used e-mail clients such as Gmail, Yahoo mail and Outlook have undergone quite dramatic changes over the last few years. The look-and-feel of these clients have completely changed and functionality that was previously completely separated from e-mail such as chatting, keeping track of tasks, or maintaining contact information has slowly integrated with most modern e-mail clients.

Are all e-mail clients equally efficient and do they all have the same functionality? Probably not. They do however have a lot in common and the choice of what e-mail client to use comes down to preference. Most students at the Radboud University however agree that the e-mail client the university uses for handling mail (called 'Share') is more difficult to use than the e-mail client they use for their personal mail. Share is an adaptation of an e-mail client called 'Zimbra' that is designed specifically for large professional organizations such as companies and universities. Zimbra is the default webmail solution for Radboud University Nijmegen. It is a webmail client that resembles a full-fledged desktop application. This means that it has lots of functionality but may come across as 'cluttered'. There are a great number of weblogs that criticize or praise Zimbra but one of them really hits the spot: 'Review: Zimbra Webmail'[8]. In short, it says that Zimbra is great because of its functionality but lacks the ease-of-use and comfort of a webmail client like Gmail. So if you need a basic webmail client for handling mail Zimbra is probably not for you.

It has however never been reported that Share systematically performs worse than other webmail clients. That is what this thesis will try to do. The research presented in this thesis will assess whether the following hypothesis is true:

Share performs worse than other, freely available, webmail clients

In this thesis, first a survey will be presented about the functionality of five of the most widely used e-mail clients (Gmail, Yahoo mail, Windows Live mail, Mozilla Thunderbird and Microsoft Outlook) and the e-mail client the Radboud University uses (Share). The results of this survey shows which features people value the most and which are, in their opinion, obsolete.

Next, an experiment has been designed to compare the usability of Gmail and Share. In this experiment, human subjects performed a set of tasks that have been based on the results of the preliminary survey. The data gathered in this experiment have been used to analyze a number of different usability aspects. Any anomalies that occurred during the experiment were noted and reported on. After the experiment the subjects had to fill in a survey that showed the subjects emotional disposition towards both Share and Gmail.

This design ensures that the dislike that some student have against Share can be traced back to a fundamental usability aspect.

2. Theory

There are two main ways of measuring the usability of a system: Usability Testing and Usability Inspection.

Usability Inspection

Usability inspection looks at the performance of software that is being used by a user who, in most cases, is not an expert in the field of usability. There are a number of different ways to perform usability inspection of a system (Lodhi, 2010)[1]:

- *Surveys and Questionnaires*: A lot of usability aspects can be studied by simply asking users about their experiences by means of a survey or a questionnaire.
- *Observation*: Simply observe users of the system. Most often there are some heuristics and the test subject is asked to perform a set of tasks. The time it takes for a subject to complete these tasks can be measured. While the subject is performing the tasks the subject might also say what he or she is thinking. This is better known as the 'Thinking Aloud Protocol' (Cooke, 2010)[2].
- *User Experience*: More and more new methods focus on the experience the user has while using the system. This is supposed to be a good representation of the overall usability of the software.

Usability Testing

With usability testing, experts in the field of usability test a software system. There are couple of different ways to do this (Nielsen 1993)[3]:

- *Heuristic evaluation*: Experts look at every aspect of usability and give their opinion. This can be done alone or in groups.
- *Cognitive walkthrough*: This makes use of a somewhat more detailed procedure in which an expert tries simulating a user and checking if this user could perform a task with the information that is available to him.
- *Formal Usability Inspection*: A six-step procedure with strictly defined roles that combines heuristic evaluation with a simple form of cognitive walkthroughs.
- *Pleuristic Walkthrough*: Users, specialists and the design team come together to go through scenario's and identify possible problems.

In this thesis Usability testing is used to assess the usability of the two e-mail clients. The most widely used technique for this purpose is developed by Jakob Nielsen (Nielsen, 1994)[4]. He states there are five attributes that together define the usability of a software system: learnability, efficiency, memorability, errors, and satisfaction.

By testing each of these five attributes one can compare its scores with that of another interface which might give some insight to which interface is better. To really get to the core of the problem however users have to be observed while using the system. It is important to notice where a user is having trouble and how observed problems may be fixed. So, even though the five attributes of Nielsen give a nice overview of the usability of a system Nielsen's method does not offer any concrete solutions.

Modern interactive designs attract users by means of their appeal (Y. Rogers et al, 2011)[5]. For example, the product range of Apple is well-known for its sexy appearance and fancy look and feel.

Similar developments have resulted in, e.g., flashy web designs and desirable smart phones. Although the assessment of "user satisfaction" following Nielsen's technique may be used to measure these novel aspects of usability, recently an additional usability attribute has gained interest, namely the User Experience. This attribute evaluates both user satisfaction and emotional aspects of the design. The AttrakDiff method (see chapter 2.2) is particularly suited to evaluate this novel attribute.

The next two sections describe the five usability attributes from Nielsen and the AttrakDiff method for assessing the user experience, respectively.

2.1 Nielsen's Usability Attributes

2.1.1 Learnability

Learnability is one of the most fundamental aspects of usability. Learnability defines how well a user performs while using the system for the first time, that is, the user has no previous experience with the system. This also means that adapting or updating a system requires new users with no previous experience to assess the learnability. Testing learnability can be done by taking a set of tasks and setting a timeframe in which these tasks should be completed. The time it takes for a subject to successfully complete all tasks within this timeframe is a benchmark for the usability of the system. A second option is to set a certain 'performance' benchmark and see how long it takes for a user to reach this benchmark. Because new users usually jump right in and often don't take the time to fully explore every possibility of a system it is a good idea to set this benchmark at a level where a user can use the system reasonably well , instead of expecting a user to fully master the system.

2.1.2 Efficiency

Efficiency is about the performance of a test subject after he or she has learned how to use the system. With some complex systems it can take a lot of time to finish the learning phase while other systems have a very short learning curve. However, there is always room for improvement. To measure the efficiency experienced users are required. The definition of an 'experienced user' is key here. Are these people that have been working with the system for years, or people that have used the system for an X amount of time? In usability testing often a test subject is asked to spend a couple of hours working with the system, and thus learning how to use it. When measuring the performance of the subject while he or she is working with the system the learning curve is finished when the performance isn't increasing anymore. Subjects can be instructed to perform a certain set of tasks. The time the subject takes is the perceived efficiency. So, measuring the efficiency of a system requires a definition of an 'experienced user', a number of test subjects and a set of tasks to measure the performance.

