Usability Evaluation: Tasks Susceptible to Concurrent Think-Aloud Protocol

Juliana Anyango Ogolla
Human-Centered Systems (HCS)
Department of Computer and Information Science (IDA)

MASTER THESIS-TQDV30
LIU-IDA/LITH-EX-A--11/034--SE

Supervisor: Johan Åberg
Abstract

Think-aloud protocol is a usability testing method whereby the participant running the usability test on an interface, thinks aloud as a way of giving feedback of the task he/she is performing on the given interface. It is one of the most researched on usability testing methods. It has attracted both praises and criticisms based on the effects it has on the participants or the tests at hand. A recently done study that used simple tasks, aimed at finding out the difference between using think-aloud protocol and not using think-aloud protocol. The study concluded that no notable differences were evident on the number of fixations and the amount of screen areas viewed when using think-aloud protocol and when not using think-aloud protocol.

As an extension and follow-up of the recently done study, this study focused on finding the type of tasks that the concurrent think-aloud protocol has effects on. The tasks were chosen based on the information scent concept and eye-tracking methodology was used in collecting the necessary results.

The study that involved twenty participants, resulted to some effects of the concurrent think-aloud protocol being noted on the low-scent tasks but not on high-scent tasks. It therefore goes ahead to conclude the tasks onto which concurrent think-aloud protocol would be more effective and the tasks that would be executed more effectively through other usability testing methods other than concurrent think-aloud protocol.

Key words:

Usability, usability testing, think-aloud protocol, concurrent think-aloud protocol, participant, eye-tracking, information scent, calibration, low-scent, high-scent, task.
Acknowledgements

This study was carried out in the Human-Centered Systems division in the Department of Computer and Information Science at Linköping University from February 2011 to August 2011.

I would like to extend my sincere gratitude to my supervisor, Dr. Johan Åberg, Linköping University, for the fruitful discussions we had during this study. His comments, directions and support have been extremely useful and resourceful over the 20-week period that I carried out this study.

I would also like to acknowledge the IKT-Studio team and their coordinator, Linnea Björk Timm, for lending me the eye-tracking machine for my practical work and for their guidance on how to operate it.

My sincere gratitude further goes to my friends: Atta Kuranchie, Zamda Mandari, Sandhya Mudduluru and Kwasi Frimpong and to all my classmates in MSc. Computer Science for their encouragement throughout my study period.

Finally to my family back home in Kenya, I may not be able to thank you enough for your input, both directly and indirectly in this study but your efforts are really appreciated. To other unnamed sources, thanks for the support and encouragement all through the period of this study.
# Table of Contents

1. INTRODUCTION .......................................................................................................................... 1  
   1.1 Background of Study .............................................................................................................. 1  
   1.2 Aim of the Study ..................................................................................................................... 1  
   1.3 Disposition ............................................................................................................................. 3  

2. LITERATURE REVIEW ................................................................................................................. 4  
   2.1 Usability .................................................................................................................................. 4  
   2.2 What is Usability Testing? ....................................................................................................... 4  
      2.2.1 Current issues with usability testing ................................................................................ 6  
      2.2.2 Good characteristics of usability recommendations ....................................................... 7  
   2.3 Existing Methods of Usability Evaluation .............................................................................. 8  
      2.3.1 Inspection based methods (Expert based methods) .......................................................... 8  
      2.3.2 Model based methods ..................................................................................................... 9  
      2.3.3 Usability testing (User based methods) ......................................................................... 9  
   2.4 Think-Aloud Protocol Method .............................................................................................. 12  
      2.4.1 Types of think-aloud protocol method .......................................................................... 12  
      2.4.2 Variety of think-aloud protocols mostly used ............................................................... 13  
      2.4.3 Concurrent think-aloud protocol and surveying behavior .......................................... 13  
      2.4.4 Some of the studies/researches that have been done so far on think-aloud protocol as a method of usability testing ......................................................................................... 14  
   2.5 Eye-Tracker as a Tool in this Study (Eye-Tracking as a Methodology in this Study) .......... 16  
   2.6 Information Scent as a Concept in this Study .................................................................... 18  

3. METHOD ...................................................................................................................................... 21  
   3.1 Type of Study .......................................................................................................................... 21  
   3.2 Participants ............................................................................................................................. 21  
      3.2.1 Participants for the pre-study (Participants during the selection of tasks based on the information scent concept) ................................................................. 21  
      3.2.2 Participants for the Pilot test procedure ....................................................................... 22
List of Figures

Figure 1: The graphical representation of the ten tasks and the average number of wrong clicks observed on the main webpage ................................................................. 26
Figure 2: Average amount of screen areas viewed in task 1 ........................................... 36
Figure 3: Average number of fixations in task 1 .............................................................. 36
Figure 4: Average total task time in task 1 ..................................................................... 37
Figure 5: Average amount of screen areas viewed in task 2 .......................................... 38
Figure 6: Average number of fixations in task 2 ............................................................ 38
Figure 7: Average total task time in task 2 ..................................................................... 39
Figure 8: Average amount of screen areas viewed in task 3 .......................................... 40
Figure 9: Average number of fixations in task 3 ............................................................ 40
Figure 10: Average total task time in task 3 ................................................................. 41
Figure 11: Average amount of screen areas viewed in task 4 ....................................... 42
Figure 12: Average number of fixations in task 4 .......................................................... 43
Figure 13: Average total task time in task 4 ................................................................. 43
Figure 14: Average amount of screen areas viewed in task 5 ....................................... 44
Figure 15: Average number of fixations in task 5 .......................................................... 45
Figure 16: Average total task time in task 5 ................................................................. 45
Figure 17: Average amount of screen areas viewed in task 6 ....................................... 46
Figure 18: Average number of fixations in task 6 .......................................................... 47
Figure 19: Average total task time in task 6

Figure 20: Task time averages (low information scent tasks)

Figure 21: Task time averages on main page (low information scent tasks)

Figure 22: Average number of fixations on main page (low information scent tasks)

Figure 23: Average number of areas with fixation on main page (low information scent tasks)

Figure 24: Task time averages (high information scent tasks)

Figure 25: Task time averages on main page (high information scent tasks)

Figure 26: Average number of fixations on main page (high information scent tasks)

Figure 27: Average number of areas with fixations on main page (high information scent tasks)

Figure 28: Amount of screen areas viewed on task one (with concurrent think-aloud protocol)

Figure 29: Number fixations on task 1 (with concurrent think-aloud protocol)

Figure 30: Amount of screen areas viewed on task 1 (without think-aloud protocol)

Figure 31: Number of fixations on task 1 (without think-aloud protocol)

Figure 32: Amount of screen areas viewed on task 2 (with concurrent think-aloud protocol)

Figure 33: Number of fixations on task 2 (with concurrent think-aloud protocol)

Figure 34: Amount of screen areas viewed on task 2 (without think-aloud protocol)

Figure 35: Number of fixations on task 2 (without think-aloud protocol)

Figure 36: Amount of screen areas viewed on task 3 (with concurrent think-aloud protocol)

Figure 37: Number of fixations on task 3 (with concurrent think-aloud protocol)

Figure 38: Amount of screen areas viewed on task three (without think-aloud protocol)

Figure 39: Number of fixations on task 3 (without think-aloud protocol)

Figure 40: Amount of screen areas viewed on task 4 (with concurrent think-aloud protocol)

Figure 41: Number of fixations on task four (with concurrent think-aloud protocol)

Figure 42: Amount of screen areas viewed on task four (without think-aloud protocol)

Figure 43: Number of fixations on task four (without think-aloud protocol)
Figure 44: Amount of screen areas viewed on task five (with concurrent think-aloud protocol) ................................................................. 81
Figure 45: Number of fixations on task five (with concurrent think-aloud protocol) ........... 82
Figure 46: Amount of screen areas viewed on task five (without think-aloud protocol) .... 83
Figure 47: Number of fixations on task five (without think-aloud protocol) .................. 84
Figure 48: Amount of screen areas viewed on task six (with concurrent think-aloud protocol) .............................................................................. 85
Figure 49: Number of fixations on task six (with concurrent think-aloud protocol) ......... 86
Figure 50: Amount of screen areas viewed on task six (without think-aloud protocol) .... 86
Figure 51: Number of fixations on task six (without think-aloud protocol) ................. 87

**List of Tables**

Table 1: Tasks selected and the average number of wrong clicks on their main webpage ...... 25
Table 2: Randomization of tasks .................................................................................. 27
Table 3: Shapiro-Wilk normality test results for each task .............................................. 49
1. INTRODUCTION

1.1 Background of Study

Think aloud protocol has been widely adopted as a usability testing method lately in the usability testing industry. Much research has been done and some are still ongoing regarding this interesting method of usability testing that requires the participants/users to perform the given tasks and still give verbal feedback concerning the task performance.

Some of the researches done have proved that think-aloud protocol method may change the way a user interacts with the system. Studies on how think-aloud protocol method affects the user’s behavior have also been on the rise. The two types of think-aloud protocol method mainly known as concurrent think-aloud protocol method and retrospective think-aloud protocol method have been targets of research as well with many concluding that the two methods resulted in discovery of similar number and types of usability problems. Also participants who worked silently and verbalized in retrospect performed more successful than their counterparts who were in the concurrent think-aloud condition. This is due to reactivity in concurrent think-aloud which affects the overall success rate of the tasks.

Eye-tracking has proved to be an effective methodology in most of the researches that aim at exposing the limitations of the think-aloud protocol method. It helps to collect the users’ eye-movement data on the screen during the testing session. This helps to analyze how the users’ eyes move across the screen during think-aloud protocol testing and thus assists the evaluator in knowing how efficiently the tasks are performed.

A recent study just completed, involved “A comparison between using think-aloud protocol method and not using think-aloud protocol method with an eye-tracker”. The study used very simple tasks. The results however showed no difference in the number of fixations and no difference in the amount of screen areas viewed. An interest to follow-up on this recently done study, led to the main aim of this thesis.

1.2 Aim of the Study

This thesis aims to answer the following primary research question:

- What effects do concurrent think-aloud protocol method, have on tasks with low information scent and tasks with high information scent in usability testing? The effects will be looked at in terms of the task time, the number of fixations and the amount of screen areas viewed as recorded using the eye-tracking methodology.

Surveying behavior is commonly observed during think-aloud protocol method of usability testing. The participants who are asked to use the think-aloud protocol starts by surveying or mapping the webpage. This means that the participant briefly looks at the whole web page before starting to perform specific tasks.
The participant’s attention is drawn to more areas and parts of the webpage that would not have otherwise been noticed.

(Pernice & Nielsen, 2009) theory explains why this happens and they state that the testing participant feels the need to really gather all the available information at hand and explain it to the test facilitator. This behavior as observed by (Pernice & Nielsen, 2009) mainly occurs when the test participant is not well acquainted to the webpage. After sometime, the test participant becomes more focused on their specific tasks and this behavior reduces.

Many users think that the test facilitator is more interested to hear comments on the aesthetics, design and general outlook of the webpage. While these can be interesting comments, they do not give much needed, pure usability feedback as compared to the feedback given when the test participant directly focuses on a given, specific task and comments on the problems or ease he/she is having while interacting with the interface to perform the task. This surveying behavior phenomenon is mostly evident when using think-aloud protocol method of usability testing (Pernice & Nielsen, 2009).

The fact that users are bound to survey the web pages during the concurrent think-aloud protocol usability testing method is a trivial point in usability research because it impacts negatively on the user’s behavior and in return affects the task success, task time task completion and number of fixations recorded during the task performance. The surveying behavior was therefore looked into during this study’s practical session to ensure more insight on when it is more evident.

As a secondary research question that will reinforce the just stated primary research question, this study will answer the following secondary question:

- While using the concurrent think-aloud protocol, in what conditions does the surveying behavior occur? Is it when performing tasks with low scent or when performing tasks with high scent?

The concept of information scent will be used as a basis for selecting websites and tasks. The selection of tasks would be a pre-study involving many tasks and then selecting the tasks with the lowest and highest information scents respectively.

The study/thesis will serve as a follow-up and extension of the recently done study stated above and will also take into consideration the surveying behavior of the users/participants involved.
1.3 Disposition

Chapter 1: Introduction

This part introduces the title in general and the area of study that this thesis focuses on. It leads the reader to the primary and secondary aims of this study.

Chapter 2: Theoretical background

This part shows the literature review of what various articles, electronic media and books highlight concerning usability evaluation, usability testing and think-aloud protocol as a method of usability testing.

Chapter 3: Method

This part shows the type of experiment that was carried out in this study and the procedures. It also shows the tools, methodologies and concepts used during this study so as to get the necessary results. The means of data collection used in this study are also covered in this part.

Chapter 4: Results presentation

This part shows the raw data got from the experiment carried out in this study.

Chapter 5: Discussion

This part analyses the results presented on the results presentation part. It explains what can be deduced from the presented data.

Chapter 6: Conclusion

This part explains what can be concluded from the analyzed data. It also shows what the analyzed results suggest in relation to some of the previous studies done in the same area.
2. LITERATURE REVIEW

2.1 Usability

Usability generally refers to the quality of being able to provide good service. It can also refer to making a product easier to use by matching its features with the user’s needs and requirements. (Norman D. A., 1999) however clearly states that *usability is much more important on websites that with physical products*. To determine the usability level of a given website, usability testing has to be carried out on the website.

2.2 What is Usability Testing?

Usability testing refers to the technique used to evaluate and determine the ease of using a given product. This is done by having users who represent a group targeted by the system also known as representative users, use the system and while on it, the usability tester observes and listens to the users’ complains or compliments as he/she takes notes. The usability tester can also ask/interview the user on the general feeling while using the product.

Once the usability tests are carried out, total quantitative data is analyzed to determine the usability of the product. Some of the quantitative data that are usually considered while measuring the usability of a product are time taken by the representative user to use the product and also the rates at which the user gets errors while using the given product. Usability tests not only measure the usability of the product but also determine the user’s satisfaction with the given product and helps to determine adjustments required on the product so as to improve user performance.

Usability testing should be done early and quite often during the product development stages so as to provide vital feedback to the developers and designers of the product while most recommended changes can still be implemented, that is before the product design and make-up becomes complicated or too concrete to change. This makes it easy to adjust the product to user’s requirements as the product’s flexibility is still high and it also makes it cheaper to make the needed adjustments. Testing often implies testing the product after each completed step.

According to (Experience Solutions Ltd., 2010), various types of usability testing exist with reference to why the usability testing needs to be carried out at any given moment. These include:

*Comparative Usability Testing*

This is usually carried out when the tester wants to compare a given product (in this case a website) to other websites. It is done to establish which website is easier to use than the other and which one has a better design.

This is usually done to help the tester to find out detailed functional needs of the user for a system. Sample websites might be used and as the representative users use the website, the tester records the user’s complains and compliments which in return act as the new requirements for an upcoming website. It helps the tester to establish the user’s needs and concerns to be included in a new product. It serves as an alternative way for analysts to gather user’s functional requirements of a website/product.

**Usability Evaluation (Evaluation Usability Testing)**

This is normally done on a new website or a website that has been upgraded and more functional features have been added. The testers usually test the updated features or new website/product with the users to “measure the user experience” and to find out if the product/website suits what the users want and if it is ready for the market. In case usability faults are found, the feedback is usually given to the designers and programmers on time for them to adjust the website to make it simpler and easier to use before launching it to the market for all the possible users.

The concept of Usability evaluation can be looked at in various ways (Dillon, 2001). Some of the major angles of perception given to this are:

**Semantics**

This is where the usability is looked at in terms of ‘ease of use’ or ‘user friendliness’ of the website. The various constructs of the website are not taken into consideration. The general ease of use to the users is given more concentration.

**Features**

When usability evaluation of a website/product is looked at from the features perspective, the various features that make up the website is what is looked into. The evaluation is mostly based on which features/constructs the users feel are present and which ones are absent. This also considers which features need change/adjustment to make it easy for the users to effectively identify and use the features. The outlook of the features and how this attracts the users is also taken into consideration. Some of the features that are usually given attention are: Windows, menus, Icons and pointing devices. The graphical design of the website is evaluated in this case.

**Operations**

The performance of the website/product and the effective levels with which users from a given/specific user group carry out a certain task is what is considered when usability evaluation is looked at in this concept. The rate/speed at which the website can be operated by a given groups of users display how suitable the website/product is for the market.

In general, usability should not only be based on the desirable interface attributes/features but should also be looked at in general as a measure of Human computer interaction.
2.2.1 Current issues with usability testing

Some of the major challenges that Usability testers have been known to face so far are:

Cost

Usability testing is relatively expensive. This is due to the facilities, staff, time, and equipments that are at times needed for the testing exercise to be carried out successfully. For instance; testing a complete or more revised web application is known to require more representative users to test the application than a new and just started application. The more the users, the more the resources needed and thus the more the expense (eVALUEd, 2006).

Sample of potential Users

More often than not, usability testing is known to be possible and easier with a smaller sample of potential representative users. The more the users involved during the testing, the higher the associated costs, time and the more complicated the tests. A larger sample of users is however known to provide more concrete and reliable results since it increases the chances of identifying a wide range of problems during usability testing of an application. It is natural that involvement of a larger sample of users as an effective tool during the usability testing is met by additional overhead costs and resources (eVALUEd, 2006).

Complexity in data analysis

Depending on how complicated the specific tasks to be tested are, the type of representative users, the tools and the methods used during the usability testing process, data collection and analysis can pose to be a challenge to the usability tester. Some users for instance; are not good at expressing themselves. Analyzing emotional or even verbal response from such users can therefore be a bit tricky, as one might be forced to ‘read too much’ into the user’s reactions. Also based on the other given factors above and depending on the scale in which the testing is done, data can be complex and at times time consuming to analyze. The analysis of the data should be allocated enough time so as to avoid inaccurate results (eVALUEd, 2006).

Commitment by participants

Both the testers and the representative users should be committed on their part to make the whole usability testing process successful. Cases where either of the participants became reluctant on their part have resulted to slow usability testing process and at times inaccurate results due to lack of devotion to the given task. Slowing down the usability testing process might have other adverse effects on the other development stages of the application because the results presented after the usability testing acts as an important feedback to the designers and programmers of the application and the feedback should be given on time for timely adjustments to the application or the design before the application/product gets to a more complicated stage (eVALUEd, 2006).
**Representation of the real scenario**

Usability testing results is not always one hundred percent (100%) representation of the real scenario. The results represent the views of the representative users involved during the usability testing and their views are assumed to represent the views of many users in the world who would use the application at some point. Even though a wide range of users representing all user groups is always carefully chosen during the testing, the assumption might not hold. The user might be biased during the testing or might give misleading results based on other external factors like personality, moods and so on. The final results got might therefore vary from a real world scenario of users’ perceptions on the application. The smaller the variation percentage is, the better and more accurate the results. A wider variation would indicate less accurate results and more the indication that most users in the real world might have a different perception on the application (Experience Solutions Ltd., 2010).