2.1.3 Memorability

Memorability is closely linked to learnability. The difference is that Learnability is only applicable the first periods that someone learns how to use a system. An example by Nielsen about a 'Kiss and Ride' sign:

"[...]Initially, the meaning of this sign may not be obvious (it has poor learnability without outside assistance), but once you realize that it indicates a drop-off zone for commuters arriving in a car driven by somebody else, the sign becomes sufficiently memorable to allow you to find such zones at other stations (it is easy to remember)(p31).'

There are two ways to test memorability. First by measuring the performance of a user who hasn't been using the system for a while to check how well he or she remembered the functionality of the system. Secondly by asking subjects who have just finished testing to recall from memory how to perform certain tasks or to describe certain actions and ask what would be the consequences. The score would then be the number of correct answers. The first method is preferable because most interfaces are very visual and make it so that the user doesn't have to remember all that much.

2.1.4 Errors
A system should be able to minimize the number of errors a user makes. An error is defined as an action that doesn't bring you closer to your goal and the 'error rate' of a system can be measured by counting the number of errors during a certain set of tasks. This means that the error rate can be measured during another usability test. A number of different types of errors can be distinguished from one another. There are, for example, errors that can be corrected by the user, but also errors that are catastrophic for the completion of a task, without the user even knowing he did something wrong.

2.1.5 Satisfaction
Satisfaction tests how pleasurable it is to use the system. The importance of this depends on the system that is being tested. It seems logical that a game would yield more pleasure then a data-processing system. It has been determined that people that feel like they are in control of a system often have a more positive attitude towards that same system (Kay 1989). The satisfaction of a user can simply be determined by means of asking questions. In the most cases a small survey will suffice.

2.2 User Experience
AttrakDiff, designed by a company called 'User Interface Design GmbH' with the help of Prof. Dr. Marc Hassenzahl, is an instrument for measuring the attractiveness of interactive products. With the help of pairs of opposite adjectives, users (or potential users) can indicate their perception of the product by means of an online survey. These adjective-pairs make a collection of the evaluation dimensions possible.

The pairs of adjectives belong to either of the following dimensions (M. Hassenzahl et al, 2000[6] and M. Hassenzahl, 2003[7]):

Pragmatic Quality (PQ):
Describes the usability of a product and indicates how successful users are in achieving their goals using the product.

Hedonic Quality – Stimulation (HQ-S):
Mankind has an inherent need to develop and move forward. This dimension indicates to what extent the product can support those in needs in terms of novel, interesting and stimulating functions, contents en interaction- and presentation-styles.

Hedonic Quality – Identity (HQ-I):

Indicates to what extent the product allows the user to identify with it.

Attractiveness (ATT):

Describes a global value of the product based on the quality perception.

Hedonic and pragmatic qualities are independent of one another, and contribute equally to the rating of attractiveness.

3. Method

To evaluate the usability of mail clients, the following research has been conducted. First, in an online survey, the most prominent aspects that users from webmail clients find important such as being able to attach a file and searching your e-mail on various attributes have been explored. This included aspects like changing the font of text and changing contact information. In addition to these questions participants were also asked to list their favorite webmail clients.

Based on these findings a second experiment was designed to compare the most popular webmail client with Share mail on the aspects that were rated as most important (details in chapter 3.2). As a part of the experiment test subjects had to fill out two AttrakDiff survey's, one for Gmail and one for Share. The results were used to measure the User Experience. The AttrakDiff survey is attached in Appendix C.

3.1 Survey on webmail client usage

The first survey was designed to find out which aspects of an e-mail client are the most important. To achieve this, six e-mail clients were compared: Gmail, Yahoo mail, Windows Live mail, Mozilla Thunderbird, Microsoft Outlook and Share. All aspects of usability that might prove interesting were listed. This was done by comparing the e-mail clients and noting all functionality that most of them had in common. This list was later refined by generalizing some functionality that most clients have, but was slightly different in each client. An example of this is the chat functionality. In Yahoo mail one can opt to chat via Facebook, in Gmail via Google Talk and in Windows Live mail via MSN Live messenger. If one of the questions of the survey was to rate each of these chat-systems on how important it was for the test subjects the results would show one's preference for each of these systems. Instead these different chat systems were generalized and the question became how important it is to have *a* form of chatting with your contacts. In the survey all questions besides the first three were represented by the following five-point likert scale:

1. Very unimportant
2. Unimportant
3. Neutral
4. Important
5. Very important

The complete survey is added in Appendix A.

3.2 Experiment

In order to acquire the data necessary to investigate the hypothesis an experiment was carried out. This experiment consisted of three main indicators: two objective variables (efficiency and errors) and one subjective variable (the User Experience). Gmail was chosen as a comparison to Share, this is because the preliminary survey showed that Gmail is most used (see chapter 4.1) and thus, the most ideal candidate for comparison. In the experiment test subjects were asked to carry out a total of eleven tasks as quickly and accurately as possible in either Gmail or Share (instructions in Appendix B). After completing the list of tasks the test subjects had to carry out the same tasks in the other e-

Theoretic variable	Indicator	Empiric variable
Usability	Efficiency	The time it takes to complete the list of tasks
	Errors	The number of errors made during the experiment
	User Experience	AttrakDiff web survey

mail client. Finally the subjects were asked to fill out the AttrakDiff survey for both Gmail and Share to subjectively measure the User Experience.

The most highly valued aspects of functionality were selected for the tasks. Note that tasks that are very similar, such as the various search functions have not all been selected. Functionality that is difficult or impossible to test such as sorting e-mail on date (in both e-mail clients, e-mail is sorted on date as default) have also been omitted. Instead the most valued tasks that are notably different have been selected to get the widest possible base for the tasks. Following is an overview of these tasks. A functional comparison of these tasks will follow in chapter 3.3.

Aspects:

1. Attaching a file
2. Searching through e-mail
3. Replying on an e-mail
4. Delete an e-mail
5. Using folders
6. Edit a received e-mail as a new e-mail
7. Mark e-mail as spam

Task overview		
	Task	*Aspect*
1	Send e-mail with attachment	1
2	Reply to all	3
3	Search e-mail on attachment	2-I
4	Delete e-mail	4
5	Create a folder	5-I
6	Search e-mail on subject	2-II
7	Move e-mail to folder	5-II
8	Search e-mail on sender	2-III
9	Edit a received mail as new	6
10	Mark e-mail as spam	7
11	Search e-mail on date	2-IV

A single examiner conducts the experiment. He sits next to the test subject and times the experiment by means of a stopwatch. He also writes down any observations on the form in Appendix D as well as the final time. Because the experiment was conducted by a single person the decision has been made to measure the total time rather than the time per task. This way, the person conducting the experiment was also able to count the number of errors, as well as take notes on any problems the test subject might have. This means that the results will show what the problem is, rather than just during which task the problem occurs.

Students from the Radboud University were chosen as test subjects. The reason for this is that it will keep the group as homogenous as possible and they will at least have some experience with the Share RU mail system. Because webmail clients are applications that are used almost every day it is not necessary to test its memorability. Because of the scope of the experiments learnability has also been omitted, it would have required too much effort to have test subjects come back after a couple

of weeks for another experiment. The satisfaction attribute of Nielsen's technique has been substituted for the User Experience and was tested by means of the AttrakDiff survey.

The experiment has a 'within subject design' meaning that test subjects performed a set of tasks within both Gmail and Share. The design was counterbalanced by randomizing which client was tested first to cancel out any uncontrolled learning effects. The versions of the clients that were used were the versions that were online in July '12. Because the experiment was meant to be as realistic as possible the e-mail accounts of a student from the Radboud University were used for the tasks (a Gmail and a Share account). The e-mails that were used for testing were real e-mails that were present on both accounts.

3.3 Functional Comparison

This chapter consists of a comparison of how you can carry out each task in Gmail and in Share.

Attaching a file

This is done in relatively the same way in both clients. You click the 'add attachment button' and select a file. In Share however you have to click an extra button 'Open file' to select a file and Gmail shows a progress bar to visualize the upload.

Searching through e-mail

Both clients have a search bar at the top of the screen with which you can search on basic terms such as sender/recipient, subject, label or words that occur within the e-mail. Advanced search works a bit different for both clients. In Gmail, you have a small drop-down window that gives you the added possibility to also search on date (or a range of dates) and on words that should not be included in the search. Also worth mentioning is that Gmail has an autocomplete function. This means that it will try to complete your search term whilst typing. It will complete with contacts and labels as well as completing terms like 'has photo' or 'has document'.
The advanced search in Share works a bit different. It lacks the auto-completion that Gmail possesses but you have far more options. You can, for instance, search on the size of an attachment, the domain form which an e-mail has been sent or whether an e-mail has been read or not.
The bottom line is that Gmail's search works more intuitively while Share's search has more possibilities. The basic search however, is comparable.

Replying on an e-mail

The position of the reply buttons differs between clients. In Gmail these are located below the e-mail as well as at the top-right corner of the e-mail in a drop-down menu. In Share these buttons are located in the menu bar above the e-mail. Gmail also provides a more intuitive way to reply on an e-mail. Below the original e-mail there is a box that says 'Click here to reply'. This works intuitively because the box is the same length as the box you would get when you write a reply.

Deleting an e-mail

In both clients you can delete an e-mail by checking it's box in the inbox and clicking on the 'delete' button. When you open an e-mail in Gmail you can delete it straight away by simply pressing the 'delete' button. Because Share opens an e-mail below the inbox rather than over the inbox it does not possess this function.

Creating and moving an e-mail to a self-made folder

In Gmail you create a new label via the folder menu on the left side of the screen. You can choose a name and a color for the label. Adding a label to a folder can be done by either dragging the e-mail to the label, or by dragging the label onto the e-mail. Clicking the label will result in a list of e-mails currently containing that label.

In Share labels are called 'folder' but the two have the same function. When creating a new folder you can only give it a name. It is however possible to give it a color by editing it after creation.

Edit a received e-mail as a new e-mail

Gmail does not have an option to edit a received e-mail as a new e-mail. You'll have to copy the contents of the e-mail, paste them in a new e-mail and manually edit the subject.

Share does have the option; you can edit a received e-mail as a new e-mail by right clicking on the e-mail and selecting 'Edit as New' from the drop-down menu.

Mark e-mail as spam

In both e-mail clients you can simply select an e-mail and click on the 'mark as spam' button in the menu.

4. Results

This chapter contains the results from the preliminary survey (4.1) as well as the experiment (4.2) and some of the more interesting observations (4.3). The results of the experiment are described in some detail and conclusions based on these results are drawn in chapter 5.

4.1 Results preliminary survey

Following are the results of the initial survey. In total, 44 participants contributed to these results with an average age of 22,86 and an average experience of over ten years. The survey was distributed among students from the Radboud university.

Regarding the tables, the numbers next to the question represents the average score, the lowest score 'very unimportant' counting as 1 and the highest score 'very important' counting as 5.

It is interesting to see that Yahoo mail is not used at all. Most people have experience with some, or all, of the other e-mail clients with Gmail and Windows Live Mail being to most popular e-mail client the test subjects are currently using. Looking at the results we see that people do not care too much about the extra functions in e-mail clients such as a chat or a calendar function. Neither do they care much for editing their contacts.

First, the results will be grouped by category and the final table gives a ranking of all questions.

General Data	
Total number of participants	44
Male	35
Female	9
Average age	22,86
Average years of experience	10,64

What e-mail clients do you have experience with?		
	#	*Percentage*
Microsoft Outlook	28	63,6 %
Gmail	39	88,6 %
Share RU mail (Zimbra)	30	68,2 %
Mozilla Thunderbird	14	31,8 %
Windows Live Mail	34	77,3 %
Yahoo Mail	3	6,8 %

Which e-mail client do you currently use the most?		
	#	*Percentage*
Microsoft Outlook	3	6,8 %
Gmail	22	50,0 %
Share RU mail (Zimbra)	1	2,3 %
Mozilla Thunderbird	5	11,4 %
Windows Live Mail	13	29,5 %
Yahoo Mail	0	0 %

Inbox	
Search through e-mail on sender	4,71
Sort e-mail on date	4,60
Search through e-mail on topic	4,40
Copy the contents of an e-mail	4,27
Search through e-mail on content	4,14
Search through e-mail on date	4,09
Delete e-mail	3,97
Move e-mail to a self-made folder	3,94
Search through e-mail on recipient	3,94
Edit a received e-mail as a new e-mail	3,74
Search through e-mail on attachment	3,68
Check e-mail as spam	3,63
Sort e-mail on sender	3,51
Organizing e-mail by assigning labels to individual e-mails	3,41
Sort e-mail on topic	3,18
Sort e-mail on recipient	3,06
Archive mail in a default folder	2,97
Having your e-mail client auto-complete a search term	2,85
Assigning priority to certain e-mails	2,82
Having multiple tabs	2,65

Composing new mail	
Being able to attach a file	4,91
Having the e-mail client auto-complete e-mail addresses	4,43
Seeing a confirmation when an e-mail has been sent	3,91
Have the e-mail client automatically save a concept every few seconds	3,71
Being able to put text bold/italic/underlined	3,46
Being able to use emoticons	3,36
Being able to change the font of text	2,97
Composing new mail in a pop-up	2,80
Being able to change the color of text	2,57