**2.2.2 Good characteristics of usability recommendations**

Once the usability evaluation of an application is done, the results (user’s feedback) are presented then analyzed so as to draw some useful conclusions from them. The analyzed results then act as the basis of the recommendations that are given to the application designers and programmers for them to make the necessary adjustments to improve the application’s usability. Usability recommendations should therefore be done in the most effective way possible since they play a big role in the overall improvement of the application’s usability state. Some of the characteristics that good usability recommendations should have are:

*Effective communication*

Each recommendation that is given to the designers and programmers of an application should be done clearly. This avoids any ambiguity and makes the programmers and designers be sure of what the exact problem is.

*Recommendations should target improvement*

The evaluators or the parties who give usability recommendations based on evaluator’s findings should structure the recommendations such that they not only help to improve usability of a particular task but help to improve the general usability of the whole application as well (Molich, Jeffries, & Dumas, 2007).

*Use of examples*

Since most evaluators have direct experience with the users during the usability testing phase, they should quote some examples while giving recommendations so as to make the recommendations can have more weight. Examples can be in form of user’s direct speech to show the user’s reaction. This helps to avoid vagueness in the recommendations and deepen the understanding of the designers and programmers into what the users really want.
Use of images and drawings

This mostly applies to usability recommendations on the application’s graphical design. The evaluator can make the recommendations more specific and clearer by providing images and shapes of some of the graphical structures that the users might be opposed to. For example the evaluator can draw the recommended shapes of icons that the users are more comfortable with. This accompanied by proper description, makes it much clearer to the designers.

Respect the Business and technical constraints

Every development team operates within specific constraints. Usability recommendations made to the development team (designer and programmers) should not only incline to what the users want but should also try and integrate these recommendations within the business and technical constraints of the development organization. This ensures that the given usability recommendations are not beyond the available resources (Molich, Jeffries, & Dumas, 2007).

2.3 Existing Methods of Usability Evaluation

Usability evaluation methods can be grouped into three distinct categories:

2.3.1 Inspection based methods (Expert based methods)

Expert based method is whereby a Human computer interaction (HCI) or usability expert assesses an application and gives feedback on its usability level. The HCI professional examines the website/application and estimates its usability for a certain user group. There are no users involved in this method of evaluation and the given results are entirely based on the HCI expert’s judgment and interpretation since the HCI expert is the evaluator. The method is known to be cheaper and to produce evaluation results faster than the user-based methods. The evaluators of the application are provided with a pre-determined structured method that they use to examine and report any noted problems with the interface. It is the evaluator’s role to make a guess on how the users would react to certain interface attributes and certain task procedures (Dillon, 2001).

There are two expert based usability evaluation methods. These are:

Heuristic evaluation

In heuristic evaluation, the HCI expert is provided with a list containing design guidelines which he/she uses to examine and evaluate every screen of the interface in a sequential order following a given path for a specific task. If the evaluator comes across any violations of these guidelines, he/she reports them as the potential user problems. The evaluator (HCI expert) assumes that any guideline not met by the interface is a likely problem that the user might face while using the interface. Heuristic evaluation focuses more on how well an interface conforms to the given design guidelines (Dillon, 2001).
Cognitive Walkthrough

In Cognitive walkthrough method, the HCI expert determines the sequence expected in a correct task performance. He/she then estimates on a screen by screen basis the likelihood of a user to perform the determined sequence correctly or the likelihood of failure performing the sequence. This estimation is assumed to be directly related to the real scenario whereby in instances where the likelihood of failure in performing a task according to a given sequence is detected, it is also assumed that the real users will fail when performing the task. Cognitive evaluation therefore focuses more on the user experience when operating a given application taking note on the difficulties the user might face while learning to perform given tasks on the application (Dillon, 2001).

2.3.2 Model based methods

In Model based methods, the HCI expert uses formal methods to predict user performance when carrying out a given task in an application. Just like in expert based methods, no users are involved during the usability evaluation.

Some models have been invented that can predict certain user performance aspects like time taken to complete a given task and the ease in learning a new task. An evaluator pre-determines an exact sequence of events that a user will have to carry out to perform a task. An analytical model is then applied to this sequence and the index of usability is calculated. The models work effectively in predicting time taken for task completion in error-free tasks and tasks that need no decision making (Dillon, 2001).

2.3.3 Usability testing (User based methods)

In user based methods, a sample of users try to use the application. During the usability testing, these users perform a set of pre-defined tasks on the application. These methods give more valid and reliable estimate of an application’s usability because users are actively involved during the testing unlike in expert based and model based methods.

The main aim of user based methods is to check the extent in which an application supports the target users in their work. It also checks how easily, effectively and satisfactorily the users perform pre-set tasks in given environments. Users are usually asked to perform a set of tasks in a given application and might employ the use of a given technology to perform the task. Based on the evaluator’s focus, success rate at which the users complete the tasks and their speed of performance during the tasks are recorded. Upon completion, users are asked to give their views (likes and dislikes) and also performance views on the application. Measures of efficiency, effectiveness and satisfaction of the application to the intended users can then be derived by analyzing the results (users’ feedbacks and the evaluator’s recorded information) using various usability metrics. From the results analysis, potential problems that the intended users might face can be identified and re-design approach can be determined.

User based tests are often constrained as a result of result limitations. Due to this, HCI experts are mostly interested in coming up with ways of how to gain the most information from the smallest sample of users. The sample size requirement however, is highly dependent on the
type of errors one seeks to identify and the probability of occurrence of such errors. A few users might identify problems in a new application but more users would be needed to identify a range of problems in a more revised or completed application/product (Dillon, 2001).

Existing User-based evaluation methods vary based on how the feedback from the users is collected. These methods include:

*Interviews and Videos*

The evaluators/testers interview the users by asking them questions about their experiences in using the application after the users use the application. The users are free to air their views about the system and through this, they point out their likes and dislikes in carrying out specific tasks on the application. A video may be taken with the users describing the performance and any perceptions they have on the application. Recorded videos also help in subsequent analysis of the navigations, transactions and problem handling that take place during the users’ interaction with the application.

*Unstructured user based tests*

This involves the user and the evaluator jointly interacting with the system to agree on what works, what does not work, what is good with the design and what might be problematic with the design. This user based method can be effective in exploring various interface options in the early stages of application design where it might be too early to employ the use of formal, quantitative assessments.

*Questionnaires*

This is a query technique that elicits the users’ perceptions about an application by having the representative users fill in questionnaires as they use the application or immediately after using the application. How to design the questionnaires is an important issue if questionnaires are to be used. The purpose of the usability testing should be clearly brought out in the questionnaires by designing the questions to fit the intended areas to be tested. The questions should also be designed in a way that they can provide measurable feedback. This makes the questionnaires more effective. The use of questionnaires guarantees less time to be spent in testing, wider user group can be targeted and the results of the questionnaires can be easily and effectively analyzed. The questionnaires must however be reliable and valid to ensure testing for efficiency and effectiveness of the application.

*Observation*

Unlike use of Questionnaires, observation method cannot be done remotely from the user. Observation method is mostly used to test for effectiveness of the system/application and user satisfaction. During the usability testing session, the evaluators observe users as they use the system to accomplish tasks and they try to get the kind of mental model the users have about the system. They observe the user’s attitude, reactions, emotions, facial expressions, verbal comments, sitting adjustments and so on to establish the user’s attitude towards the
system/application. The evaluators (observers) note down their observations and use them as the results to be analyzed. Observation method is used to obtain qualitative data and not quantitative data.

Think-aloud protocol

According to (Po-Yin Yen & Suzanne Bakken, 2009), *Think-aloud protocol was developed by Lewis in 1982 to understand cognitive process.* In the think-aloud protocol method, the evaluator observes while the user works with the interface and encourages the user to speak out his/her thoughts (think-aloud) as he/she navigates through the interface to carry out a specific task or general tasks on the interface. The user should think-aloud so as to voice out what he/she is thinking or wondering about the application/interface at each moment during the testing session. One major setback with this method is that many are times that the users cannot communicate as fast as they think and act due to divided attention. It is therefore not very easy for the evaluator to connect the user’s comments with his/her respective actions. This problem is usually solved when the user incurs a problem on the application which makes him/her slow down on their actions. This in return gives the evaluator time to take notes and to correlate what the user is saying (thinking aloud) and the action at that given moment (Norgaard & Hornbaek, 2006).

To make this method more effective, the user should comment liberally on his actions and thoughts without any bias and the evaluator should create an informal environment for the user so as to make the user comfortable without any tensions given the method itself is an informal method. The evaluators are known to collect the feedback by taking notes when the user comments on specific tasks or on key tasks and also by audio/video recording as the user thinks aloud. The video evidence is known to give a more quantitative data.

The think-aloud protocol method is known to give faster feedback as compared to questionnaires. Other ways the evaluator mostly uses to collect data when using this method in usability testing are: through observation of what the user does and hearing of what the user says; he/she must however correlate the two and know which comment refers to what observed action (Norgaard & Hornbaek, 2006).