Contacts	
Save e-mail addresses that you have ever mailed to	3,97
Have the ability to automatically join contacts that appear twice	3,59
Being able to assign basic information such as Name, Address, etc.	3,49
Being able to organize your contacts in categories	3,23
Import contacts from other e-mail clients	3,23
Being able to add a picture of the contact	2,29

General	
Being able to 'reply to all'	4,69
Being able to forward mail	4,56
Have your privacy safeguarded	4,43
See the exact time an e-mail was sent	4,09
Having conversation threads	3,86
Being able to use multiple accounts within the e-mail client	3,83
The ability to use the e-mail client on your smartphone	3,73
Viewing office attachments such as text documents within the e-mail client	3,57
Being able to print the e-mail	3,37
Being able to Share pictures	2,79
Being able to use keyboard shortcuts	2,71
Being able to use a calendar within the e-mail client	2,51
Being able to use some form of chat to chat with your contacts	2,51
Being able to set a personalized signature	2,43
Being able to keep a list of tasks within the e-mail clients	2,17
See a picture of the person that sent the e-mail	2,03

All questions (ranked)		
1	Being able to attach a file	4,91
2	Search through e-mail on sender	4,71
3	Being able to 'reply to all'	4,69
4	Sort e-mail on date	4,60
5	Being able to forward mail	4,56
6	Having the e-mail client auto-complete e-mail addresses	4,43
7	Have your privacy safeguarded	4,43
8	Search through e-mail on topic	4,40
9	Copy the contents of an e-mail	4,27
10	Search through e-mail on content	4,14
11	Search through e-mail on date	4,09
12	See the exact time an e-mail was sent	4,09
13	Delete e-mail	3,97
14	Save e-mail addresses that you have ever mailed to	3,97
15	Move e-mail to a self-made folder	3,94
16	Search through e-mail on recipient	3,94
17	Seeing a confirmation when an e-mail has been sent	3,91
18	Having conversation threads	3,86
19	Being able to use multiple accounts within the e-mail client	3,83
20	Edit a received e-mail as a new e-mail	3,74
21	The ability to use the e-mail client on your smartphone	3,73
22	Have the e-mail client automatically save a concept every few seconds	3,71
23	Search through e-mail on attachment	3,68
24	Check e-mail as spam	3,63
25	Have the ability to automatically join contacts that appear twice	3,59

26	Viewing office attachments such as text documents within the e-mail client	3,57
27	Sort e-mail on sender	3,51
28	Being able to assign basic information such as Name, Address, etc.	3,49
29	Being able to put text bold/italic/underlined	3,46
30	Organizing e-mail by assigning labels to individual e-mails	3,41
31	Being able to print the e-mail	3,37
32	Being able to use emoticons	3,36
33	Being able to organize your contacts in categories	3,23
34	Import contacts from other e-mail clients	3,23
35	Sort e-mail on topic	3,18
36	Sort e-mail on recipient	3,06
37	Archive mail in a default folder	2,97
38	Being able to change the font of text	2,97
39	Having your e-mail client auto-complete a search term	2,85
40	Assigning priority to certain e-mails	2,82
41	Composing new mail in a pop-up	2,80
42	Being able to Share pictures	2,79
43	Being able to use keyboard shortcuts	2,71
44	Having multiple tabs	2,65
45	Being able to change the color of text	2,57
46	Being able to use a calendar within the e-mail client	2,51
47	Being able to use some form of chat to chat with your contacts	2,51
48	Being able to set a personalized signature	2,43
49	Being able to add a picture of the contact	2,29
50	Being able to keep a list of tasks within the e-mail clients	2,17
51	See a picture of the person that send the e-mail	2,03

4.2 Results Experiment

This chapter will present the results of the experiment. Each table or figure is accompanied by a short description of the results, pointing out some of the more interesting results.

A total of eleven subjects participated in the experiment, three female and eight male, with various degrees of experience and success. The average age was 22.1 years.

Results Gmail				
	Time	Errors	1st/2nd	Experience
Subject 1	0:05:59	13	2	2
Subject 2	0:06:32	2	2	3
Subject 3	0:10:21	14	1	1
Subject 4	0:06:06	3	1	2
Subject 5	0:06:33	10	1	1
Subject 6	0:08:23	5	1	3
Subject 7	0:08:52	10	1	3
Subject 8	0:05:10	1	2	3
Subject 9	0:08:00	5	2	1
Subject 10	0:03:52	3	2	3
Subject 11	0:04:36	1	1	3

Table 1: Results Gmail

Averages Gmail		
	Time	Errors
Total average	0:06:46	6,09
Average 1st	0:07:28	7,17
Average 2nd	0:05:55	4,80
Average exp1	0:08:18	9,67
Average exp2	0:06:02	8,00
Average exp3	0:06:14	3.67
Average experience	2,27	

Table 2: Averages Gmail

When looking at the results in table 1 one would expect that the amount of errors would increase as the time increases. We can however see two exceptions to this idea. Subject 1 notes a relatively fast time (5:59) with a relative high amount of errors (13). This was due to the fact that the subject had difficulty finding some of the more advanced options, and tried clicking everything to see where such options might be found. Because the subject tested Gmail second he already knew what to do so when an option was found the subject quickly finished the task.

The other exception is subject 6, this person was relatively slow (8:23) while making little errors (just 5). The reason for this is that this person took it's time to accurately read the instructions.

When looking at the averages (table 2), most of the results are as expected, the subjects that tested Gmail first were slower (7:28 vs. 5:55) and made more mistakes (7.17 vs. 4.80). However when looking at averages of the subjects with medium (exp2) and high (exp3) experience we can see that the time it took to complete the experiment doesn't differ very much (6:02 vs. 6:14) but the errors that were made do differ (8.00 vs. 3.67). Although there seems to be no apparent reason for these results it is worth noting that there were only two subjects that had medium experience with Gmail which might not represent the entire group.

	Results Share			
	Time	Errors	1st/2nd	Experience
Subject 1	0:07:22	8	1	1
Subject 2	0:07:51	3	1	1
Subject 3	0:09:07	10	2	1
Subject 4	0:04:18	3	2	3
Subject 5	0:04:13	2	2	1
Subject 6	0:07:08	2	2	1
Subject 7	0:05:28	3	2	1
Subject 8	0:05:41	1	1	2
Subject 9	0:07:38	6	1	2
Subject 10	0:04:49	1	1	1
Subject 11	0:04:25	1	2	3

Table 3: Results Share

Averages Share		
	Time	**Errors**
Total average	0:06:11	3,64
Average 1st	0:06:40	3,80
Average 2nd	0:05:46	3,50
Average exp1	0:06:34	4,14
Average exp2	0:06:39	3,50
Average exp3	0:04:22	2,00
Average experience	1,55	

Table 4: Averages Share

The results from the Share experiment are somewhat more in the lines of what one would expect. The only exception is subject 6 who, once again, has a relatively slow time (7:08) while making few mistakes (2). This time this wasn't caused by taking a long time reading as this subject tested Share second but the subject was carefully interpreting the interface to look for the right button hence not making many mistakes.