Think-aloud protocols therefore help to shed light on the user’s thoughts while he/she is interacting with the application interface. This helps to address issues of user cognition and comprehension.

A noted disadvantage of this method is that *participants can be resistant to verbalizing the problem and that it sometimes can be difficult to identify changes in behavior due to learning* (Holzinger, 2005). The users (participants) at times also lose their confidence and track of information and link location due to divided attention between using the system and reporting the results. Some of these major setbacks of this method therefore have to be taken into consideration if the method is to be used effectively (Norgaard & Hornbaek, 2006).
2.4 Think-Aloud Protocol Method

2.4.1 Types of think-aloud protocol method

There are two types of the Think-aloud protocol method. These are:

*Concurrent Think-Aloud Protocol*

In concurrent think-aloud protocol; more problems are detected through observation. This type of think-aloud protocol enables the evaluator to see into how the user (participant) is thinking. Apart from just the problems that the evaluator observes as the user thinks aloud, the evaluator also gets to capture a complete overview of the problems encountered by the user for instance how the user is distracted from optimal performance due to handling two roles at ago, these being performing the task at hand and giving the verbal feedback at the same time.

A well known advantage that is usually associated with the concurrent think-aloud protocol is that the verbal feedback that the user gives as he/she carries out the task may exhibit any surprises, irritation, doubts, satisfaction or other feelings that may arise during the task performance process. This helps the evaluator to get a more genuine feedback on how the user feels while using the application. The user is not able to hide or fabricate his/her feelings concerning how easy/difficult it is to use the application. The verbal output given by the user during the task performance is more in a reactive manner.

Often considered as some sort of disadvantage of the concurrent think-aloud protocol is the fact that handling two things at the same time, in this case performing the task and verbalizing one’s thoughts concerning the tasks may result to reactivity especially in cases of high-task complexity. Reactivity implies that the users work differently from normal and this contributes to deviation from optimal working level which generally affects task performance. (Maaike & Menno, 2003) however argue that the risk of reactivity can be avoided by imposing strict guidelines (Maaike & Menno, 2003).

*Retrospective Think-Aloud Protocol*

In retrospective think-aloud protocol; more problems are detected through and during verbalization. During this type of think-aloud process, users perform the tasks in silence then after that they give verbal feedback of their thoughts in retrospect. In certain cases, especially when the duration between the task performance and verbalization of thoughts is long, the retrospective verbalization takes place without any stimuli or incitement. This leads to production or giving of exhausted comments. To avoid this, usability professionals (evaluators) support the retrospective verbalizations by recording of the task performance. (Nielsen, 1994) for instance recommended the use of video recording to record the performance.
In cases where the verbalization of the participant’s thoughts is evoked by stimuli then the retrospective think-aloud method will be a combination of the benefits of both working silently and thinking aloud. This is due to the fact that the stimuli will be an indication of fresh thoughts of the task performance still in the participant’s mind. Nevertheless, it remains a challenge in the usability industry for a participant to remember everything he/she thought of during the task performance. Some users/participants might come up with fabricated thoughts just to please the evaluator or just for the sake of giving some verbal feedback. One main advantage of the Retrospective think-aloud protocol is that its validity is not affected by task complexity (Maaike & Menno, 2003).

2.4.2 Variety of think-aloud protocols mostly used

There exists a variety of think-aloud protocols that are usually used quite often by usability professionals (Olmsted-Hawala, Murphy, Hawala, & Ashenfelter, 2010). The protocol refers to the way an evaluator handles the participants during the think-aloud method of usability evaluation. It is some form of guidance on what sort of feedback the evaluator expects from the participant. This in return determines the results got during the given think-aloud testing. The most commonly applied protocols are:

- Instruction: This is whereby the evaluator tries to give some sort of indirect instruction so as to provoke some feedback from the user/participant. For example; “Tell me why you clicked on the Tab”, “Tell me why you scrolled on the menu”. “Tell me if you are looking for something”; “Tell me what you are looking for and whether you can find it”

- Intervention: This is whereby the evaluator intervenes as the participant verbalizes his/her thoughts. This helps to probe or prompt the user/participant to give more expounded verbalized thoughts. For example; “Is that what you expected to happen?” “Keep talking”, “What do you think of the shape of the icon?” “What are you thinking about now?” “What do you think this icon means?”

- Prompting: This is whereby the evaluator induces prompts to the user at intervals of given period of time for instance after every 10 minutes of prolonged silence. This re-energizes the user to give any existing thoughts in his mind about the task performance.

2.4.3 Concurrent think-aloud protocol and surveying behavior

Surveying behavior is one major factor that cannot be ignored during concurrent think-aloud protocol usability testing method. According to (Pernice & Nielsen, 2009), surveying behavior is whereby the test participants survey and appraise the whole webpage before they
really try to do the actual task work. This eventually has an impact on the task success rate, task completion rate and the task time of a given task that the participant/user is supposed to perform. Surveying the webpage before performing the given, specific task also leads to more fixations being recorded by the eye-tracker than would have been the case if the participant could have directly performed the given task.

According to (Pernice & Nielsen, 2009) article titled “Eye-tracking methodology”, one of the major reasons given for surveying behavior is that during the concurrent think-aloud protocol usability testing session, most users will look around the webpage to make themselves familiar and better-versed on the subject since many people would not be comfortable to talk about a subject they are not well acclimatized to. They therefore feel that they would give the facilitators better feedback if they generally know the main items on the webpage. Some users also use the surveying behavior as a means of warming up to the webpage before they focus on performing the specific tasks they are asked to perform. Users generally forget that it is more important to talk about their actions concerning the given tasks being tested than criticizing or giving feedback of the whole page in general.

The surveying behavior is usually evident by the average number of dispersed fixations recorded in the first few seconds and that are located mostly on the periphery of the webpage and not specifically on a given task.

2.4.4 Some of the studies/researches that have been done so far on think-aloud protocol as a method of usability testing

Study 1

A study carried out by (Maaike & Menno, 2003) aimed at comparing the two variants of think-aloud protocol namely concurrent and retrospective think-aloud protocols. The study was designed to empirically investigate the value of both variants by highlighting their benefits and drawbacks (Maaike & Menno, 2003).

The conclusion of the study was that the verbalization of thoughts was more substantial in the retrospective think-aloud method though the two methods resulted in similar number and types of problems. Also participants who worked silently and verbalized in retrospect performed more successful than their counterparts who were in the concurrent think-aloud condition. This is due to reactivity in concurrent think-aloud which affects the overall success rate of the tasks (Maaike & Menno, 2003).
Study 2

The main objective of the study carried out by (Guan, Lee, Cuddihy, & Ramey, 2006) was to investigate “the validity of the stimulated retrospective think-aloud method as measured by eye tracking”. This was done by comparing the verbalization of participant’s thoughts with their eye movements as indicated by the eye tracker. The eye tracker helped to indicate the specific actions that the participants carried out in order to complete given tasks (Guan, Lee, Cuddihy, & Ramey, 2006).

The study concluded that stimulated retrospective think-aloud method had a low risk of the participants giving fabricated/invented verbalized thoughts. It also states that the validity of the stimulated retrospective think-aloud method is not affected by task complexity since the participants have full concentration when performing tasks and verbalize their thoughts later. The study further indicates that the retrospective think-aloud method also provides info on the user’s way of reasoning and strategies while carrying out tasks. This study therefore supported retrospective think-aloud as a reliable and valid method to be applied by usability professionals (Guan, Lee, Cuddihy, & Ramey, 2006).

Study 3

Another study was done by (Hertzum, Hansen, & Andersen, 2009) to investigate whether participants that think-aloud in a classic or relaxed way behave differently compared to performing in silence. The study titled “Scrutinizing usability evaluation: Does thinking aloud affect behavior and mental workload?” concluded that apart from prolonging tasks, classic think-aloud has little or at times no effect on the participant’s behavior. However, behavior is affected in several ways when using relaxed think-aloud (Hertzum, Hansen, & Andersen, 2009).

The study found out that in a relaxed think-aloud setting, participants spent a larger part of task performing time on general distributed visual behavior, issued more commands in navigating within and between the websites’ pages, took a longer time to perform tasks and generally experienced higher mental workload. These had some negative impact on the usability evaluation (Hertzum, Hansen, & Andersen, 2009).

Study 4

A recent study just completed, involved “A comparison between using think-aloud protocol method and not using think-aloud protocol method with an eye-tracker”. The study which used very simple tasks, showed no difference in the number of fixations and no difference in the amount of screen areas viewed.
Study 5

An article done by (Pernice & Nielsen, 2009) titled “eye-tracking methodology”, points on the Surveying behavior of the participants during usability testing session. The surveying behavior is mostly observed when participants use think-aloud protocol method of usability testing, during which the participants survey and appraise the whole webpage before they try to perform the actual task on the webpage. An advantage associated with the surveying behavior is that it helps the participants/users to identify features that they might not have easily identified if they were to be directly involved in doing the specific tasks.

The article however concludes that the surveying behavior affects the given task success rate, the task completion and task time during the usability testing since the users scan much of the webpage while paying more attention to the main details than to the specific given tasks. More fixations and screen areas looked at are recorded by the eye-tracker when the user surveys the whole web page before performing a specific task compared to the number of fixations recorded when the user directly performs the given tasks.

This study

As a follow-up and extension of Study 4 in combination with Study 5, this study aims at finding out the “effects that concurrent think-aloud protocol method has on tasks with low information scent and tasks with high information scent in usability testing. The effects will be looked at in terms of the task time on the main page, the number of fixations on the main page and the amount of screen areas viewed on the main page as recorded using the eye-tracking methodology”. The study uses the concept of information scent as a basis for selecting websites and tasks therefore choosing a selection of tasks with low information scent and also tasks with high information scent. The high scented and low scented websites/tasks are objects on which usability testing using concurrent think-aloud protocol in one group and not using concurrent think-aloud protocol in another will be carried out. The differences will then be analyzed and compared.

The surveying behavior will also be taken into consideration and whether the surveying behavior occurs during the usability testing of the high information scent tasks or during the testing of the low information scent tasks will be noted and considered. The procedure of carrying out this study is given in the method section.

2.5 Eye-Tracker as a Tool in this Study (Eye-Tracking as a Methodology in this Study)

A Standard English dictionary describes Eye-tracking as the process of measuring either a point of gaze (where one is looking) or the motion of an eye relative to the head. An Eye-tracker therefore refers to the device that measures eye positions and eye movements. According to (Nielsen & Pernice, Eyetracking web usability, 2010), Eye-tracking is simply following the trail of where a person is looking.
The eye-tracking technology makes it easier for the evaluator to observe the path that the participant looks on the computer screen. The eye-tracking camera/equipment being a physical device, can be built into the computer monitor while the eye-tracking software keeps track of what is being displayed on the screen as the participant looks at the screen (Nielsen & Pernice, Eyetracking web usability, 2010).

The operations of the eye-tracking technology are based on a few concepts outlined as follows (Nielsen & Pernice, Eyetracking web usability, 2010):

- Whatever falls in the peripheral rather than the foveal vision (a central point with high resolution) is blurred.
- *Fixation* is a term used in the eye-tracking environment to refer to when the eye is resting on something specific.
- *Saccades* are rapid eye movements from one fixation to the next.
- The *mind-eye hypothesis* holds that people tend to think of what they are looking at. Though this hypothesis is not 100% true, it holds enough for eye-tracking to indicate what the participants/users pay attention to on the web pages. The hypothesis is used to determine what the users might be thinking of. It indicates that fixation equals attention.
- For a certain element to attract fixations or be ignored is not an obvious indication of users’ good or bad thoughts concerning the element.
- Users tend to be more attracted to those websites that allow them to easily focus on what they want and also to those that allow them to easily ignore what they do not need.
- Eye-tracking results are usually visualized in three main ways: By watching slow-motion gaze replay videos. This is known to be quite time consuming though. The other ways is by heat maps and gaze plots which are usually used to represent movements in time as users navigate through the website and as their eyes move rapidly across the web page.
- Heat maps are used to show the combined fixations of many users on a webpage. The red areas represent the places that users look at the most, the yellow areas indicate fewer fixations compared to the red and the gray areas are the areas that did not attract any fixations at all thus are the areas that the user did not look at.
- A gaze plot on the other hand, indicates the experience of the visit of a single user on a web page. Each blue dot on a gaze plot represents a fixation. The size of the blue dot is directly proportional to the duration of the fixation thus the larger the dot, the longer the fixation. The thin lines between the blue dots represent the saccades that are recorded as the eyes moved from one location to another. The numbers on the blue dots represent the sequences of the fixations; they show the order in which the fixations occurred.
- Experiences that users have on the other pages of the website and the tasks that the users try to perform on the web site greatly influence how the users look at the web page.
According to (Pernice & Nielsen, 2009), certain benefits associated with the eye-tracking methodology makes it appropriate for use in most usability testing studies especially in studies that aim to point out the limitations of the think-aloud protocol method. It is therefore a reliable tool of study in most researches carried out in the think-aloud protocol method. Some of its major benefits include (Pernice & Nielsen, 2009):

- Eye-tracking enables the usability professionals to have in-depth understanding of the users’ experiences on a webpage.
- Eye-tracking helps test facilitators to avoid common mistakes; for instance interrupting a participant when he/she is quiet for a long duration.
- Eye-tracking makes the usability practitioners identify certain user behavior that are usually not easily identified in a normal traditional usability testing set-up. These behaviors include: Exhaustive review referring to when users look repeatedly at areas that seem helpful but are not helpful, Selective disregard referring to when users intentionally ignore/tune-out areas of the website at given times and Miscues referring to the interface items that erroneously call for attention. At first, users tend to waste some time on miscues thinking they are of importance. This at times chews up users’ time that could have been spent on more important items.

Despite the benefits associated with the eye-tracking technology, some major drawbacks associated with it cannot be ignored. Eye-tracking for instance makes a normal usability test to be more time consuming, expensive, difficult and easy to get non-accurate/misleading results (Pernice & Nielsen, 2009).

A regular usability test is known to be simple and at times just needs paper and pen to get the exercise done and record down the noted observations. Eye-tracking however, helps the evaluator to pay close attention to details during the usability research. It is because of these benefits stated here that eye-tracking proved to be a very important tool in this study. In this study, Eye-tracking will assist in getting the in-depth details of the results that will be displayed on the screen using the eye-tracking software when usability tests are carried out on both the low information scented tasks and high information scented tasks with one group using think-aloud protocol and the other one not using think-aloud protocol method. The eye-tracking methodology will help to compare and analyze the results to determine which type of tasks are affected by think-aloud protocol method and how.

### 2.6 Information Scent as a Concept in this Study

A ‘catchy’ website is all that users want. This entails a combination: First, users will leave a website if the content of the website is good but difficult to find. Secondly, users will leave a website if the content is easy to find but offers very little or empty of the information needed (unnecessary content). A website’s content should therefore be good and easy to find thus the combination for a catchy website. According to (Nielsen, useit.com, Jakob Nielsen's Alertbox, Information Foraging, 2003), People like to get maximum benefits for minimum efforts. (Nielsen, useit.com, Jakob Nielsen's Alertbox, Information Foraging, 2003) clearly states that Information scent refers to the act of predicting a path’s success. This means assessing
whether a task’s given path exhibits cues related to the desired outcome. Information scent is a concept that is widely applied in information foraging (Nielsen, useit.com, Jakob Nielsen's Alertbox, Information Foraging, 2003). Information scent is the strength of local cues such as text labels in providing an indication of the utility or relevance of a navigational path leading to some distal information source (Pirolli, Card, & Wege, 2001).

From the sources (articles) read during this study, it can therefore be concluded that an information scent is a feature on a website element (tab, a link, a label etc.) that makes the user to easily predict that the element is leading them along the right path towards finding the desired destination. Use of words that relate to the outcome on a website element for instance a link or a menu, would serve as a good scent. Made-up words might make it hard for the user to find the sought after path.

High scent task

A high-scent task is whereby the path to the outcome (needed information) consists of labels with high information scent. The labels have a high indication that strongly relates to the destination information. This means that by seeing the labels, the users can easily predict the path.

Low scent task

A low-scent task is whereby the path to the outcome consists of labels with low information scent. The labels do not relate easily to the destination information and thus it becomes difficult for the user to know or to easily predict the correct path. Low scent tasks therefore cause users to try more paths on their way to get the desired outcome; thus resulting to a more costly visual search (Pirolli, Card, & Wege, 2001).

(Nielsen, useit.com, Jakob Nielsen's Alertbox, Information Foraging, 2003) suggests that for a good website, the information scent should keep getting stronger to keep signaling the user that he/she is near the outcome/destination. The progress must also seem rapid to be worth the predicted effort. One of the design lessons from information scent is to make links and category descriptions to explicitly describe what users are likely to find at the destination (Nielsen, useit.com, Jakob Nielsen's Alertbox, Information Foraging, 2003). Since users are usually faced with many navigation options; a path that easily hints the outcome would be more likely to be chosen by the users. This can be made possible by making the right trail to the outcome easy to identify so that the other paths can easily be rendered void or unnecessary for the given outcome by the user. The web pages also serve as a means to increase the information scent (path’s predictability) if each web page that a user advances/moves next to, towards the outcome displays that the user is on the right path and should even indicate the current path position the user is at, thus providing feedback.

In a previous research done by (Pirolli, Card, & Wege, 2001), closer inspection of the eye-tracking data suggested that while performing low scented tasks, participants ended up scanning the dense areas of the display with very short distances between eye fixations. While performing high scented tasks though, participants scanned with longer distances between
fixations. The high-scent task fixation movements also proved to be longer than the low scent movements by about 25%. The effective area of visual attention therefore changes with information scent and density (Pirolli, Card, & Wege, 2001).

Attentional spotlight being the display area surrounding an eye fixation, narrows with increasing display density when the information scent is low and broadens when the information scent is high (Pirolli, Card, & Wege, 2001).

In this study, Information scent will be used as a basis for selecting websites and tasks. A selection of tasks with low information scent and also tasks with high information scent will be identified. The study will be performed with two groups of users, one doing think-aloud and one not, then comparing the measures taken and see whether there are any differences.

The selection procedure of high information scent tasks and low information scent tasks for this study has been elaborated in the method section.
3. METHOD

3.1 Type of Study
This study was a quantitative and qualitative usability testing to evaluate the effects that concurrent think-aloud protocol method has on tasks with low information scent and tasks with high information scent in usability testing. The effects were looked at in terms of the task time, the number of fixations and the amount of screen areas viewed as recorded using the eye-tracking methodology”.