When looking at the averages the only thing that stands out is the minute difference in time between the subject with low (exp1) en medium (exp2) experience. One would expect the subject with medium experience would be somewhat faster. This might be explained by the fact that there were only two subjects with medium experience which might not be representative for the entire group.
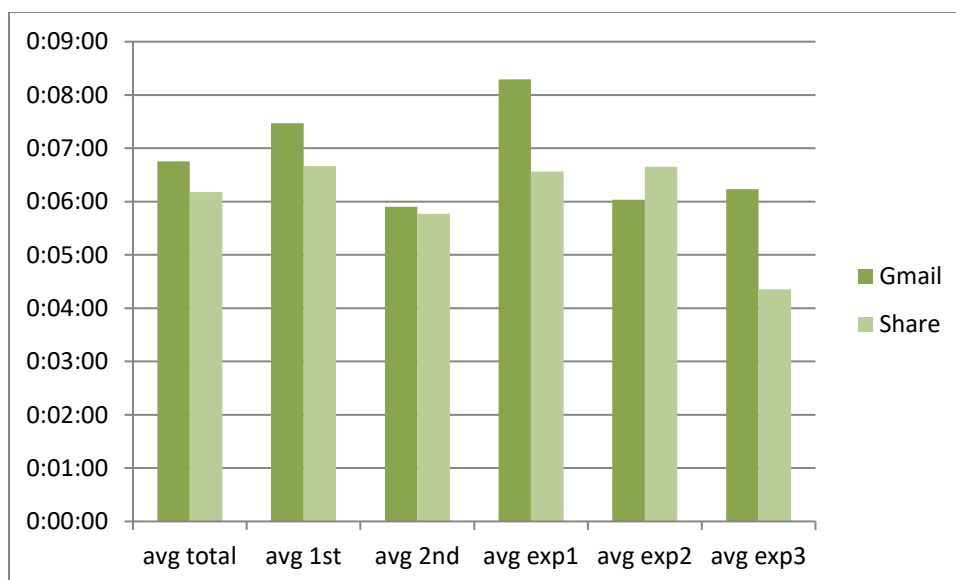

Figure 1: Time comparison

When comparing the average times of Gmail and Share (figure 1) we notice the average times of the low (exp1) and high (exp3) experienced subject is higher with Gmail than with Share. The reason for this is explained in the first part of the conclusion (chapter 5)
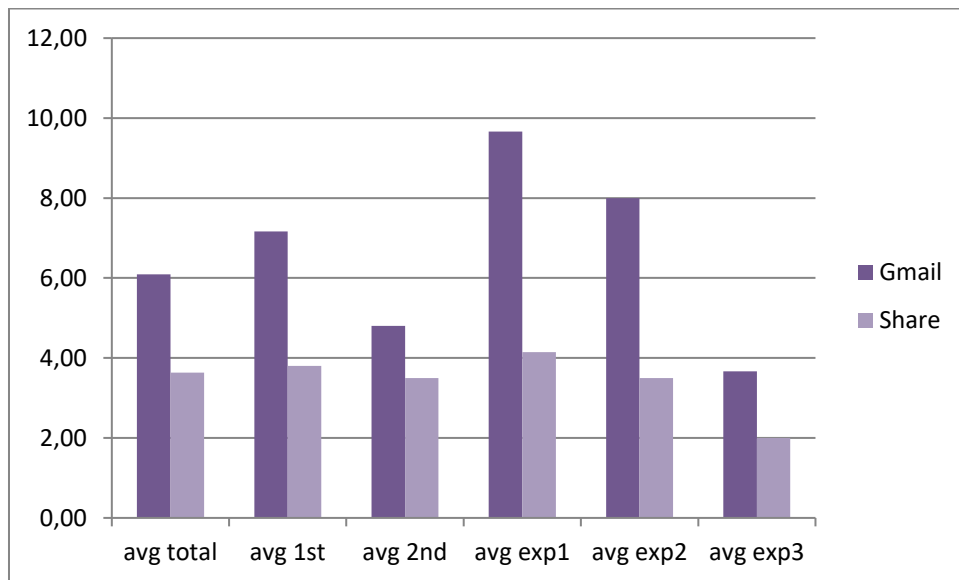
Figure 2: Error Comparison

The comparison of the errors between Gmail and Share (figure 2) shows that the error-rate with Gmail is consequently higher than with Share. The reason for this is explained in the second part of the conclusion (Chapter 5)
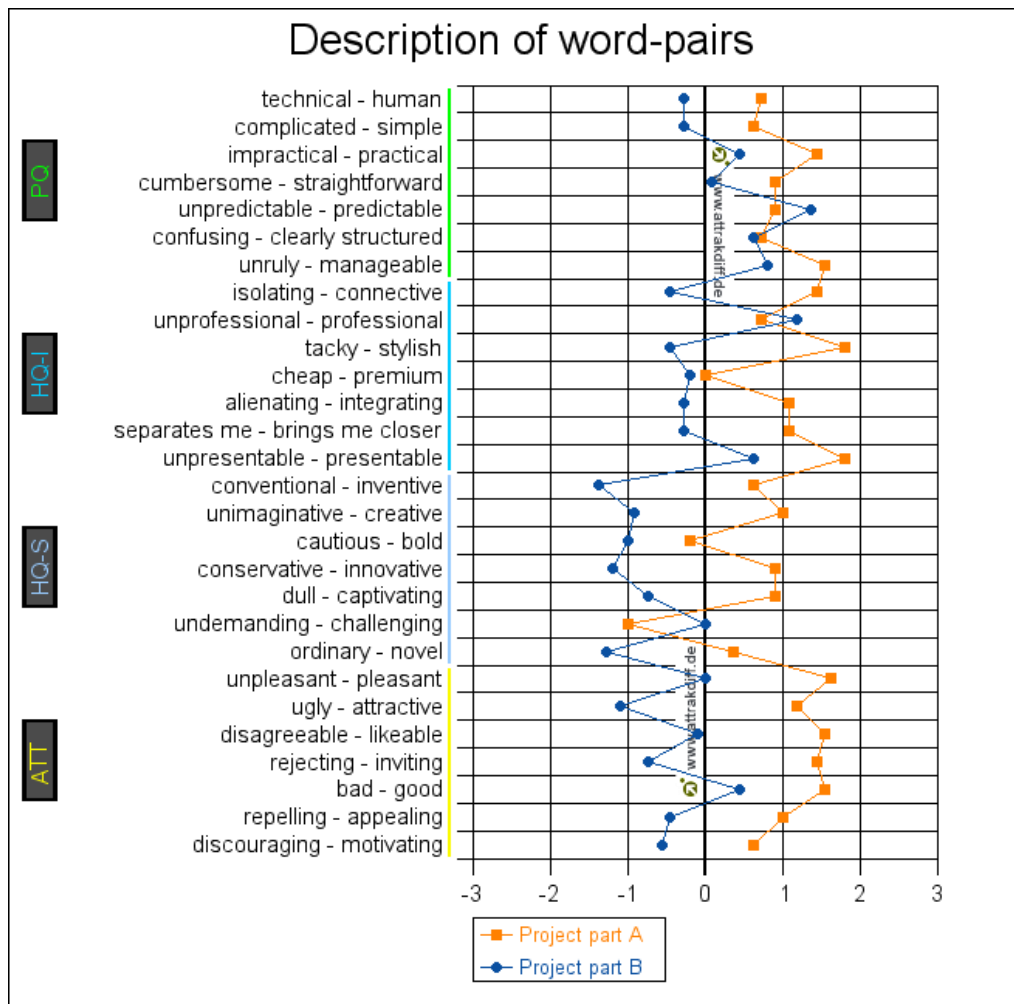
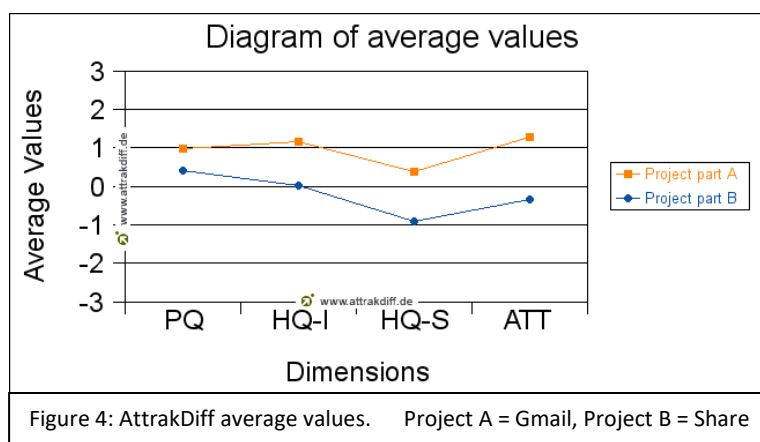Figure 3: AttrakDiff word-pairs.   Project A = Gmail, Project B = Share

Figure 4: AttrakDiff average values.   Project A = Gmail, Project B = Share

The results from the AttrakDiff survey show that Gmail scores better on most word-pairs (figure 3). Some of the more interesting pairs are for instance 'rejecting-inviting', and 'ugly-attractive' where the distance between Gmail and Share is largest (roughly 2 points). This, in combination with the rest of the ATT values shows that the test subjects think that Gmail look more attractive than Share. Even though Gmail scores better on most pairs there are three of them where Share has the upper hand. These are 'undemanding-challenging', 'unprofessional-professional' and 'unpredictable-predictable'. The reason for this might be that Share shows more options in its interface, making it look more professional (only a professional would demand some options), more predictable (no need to click through menu's) and more challenging (more options increases difficulty).
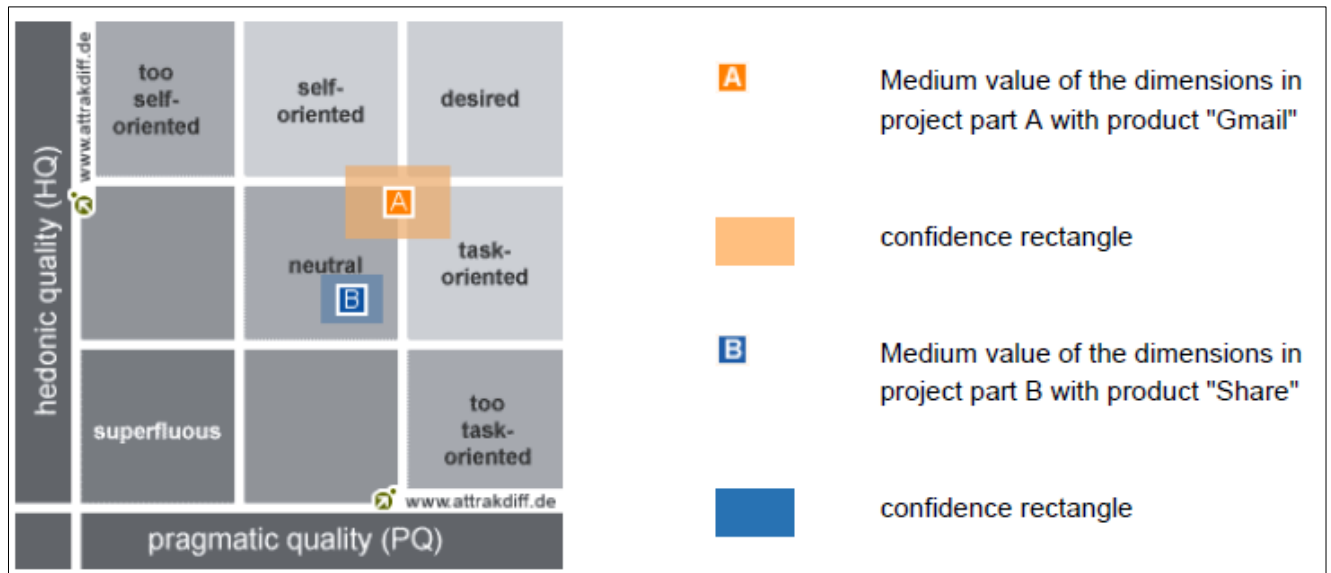
Figure 5: AttrakDiff confidence rectangle

Figure 5 shows the so-called confidence rectangle that results from the AttrakDiff survey. This image shows pragmatic quality on the horizontal axis and hedonic quality on the vertical axis. This means that the further north-east the rectangle is placed the more desirable the product is. The larger the rectangle the more the individual result vary.

We can clearly see that even though the confidence rectangle belonging to Gmail (A, orange) is larger, meaning the individual test subjects don't agree as much as they do with Share, it is still further to the north-east than the rectangle belonging to Share (B, blue).

4.3 Observations

In this chapter some of the more interesting observations that were made during the experiment are listed. For structure purposes these will be presented per task, and only if there were observations of interest.