The type of concurrent think-aloud protocol that was used in this study was level 1 verbalization concurrent think-aloud protocol because it was easier for the participants/users to vocalize their thoughts that they already have in focus of attention than to first process their thoughts mentally, link to the previous thoughts then verbalize.

The experiment therefore used the normal observation usability test method and the concurrent think-aloud protocol described in this report, the eye-tracking methodology and the information scent concept. The independent variables in this study were: testing with think-aloud protocol and testing without think-aloud protocol. The dependent variables were: the task time, the number of fixations and the amount of screen areas viewed on the main page.

3.2 Participants
It is important to note that the twenty participants, who participated during the pre-study, were different from the twenty participants who participated during the main test procedure and the three participants who participated during the pilot study procedure. The three participants who participated during the pilot study procedure did not participate during the main test procedure.

3.2.1 Participants for the pre-study (Participants during the selection of tasks based on the information scent concept)
Twenty participants, chosen among the students of Linköping University, participated in the pre-study. They were all aged between twenty one and forty. The participants had to be willing to participate in the pre-study, frequent users of the internet, capable of understanding a given task and concentrating on performing the task on a given website. The participants at the pre-study were therefore good internet users with good vision and comfortable to perform the tasks in the presence of the facilitator. They were ten males and ten females who understood both English and Swedish. They therefore perfectly understood and carried out all the instructions given by the test facilitator.

Choosing participants that were representative users for the websites chosen in the pre-study was the main goal that was considered in choosing the participants for the pre-study. The websites and tasks chosen during the pre-study and the main study were therefore tasks that an average Linköping University student can relate to and can perform well.
3.2.2 Participants for the Pilot test procedure

The main aim of the pilot test procedure was to find out if the eye-tracker and all the associated parts (the monitor, the camera, the recording software and the analyzing software) that were to be used for the main test procedure, worked as needed. Three participants were used for the pilot test procedure. They were between ages twenty to forty (20–40), were good at using the internet and following instructions given by the test facilitator and had good vision. They were also willing to participate in the pilot test procedure.

3.2.3 Participants for the main test procedure (Main study)

The total number of test participants was twenty comprising of ten females and ten males. The participants were between ages 21 to 40. To participate in the study, participants needed to be willing to participate in the main test procedure, have good basic internet skills and especially it was a must for the participants to be real target users of the chosen websites/tasks. Each participant therefore needed to be a normal website user who could perform the needed tasks on a website and be in a good position to judge the quality of the task performance and give appropriate verbal feedback. The frequency of internet use was therefore a key factor to be considered so as to ensure that the participant is familiar with the internet. It was also a must for each participant to have a good vision. The twenty participants were recruited among the students of Linköping University.

There were two groups of participants each consisting of ten people. One group performed the concurrent think-aloud protocol usability testing on the given tasks and the other group performed normal usability testing on the given tasks without thinking aloud. Each participant of each group performed their respective usability testing (either concurrent think-aloud or normal usability testing) on the tasks with high information scent and on the tasks with low information scent, meaning each participant performed two types of tasks (tasks with high information scent and tasks with low information scent). The type of given tasks that were performed were the same for both groups (each group had the same high scent tasks and the same low scent tasks).

The number of participants in this experiment played a major role in getting consistent, reliable results, upon which meaningful judgment was made. All the participants understood both English and Swedish thus no cultural differences affected the results in terms of the WebPages that were in Swedish and needed English translations. The Age difference between the participants was also not wide and this ensured consistent performance from the participants. Gender preference was also balanced during the experiment so as to avoid any effects that gender imbalance could have on the results. Each group of participants (the ones that performed concurrent think-aloud protocol and the ones that did not perform think-aloud protocol on the tasks) had five females and five males.

Participants chosen for the main study were participants who would likely visit the given websites and would know how to go about various tasks for any given website. This ensured that all the chosen participants were real target users. The participants were made to feel comfortable to converse freely with the facilitator during the tests by encouraging them before
the test to speak their mind concerning the task at hand despite the facilitator’s presence in the room. This ensured that none of the participants was tensed during the tests and were all comfortable performing the tests and giving verbal feedback (for the concurrent think-aloud protocol participants) during the tests. These considerations played a big role during the tests and especially in cases whereby the participants used the concurrent think-aloud protocol because none of the participants were tensed and they performed the tests comfortably thus giving necessary feedbacks.

The test experiments essentially needed sixteen participants but twenty were needed for the actual participation because the extra four participants (two from each group) were considered as back-up and would be considered in case any of the sixteen participant’s test results were not reliable. The extra participants would have also come in handy in times when one of the sixteen participants could have faced a physical problem for instance vision problems that would affect the tests at hand.

3.3 Test Duration

Each participant spent at most 45 minutes for all the tasks/tests he/she was supposed to carry out. The facilitator carried out the exercise with 5 to 6 participants per day.

3.4 Selection of Tasks to be performed

The tasks and the web sites to be tested were selected based on the concept of information scent. The selection of tasks was a pre-study involving many tasks and then selecting the tasks with the lowest and the highest information scents respectively. The pre-study was performed by graphically finding tasks with strongest path predictability and also tasks with weakest path predictability based on a graph. Tasks with strongest path predictability were selected as the high scent tasks while the tasks with weakest path predictability were selected as low scent tasks. Three high scent tasks from high scent websites and three low scent tasks from low scent websites were selected. Each group (one using think-aloud protocol and another one not using think-aloud protocol) performed the same high scent tasks and the same low scent tasks so as to enable uniform comparison of results. The tasks were completely independent of each other. One of the factors that had to be considered so as to help facilitate the information scents concept was that the execution of each task that was chosen had to start at the homepage and be completed in one of the subsequent web-pages after the home page and not completed on the homepage.

The graphical method was used in the pre-study that was done to select high information scent and low information scent tasks because it proved to be more accurate and unbiased than the analytical method. The analytical method would have only represented one person’s (the facilitator’s) view on which tasks are considered high scent and which ones would be considered low scent. Views may however vary depending on one’s angle of perception. The graphical method therefore proved to be more unbiased because the determination of whether a task was low scent or high scent was determined by how a number of participants carried out the task. The graphical method was carried out as follows:
3.4.1 Graphical method used to determine and select the high information scent and low information scent tasks

- A large set of tasks were collected (ten tasks were collected). The tasks were on websites with one task per one website (one website per task – a task at each site).
- As a pre-study prior to the main study, twenty participants were randomly given the collected tasks to perform. Each participant randomly performed the ten tasks implying that each of the ten collected tasks was performed twenty times during the pre-study.
- As these participants carried out the tasks, the facilitator observed from the participant’s behavior on the main webpage to determine whether the task at hand was low information scent or high information scent for the respective participants. Number of wrong clicks on the main webpage was used as a determining parameter of whether a given task was low scent or high scent for a participant. The more the number of wrong clicks on the main webpage, the lower the information scent of a given task to a given participant. According to (Pirolli, Card, & Wege, 2001), visual search in focus plus context displays is highly affected by information scent. When users detect a strong information scent, they are able to reach their goal faster but when users detect a weak or no scent, they exhibit less efficient search. The number of wrong clicks was therefore a really good and reliable way of measuring the probability of a participant to find a suitable path to follow towards task completion by making the right choice (click) on the main webpage of any given task website. The number of wrong clicks indicated unsuccessful attempts towards task execution and this directly implied the difficulty of a participant to determine the right path through the elements on the webpage. The easier it was to find the right, first click on the main webpage of any given task, the faster the participant started on the right path to task completion and the fewer the number of wrong clicks.
- After noting down the total number of wrong clicks on the main webpage from all the twenty participants for each of the ten tasks, average number of wrong clicks on the main webpage per task was then found by dividing the total number of wrong clicks on the main webpage per task by twenty because there were twenty participants for the pre-study. The result was therefore the average number of wrong clicks on the main webpage per task.
- These observations were then plotted on a graph with the tasks on the x axis and the average number of wrong clicks on the main webpage for each task on the y axis. The higher the average number of wrong clicks on the main webpage that a task had, the lower the information scent of the task and the lower the average number of wrong clicks on the main webpage that a task had, the higher the information scent of the task. The three tasks with the lowest information scent and the three tasks with highest information scent were then selected from the graph. Out of the 10 tasks that were collected, 6 tasks were selected for the usability testing in this study as seen in Table 1 and Figure1 below.
Table 1: Tasks selected and the average number of wrong clicks on their main webpage