Task 2: replying on an e-mail.
There was a mix of ways that people used to reply on the e-mail. In Share both the toolbar and the reply button you get when double-clicking the e-mail were used and in Gmail both the quickbox below the e-mail and the arrow in the upper-right corner were used. It seems different users prefer different ways to send a reply. There is no definitive preferred way.

Task 5: creating a folder
A lot of test subject had difficulties with this task in Gmail. The reason for this is that in order to create a new label one has to click on 'more' beneath the current labels and then, all the way at the bottom of the menu, there is a blue html link that allows you to create a new label. The way this link is positioned and because it is in an old html-style people tend to overlook it. A second way of creating a new label is in the 'settings' menu, which one test subject cleverly made use of because he couldn't find the normal button.

Task 7: Moving an e-mail to a folder
Just two test subjects dragged the e-mail to the folder in Gmail and none of the test subjects used this feature in Share. The reason for this might be that most test subjects opened the e-mail before trying to add a label. In Gmail this means you can no longer drag the e-mail because you need to be in the inbox-view to do so.

Task 9: Editing as new
When asked, only a single test subject was aware of the feature to edit an e-mail as your own. This strongly contradicts the result of the preliminary survey that valued this functionality with an average of 3,74 on a 5-point scale (see chapter 4.1). There seems to be no explanation for the discrepancies between the test subjects and the participants of the preliminary survey.

Task 11: Searching on date
The final task was by far the most difficult for most test subjects. Nearly all subjects tried typing the date in the search box which yield no results in either webmail client. In Gmail the subjects had a lot of difficulty finding the advanced search option where you can search on date and when they found it had a lot of difficulty with the fact that this option uses the middle-endian style of dates, starting with the month, rather than the little-endian style, starting with the day, that is used in Europe. The test subjects can't be blamed for this as the resulting confusion is through no fault of their own.

General Observations
Users that were unfamiliar with Gmail tend to think the Gmail search bar is actually the Google search bar, meaning they think it will search on the web rather than through your e-mail. This confusion is understandable seeing as the search bar has the Google logo in front of it closely resembling the actual Google search bar. Also, once you start typing in the search bar it will prompt you if you want to search on the web, adding to the confusion.

5. Conclusion

Gmail and Share are very much alike when it comes to efficiency. There are just two things that stand out: the results of the test subjects with Low and High experience (figure 1). Both groups seem to perform better in Share. The case of the Low experience users can be explained by the fact that Gmail uses as few buttons and text as possible whereas with Share a lot of options are represented by both a button and text. This makes is easier for first time users (or users with little experience) to find the right button easily. With Gmail you have to hover over the buttons to see what they do and click on menu buttons to get more options to appear. This consequently explains the huge difference in errors with low experience users (figure 2).

The difference in time between high experience users might be explained by the groups themselves, of the eleven test subjects only 2 had high experience with Share which might not be representable.

When looking at the errors we can see that Share performs a lot better, again this can be explained by the lack of text in Gmail, and the abundance of buttons in Share. However, when looking at the experience the users have, the amount of errors drops more significantly with Gmail than with Share. This, again, supports the idea that once users get familiar with the buttons in Gmail they no longer need to hover over the button to know what the icon on the button stands for.

Regarding the AttrakDiff results we can see that on all four aspects, Pragmatic Quality (PQ), Hedonic Quality – Identity (HQ-I), Hedonic Quality – Stimulation (HQ-S) and Attractiveness (ATT) Gmail scores better then Share. On three of those aspects (all but PQ) the difference is statistically significant. This was tested by means of a T-Test with a significance standard of 0,05. The fact that the difference in PQ is not as large as the other three is easily explained when looking at the results of the experiment. The experiment purely tested pragmatic qualities and, just like the AttrakDiff survey, showed little difference between the two. The fact that the difference between other three values is significant shows that the test subjects can identify themselves more easily with Gmail (HQ-I), think Gmail is more interesting and novel (HQ-S) and overall looks better (ATT).

When looking at the confidence rectangle (figure 4) we can see that Gmail edges more to the 'desirable' region and Share is set in the 'neutral' region. Because the difference between that two is larger on the vertical plane (Hedonic) than the horizontal plane (Pragmatic) is once again supports the idea that pragmatically the two don' t differ very much, but hedonically they do.

Although these results seem to reliably indicate that Gmail is superior in terms of look and feel there is another aspect one has to take into account. The reason that most students don't actively use Share is because they already use a different e-mail client. And why would one switch to Share if they are already familiar with a different e-mail client and are able to simply handle all of their university mail in that client?

The question this thesis posed was whether *Share performs worse than other, freely available, webmail clients*. By means of a preliminary survey, an experiment and the AttrakDiff survey we can conclude that even though pragmatically Share performs the same, and in some area's better, than Gmail people find the experience that Gmail offers, in terms of looks and hedonic properties, to be superior of Share's. This, combined with the notion that most student already use a different e-mail client when arriving at the university explains why they do not actively make use of Share.

6. Discussion

The results of this thesis very promising and provide a solid basis on further research. The hypothesis posed in the beginning has been answered by means of careful research. The results pose some interesting new questions such as 'why do people find Gmail more attractive?' and 'why do people who use Gmail for the first time make so many errors?'.

However, because time was limited in this Bachelor Thesis it was not possible to fully examine all aspects of the subject. There are a couple of aspects one can be improved upon. First the general scope. In this thesis only eleven subjects were tested. If more subjects were to participate in the experiment the differences between the levels of experience will become more clear and might prove to give some more insight.

Second, one could test more usability metrics than just errors and efficiency. Testing the learnability and memorability was omitted from this thesis because it was both time-consuming and not strictly necessary. However, testing these two attributes might better explain why people with different levels of experience have more or less difficulty in either Gmail or Share seeing as this is typically caused by learnability and memorability.

Third, one could time each task individually to quantitatively show which specific task proves more difficult in either Gmail or Share. This might however require a different design because timing a single search task might prove to be too quick to time accurately.

Overall this thesis provides a solid basis for further research, not only for e-mail clients but for any type of software.

7. Literature

[1]   A. Lodhi (2010); Usability Heuristics as an Assessment Parameter: for Usability Testing; 2nd International Conference on Software Technology and Engineering (ICSTE); pp. 256 - 259

[2]   L. Cooke (2010); Assessing Concurrent Think-Aloud Protocol as a Usability Test Method: A Technical Communication Approach; IEEE TRANSACTIONS ON PROFESSIONAL COMMUNICATION, Volume 53 Issue 3; pp. 202 - 215

[3]   J. Nielsen (1993); Usability Engineering J. Nielsen (1994); Usability Inspection Methods; Conference: Computer-human interaction conference (CHI '94); pp. 413 - 414

[4]   J. Nielsen (1993); Usability Engineering

[5]   Y. Rogers, H. Sharp, J. Preece (2011); Interaction Design: Beyond Human - Computer Interaction

[6]   M. Hassenzahl, A. Platz, M. Burmester, K. Lehner (2000); Hedonic and Ergonomic Quality Aspects Determine a Software's Appeal; Conference: Computer-human interaction conference (CHI 2000); pp. 201 - 208

[7]   M. Hassenzahl (2003); The thing and I: understanding the relationship between user and product; M. Blythe, C. Overbeeke, A.F. Monk, P.C. Wright (Eds.), Funology: From Usability to Enjoyment, Kluwer, Dordrecht (2003), pp. 31 - 42

[8]   Paul Stamatiou (2008); Review: Zimbra Webmail; http://paulstamatiou.com/review-zimbra-webmail

8. Appendix

A. Survey

"I'm a student Information Science and I'm doing a study about the usability of e-mail clients. That means that I'm trying to find out what functionality of e-mail clients people think is important. With this information I will then be able to assess the usability of various e-mail clients. I made a list of functionality that most e-mail clients have in common but if you feel I missed something please feel free to write down your comments in the final question. The entire survey will take about 5 minutes to complete. I want to thank you in advance for filling in the survey."

General Questions

* Male/Female
* Age
* Number of years of experience with e-mail
* Which e-mail clients do you have experience with?   a. Microsoft Outlook
b. Gmail
c. Yahoo Mail
d. Share RU mail (Zimbra)
e. Mozilla Thunderbird
f. Windows Live Mail

* Which e-mail do you currently use the most?   a. Microsoft Outlook
b. Gmail
c. Yahoo Mail
d. Share RU mail (Zimbra)
e. Mozilla Thunderbird
f. Windows Live Mail

(The order of the e-mail clients was randomized)

How important are the following aspects of an e-mail client to you?

Inbox
The possibility to:
1. check e-mail as spam
2. archive e-mail in a default archive folder
3. delete e-mail
4. move e-mail to a self-made folder
5. copy the contents of an e-mail
6. edit a received e-mail as a new e-mail
7. organize my e-mail by labeling individual e-mails
8. assign priority to certain e-mails
9. search e-mail on: a. date
b. sender
c. topic
d. content
e. with or without attachment
10. have an e-mail client auto complete a search-term
11. have multiple tabs open, allowing you to quickly switch between e-mails
12. sort your e-mail on: a. date
b. sender
c. topic
d. recipient

Composing new mail

The possibility to:

1. compose new mail in a pop-up
2. have the e-mail client auto complete e-mail addresses of contacts
3. personalize e-mail: a. font
                                   b. color
                                   c. bold/italic/underline
                                   d. emoticons
4. attach a file
5. have a confirmation the e-mail has been sent
6. have the e-mail client automatically save a concept of your mail every few seconds

Contacts

The possibility to:

1. save the e-mail addresses of people that you have mailed or have mailed you
2. have lists or different categories of contacts
3. import contacts from another e-mail client
4. be able to assign information such as Name, Address, Phone Number, etc.
5. add a picture of the contact
6. have the e-mail clients join double contacts together

General

The possibility to:

1. make use of keyboard shortcuts
2. have a calendar
3. be able to keep a list of tasks (a to-do list)
4. chat with your contacts within the e-mail client (through Google Talk, Live Messenger, Facebook Chat, etc.)
5. Share pictures
6. have multiple accounts within the same client (for instance, access to your Gmail account within your Windows Live mail account)
7. have easy access to your account via smartphone
8. have your privacy guarded
9. view office attachments such as text documents within the e-mail client
10. set a personalized signature
11. have conversation threads (grouping all reply's with their original message in order)
12. reply
13. reply to all
14. forward
15. see the time the a-mail was sent
16. see a picture of the person who send the e-mail
17. print the e-mail

B. Instructions

"Please perform the following tasks in order. After you finished a task, please continue with the next. If you are having difficulty understanding the task you are free to ask for help. However, it may occur that certain questions cannot be answered. You are encouraged to say what you are thinking while performing the tasks. Any comment you give may be helpful for the research."

1. Send this e-mail:

(The attachment can be found on the <u>desktop</u>):

> **To:** John Doe
> **Subject:** Greetings
> **Attachment:** Greetings.doc
>
> Hello World

2. Open the e-mail with the subject 'Budget Report' in the inbox and send the following reply:

> I've underlined all the savings.

3. Search for an e-mail with an attachment called 'invitation.pdf' and open it.

4. Delete the e-mail you just found.

5. Create a folder/label named 'Important!'

6. Search for an e-mail with the subject | 'BFCA-INIII-3A-2003: rectificatie bericht tentamen wiskunde 1b' |.

7. Move the e-mail to the 'Important!' folder.

8. Search for all e-mails that were ever send by a person called 'Paul Groot'

9. Edit the e-mail with the subject | 'NWI-HC003-2011-KW3-V: Overzicht college Het Heelal' | as if it were your own e-mail and change the name 'Paul Groot' to 'Iston Volpe' at the end. Then send it to 'John Doe'

10. Mark the e-mail with subject | `Wil Weg Dagen 2012´ | as spam.

11. Search for the e-mail that has was sent on 07/03/**2011** and open it

Thank you for your cooperation.

C. AttrakDiff Survey

The survey consists of the following dimensions on a 7-point scale.

Human	↔	Technical
Isolating	↔	Connective
Pleasant	↔	Unpleasant
Inventive	↔	Conventional
Simple	↔	Complicated
Professional	↔	Unprofessional
Ugly	↔	Attractive
Practical	↔	Unpractical
Likeable	↔	Disagreeable
Cumbersome	↔	Straightforward

Stylish	↔	Tacky
Predictable	↔	Unpredictable
Cheap	↔	Premium
Alienating	↔	Integrating
Brings me closer to people	↔	Separates me from people
Unpresentable	↔	Presentable
Rejecting	↔	Inviting
Unimaginative	↔	Creative
Good	↔	Bad

Confusing	↔	Clearly structured
Repelling	↔	Appealing
Bold	↔	Cautious
Innovative	↔	Conservative
Dull	↔	Captivating
Undemanding	↔	Challenging
Motivating	↔	Discouraging
Novel	↔	Ordinary
Unruly	↔	Manageable

D. Results form

Name:	Age:	M / F	Date:
Gmail / Share	1st / 2nd		Low / Medium / High
Errors:			Time:
1.			
2.			
3.			
4.			
5.			
6.			
7.			
8.			
9.			
10.			
11.			