<table>
<thead>
<tr>
<th>Website</th>
<th>Task</th>
<th>Average number of wrong clicks on the main webpage</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://www.yale.edu">www.yale.edu</a></td>
<td>Find out from the website, the tuition fees for full-time study in the Yale Graduate School of Arts and Sciences.</td>
<td>23</td>
</tr>
<tr>
<td><a href="http://www.alcatel-lucent.com">www.alcatel-lucent.com</a></td>
<td>Find the latest list of press releases on the website.</td>
<td>19</td>
</tr>
<tr>
<td><a href="http://www.ki.se">www.ki.se</a></td>
<td>Look up for the vice-president of Karolinska Institute and get his fax number from the website without using the search function.</td>
<td>17</td>
</tr>
<tr>
<td><a href="http://www.imdb.com">www.imdb.com</a></td>
<td>Look up on the website and find a list of jobs offered at imdb</td>
<td>13</td>
</tr>
<tr>
<td><a href="http://www.migrationsverket.se">www.migrationsverket.se</a></td>
<td>Look up for the head of the press unit and get his telephone number from the website</td>
<td>10</td>
</tr>
<tr>
<td><a href="http://www.bbc.co.uk">www.bbc.co.uk</a></td>
<td>Look up for the comedy page and find the current most popular comedy clips</td>
<td>9</td>
</tr>
<tr>
<td><a href="http://www.liu.se">www.liu.se</a></td>
<td>Find the online LIU magazine 2010 - English version from the given website</td>
<td>7</td>
</tr>
<tr>
<td><a href="http://www.cnn.com">www.cnn.com</a></td>
<td>Convert 50 Japanese Yen to Swedish Kronor on the cnn website. The main task was to get to the function that helps to perform this currency conversion on the CNN website (Find the currency converter from the CNN website).</td>
<td>6</td>
</tr>
<tr>
<td><a href="http://www.studera.nu">www.studera.nu</a></td>
<td>Find out from the website the requirements needed for one to be admitted as a Masters student in Sweden.</td>
<td>3</td>
</tr>
<tr>
<td><a href="http://www.hilton.com">www.hilton.com</a></td>
<td>Find the list of Hilton experience rewards that are offered to members.</td>
<td>1</td>
</tr>
</tbody>
</table>
The final six selected tasks were therefore:

- The three with the lowest information scent which were: Yale, Alcatel and Karolinska
- The three with the highest information scent which were: CNN, Studera and Hilton

### 3.5 Order of Test Executions

The participants executed the task tests in random order. The random order was chosen to confirm consistency and increase the accuracy of the results. Each participant therefore performed all the tasks (both high scent tasks and low scent tasks) though in different order for each participant. The independent nature of the chosen tasks made it possible to randomize the order of tasks’ execution for each participant. The random orders in which the tasks were performed were based on the randomization Table 2 below.
### Table 2: Randomization of tasks

<table>
<thead>
<tr>
<th>Observation Usability Testing Method</th>
<th>Order of task execution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First Participant</td>
</tr>
<tr>
<td></td>
<td>1 2 3 4 5 6</td>
</tr>
<tr>
<td></td>
<td>Second Participant</td>
</tr>
<tr>
<td></td>
<td>4 5 6 1 2 3</td>
</tr>
<tr>
<td></td>
<td>Third Participant</td>
</tr>
<tr>
<td></td>
<td>1 4 2 5 3 6</td>
</tr>
<tr>
<td></td>
<td>Fourth Participant</td>
</tr>
<tr>
<td></td>
<td>2 3 5 4 6 1</td>
</tr>
<tr>
<td></td>
<td>Fifth Participant</td>
</tr>
<tr>
<td></td>
<td>3 6 1 2 4 5</td>
</tr>
<tr>
<td></td>
<td>Sixth Participant</td>
</tr>
<tr>
<td></td>
<td>5 1 4 6 3 2</td>
</tr>
<tr>
<td></td>
<td>Seventh Participant</td>
</tr>
<tr>
<td></td>
<td>6 2 1 3 5 4</td>
</tr>
<tr>
<td></td>
<td>Eighth Participant</td>
</tr>
<tr>
<td></td>
<td>2 1 6 4 5 3</td>
</tr>
<tr>
<td></td>
<td>Ninth Participant</td>
</tr>
<tr>
<td></td>
<td>1 6 5 4 3 2</td>
</tr>
<tr>
<td></td>
<td>Tenth Participant</td>
</tr>
<tr>
<td></td>
<td>2 1 6 5 4 3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Concurrent Think-Aloud Protocol Usability Testing Method</th>
<th>Order of task execution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Eleventh Participant</td>
</tr>
<tr>
<td></td>
<td>3 2 1 6 5 4</td>
</tr>
<tr>
<td></td>
<td>Twelfth Participant</td>
</tr>
<tr>
<td></td>
<td>4 3 2 1 6 5</td>
</tr>
<tr>
<td></td>
<td>Thirteenth Participant</td>
</tr>
<tr>
<td></td>
<td>5 4 3 2 1 6</td>
</tr>
<tr>
<td></td>
<td>Fourteenth Participant</td>
</tr>
<tr>
<td></td>
<td>6 5 4 3 2 1</td>
</tr>
<tr>
<td></td>
<td>Fifteenth Participant</td>
</tr>
<tr>
<td></td>
<td>3 4 5 6 1 2</td>
</tr>
<tr>
<td></td>
<td>Sixteenth Participant</td>
</tr>
<tr>
<td></td>
<td>2 3 4 5 6 1</td>
</tr>
<tr>
<td></td>
<td>Seventeenth Participant</td>
</tr>
<tr>
<td></td>
<td>5 6 1 2 3 4</td>
</tr>
<tr>
<td></td>
<td>Eighteenth Participant</td>
</tr>
<tr>
<td></td>
<td>1 6 3 2 4 5</td>
</tr>
<tr>
<td></td>
<td>Nineteenth Participant</td>
</tr>
<tr>
<td></td>
<td>3 5 2 6 1 4</td>
</tr>
<tr>
<td></td>
<td>Twentieth Participant</td>
</tr>
<tr>
<td></td>
<td>4 5 3 1 2 6</td>
</tr>
</tbody>
</table>

**Key:**
- 1 represents Yale : low scent task
- 2 represents Alcatel : low scent task
- 3 represents Karolinska : low scent task
- 4 represents CNN : high scent task
- 5 represents Studera : high scent task
- 6 represents Hilton : high scent task

### 3.6 Tools and Methodology

Eye-tracking was used as a methodology to collect in-depth results concerning the participant’s eye movements on the screen during the testing session for both groups. Eye-tracker was therefore used as a tool consisting of both a camera that was used to capture the eye movements and also eye-tracker software that output the eye-movement results on the evaluator’s screen in form of heat maps, gaze plots and numerical statistics. Measures taken during the tests were compared and the differences noted. Results collected through the eye-tracker assisted to determine which type of tasks (between the low scent and high scent tasks) were affected by the concurrent think-aloud protocol method of usability testing.

The hardware and software equipments used during the tests had the following specifications:

- The eye-tracking machine was an iView X RED250 System from Senso Motoric Instruments (SMI).
The eye-tracking camera was attached to a flat screen Samsung 943B computer monitor model.
The stimulus computer used was of the model: Fujitsu Siemens LIFEBOOK E Series
The programs/Software used in the stimulus computer for recording and analysis of the results were: SMI iView X, SMI BeGaze 3.0 and SMI Experiment Center 3.0.
The Stimulus computer’s monitor display was set to 2 (1280 x 1024).
The participants had access to a Swedish keyboard and mouse.
The Eye-tracking machine, camera and flat screen monitor were placed on an adjustable table and the participant sat on a still chair. The adjustable table helped to adjust the position of the screen for the participant to have a comfortable view and the right calibration and the still chair ensured that the participant was not subjected to body movements that could have affected the end results.
A stop watch was used to measure the task time (the amount of time taken by a participant to accomplish a given task).

3.7 Test Procedure/Process
The test facilitator/evaluator first carried out a pilot test of the available tasks so as to ensure that no unexpected problems would be faced by the real test participants and that everything would run smooth. The pilot study, which served as a feasibility test for this study, was as follows:

3.7.1 Pilot test procedure

- The main facilitator confirmed that the other facilitators/evaluators (if any are needed for the study) are sufficiently skilled in the requirements and procedures of the test.
- The facilitator got a few participants (about three) for the pilot test
- The facilitator asked the three participants to fill out an initial form whereby the test participant was meant to give details about his/her background, vision health status and contact details. The facilitator then went through the information the participants had given through these forms so as to ensure that the form was designed in a way that the participants could easily give the right/needed information (ensuring the form lacks ambiguity).
- The facilitator then went through the general instructions for the test keenly and once he/she ensured no grammatical errors and that the instructions suit this study, he/she read out the instructions to the three participants and then asked for their opinion concerning the ease in understanding those instructions. This helped the facilitator to prove whether the general instructions were comprehensible.
- The facilitator checked and familiarized himself/herself with all the equipments and apparatus involved in the study. The facilitator therefore checked the eye-tracking
machine and operated it (both as a participant-input and as an evaluator-output), the monitors (screens) in the test room and the camera attached to the screen and the eye-tracking machine. This helped the facilitator to ensure that the equipments provided for the test met the needed specifications and operated correctly.

- The facilitator then asked the three pilot test participants to sit comfortably and then to begin their tests on the tasks that were to be used for the real/main testing. Through this, the facilitator was confirming that an average user/participant would be conversant with the equipments/apparatus, for instance, keyboard, mouse and the screen.

- The facilitator performed the calibration on the pilot test participants then asked the pilot test participants to start performing their tasks based on the general instructions given to them for each task. This proved to the facilitator whether an average user could perform the given tasks successfully and with their eyes fixed on the screen as needed so as to record results. Any difficulty faced by any participant at this point was noted and the observation used to improve on the main test session.

- One of the pilot test participants performed the normal usability testing while two of them performed the concurrent think-aloud protocol on their given tasks. This helped the facilitator to determine whether an average participant in the study would need some level of intervention from the facilitator (especially for the concurrent think-aloud protocol users) for instance asking a few leading questions so as to get the much needed results.

- Once the pilot test was over, the facilitator thanked the pilot test participants and then they were free to leave the room.

- After the trials with all the pilot test participants, the test facilitator tested the analysis software by analyzing the participant’s results so as to ensure that the analysis software works. The facilitator therefore analyzed the results as recorded by the eye-tracker and determined whether the results got were reliable and valid even though it was on a small scale. At this point, the facilitator also confirmed if the analyzing software worked correctly and whether some valid data could be got out of it.

- The facilitator then analyzed the whole testing procedure (involved ideas or advice from other experts) to identify any adverse effects caused by the procedure as a whole and the effectiveness of actions to reduce them.

After the pilot test, the real test participants proceeded to carry out the real tests as follows:
3.7.2 Main test procedure

- The test facilitator met each of the twenty test participants (one at a time) outside the experiment room and showed them to the right place in the test room. The facilitator then explained a little bit about what test would go on, in this case, usability test.

- Each chosen participant met the necessary background requirements and had good vision. A participant also read and signed an information sheet to show his/her approval to participate in the test. In this form, the participant also filled in his/her email address to allow for continued contact if necessary concerning the usability testing.

- The facilitator read out and explained general instructions of the test to the test participant.

- A participant was then asked to sit comfortably in front of the computer screen. The screen was set to perform the calibration of the eye-tracking camera in a satisfactory manner. Before the calibration (calibration was done before each new task), the participant was asked to focus his/her gaze on a plus sign in the middle of the screen. The calibration was started after he/she focused his/her gaze on the plus sign for about 1 second. This ensured that all participants started their calibration from the same point of perception. The participant was then asked to follow a mobile/moving red dot on the screen with his/her eyes for a few seconds and when the dot stopped, the calibration results (showing the current deviation of the participant’s eye and the red dot) were displayed on the test facilitator’s (results’) screen. The current deviation value should not be above +0.07 and should not be below -0.07 with 0.00 considered the best results for the calibration. This step was necessary because it helped to set the eye-tracking camera and adjust the necessary factors till satisfactory calibration settings were achieved. Measures that were taken to achieve the best calibration results included: asking the participant to be still, sit upright and comfortably on his/her seat with his/her eyes constantly focused on the screen, adjusting the adjustable table either to move up or down and at times asking the participant to move closer or further from the screen depending on the settings. Knowing the right sitting position of the participant was made possible by the iViewX Red4 Firewire software.

- Once a participant’s current deviation was noted to be within the right range and the calibration results were satisfactory, the participant was signaled by word of mouth to start the tests of the given tasks.

- The websites that were already uploaded on the SMI Experiment 3.0 software with the order based on the randomization table, each having a description of the task before the website, showed on the participant’s screen. The participant read each description
of a given task before performing the task on the given website. Each task had its own website and the participant performed one task at a time.

While performing all the tasks, the participant was not supposed to get his/her eyes off the screen and calibration was performed before each task so as to ensure accurate results were recorded for each task performed.

**For the participants not using Think-Aloud protocol method:**

- The test facilitator sat silent in the room and observed the participant as the participant performed the tests on each given task.

- The participant performed the given tasks silently without giving any concurrent verbal feedback during the test and only notified the facilitator when he/she had finished performing all the tasks.

**For the participants using concurrent Think-Aloud protocol method:**

- Before the test started, the test facilitator described in detail what the concurrent think-aloud process is and how it should be performed. The facilitator at this stage told the participant that they should verbalize their thoughts as they perform each given task.

- The test facilitator then demonstrated to a participant how a concurrent think-aloud protocol is carried out. During this demonstration, the test facilitator acted as the participant and carried out a sample task (not one of the chosen tasks) as she verbalized her thoughts during the task performance and the real participant that was supposed to use a concurrent think-aloud protocol watched her.

- After the practical demonstration of how a concurrent think-aloud protocol is be done, the facilitator asked the participant to try performing another sample task (not one of the chosen tasks) as he/she verbalized their thinking process while carrying out the task. The participant repeated this step thrice in the presence of the test facilitator till the facilitator was comfortable that the participant performed the concurrent think-aloud protocol process well while performing a given task.

- Since the chosen participants were encouraged not to be tensed during the tests, they were comfortable to verbalize their thoughts as they carried out the tests in the presence of the facilitator without any tensions. (Refer to section: 3.2.3).
Calibration was then done to the participant and once the calibration results were satisfactory, the participant started the real test using the concurrent think-aloud protocol method.

The test participant verbalized his/her thinking process when carrying out each task. The participant in this case freely gave a verbal feedback to the test facilitator on what he/she was thinking of as he/she carried out the test with his/her eyes focused on the screen.

The test facilitator played a minimal and passive role during the tests and gave short encouraging responses when the participant tried to involve her during the task performance so as not to interrupt the participant’s natural thinking process. The short answers like; ok”, “mm”, and “ah”.

While each participant performed the given tasks, the participant’s performances were recorded with the iView X software.

Once any participant (either using concurrent think-aloud protocol or not using think-aloud protocol) was done with the tests, he/she told the test facilitator that he/she was done and the facilitator thanked him/her for participating in the test exercise and revealed the real purpose of the study; mainly to evaluate the effects that concurrent think-aloud protocol method has on tasks with low information scent and tasks with high information scent in usability testing.

Finally, the test facilitator asked the participant not to reveal the real purpose of the study before the outcome of the study is presented to the public and the participant was thanked once again for their participation.

The whole process in this “Main test procedure” was carried out for each of the twenty participants that participated in the test. The part “For the participants not using Think-Aloud protocol method” applied for the participants that did not use concurrent think-aloud protocol during the tests and the part “For the participants using concurrent Think-Aloud protocol method” applied for the participants that performed the given tasks using concurrent think-aloud protocol.

The facilitator then analyzed the test results using eye-tracking analysis software called SMI BeGaze 3.0.
3.8 Dependent Measures

The results were measured mainly using task times and eye movements. Eye movements were recorded by the eye-tracker. Fixations, task time and amount of screen areas viewed on the main webpage were analyzed using the SMI BeGaze 3.0 eye-tracking software. Saccades which refer to the rapid eye movements from one fixation to the next were also recorded during the experiment and this helped to see the sequence in which a participant’s eyes moved across the screen while performing the task. The saccades joined the numbered fixations to each other and the fixations were numbered according to their sequence of occurrence.

The independent measures in this study test were: testing while using concurrent think-aloud protocol and testing without using think-aloud protocol. The dependent measures were: the task time, the number of fixations and the amount of screen areas viewed on the main page.

A task time was measured from the time a participant starts to look at the main webpage to the time when the participant makes the first clicks on a tab or menu or any tool on the webpage that takes the participant to the next page from the main webpage, as a means of performing the task. The task time measured in this case is therefore the task time spent on the main webpage thus the time taken by a participant on the main webpage while performing a task.

The number of fixations measured in this study refers to the number of fixations recorded on the main webpage when a participant performs a given task on a given website.

The amount of screen areas viewed in this study refers to the amount of screen areas recorded on the main webpage when a participant performs a given task on a given website.

3.9 Expected Results

As stated in this study, a previous study done by (Pirolli, Card, & Wege, 2001), proved that the effective area of visual attention changes with information scent and density; with high scent fixation movements being longer than low scent by about 25%. This, in combination with another study that proved no noted differences between using think-aloud protocol and not using think-aloud protocol as a usability testing method on normal, smaller tasks whereby the number of fixations and number of screen areas viewed were the same, would help us to know the effect that concurrent think-aloud protocol would have on the tasks that are not just normal tasks but have either high attraction (high scent) or low attraction (low-scent) to the participant.

How would concurrent think-aloud protocol affect the number of fixations in high scent tasks and in low scent tasks?

How would concurrent think-aloud protocol affect the previously stated difference in length of fixation movements between high-scent and low-scent tasks?

It is expected from the background of this study that the length of fixation movements would reduce when using concurrent think-aloud protocol on high-scent tasks and that it would increase when using concurrent think-aloud protocol on low-scent tasks; thus reducing the
estimated 25% margin earlier stated to be the difference in length of fixation movements between high-scent tasks and low-scent tasks.

The inner strength of the navigational path’s clues that is evident in high-scented tasks would keep the participant glued to the task (*scent following*) so as to give relevant feedback to the facilitator when using concurrent think-aloud protocol on high scented tasks. This would mean that the participant would be viewing mostly one area and the close areas around it thus leading to less number of fixations and shorter lengths of fixation movements.

The weak clues to the destination information that is evident in low-scent tasks would keep the participant searching for more relevant and “right path through the task” on the webpage/screen (*scent searching - scent finding*) so as to collect the relevant information concerning the task at hand and give relevant feedback to the facilitator. This would mean more screen areas viewed, more fixations recorded during the search for the right task-path due to low-scent and longer time taken to perform the given tasks unlike in high scent tasks.

With the surveying behavior observed in some task executions during the study; information scent will prove to be quite an interesting twist to the much expected results of this study.
4. RESULTS

The task on www.imdb.com where “a participant was supposed to look up on the website and find a list of jobs offered at imdb” was used as a warm up task so as to make a participant feel comfortable, familiar and understand the experiment environment and how the test was to be executed based on the given instructions. The results of the warm-up task were not recorded.

Fortunately, no participant was excluded from the tests. All the twenty participants participated and their results were used for analysis. Samples of images of the recorded test results (fixations and screen areas viewed) have been listed at the end of this document as attachments in the appendix section.

The total number of areas of interest (AOIs) was determined by the test facilitator as sixty four (eight rows and eight columns) for each of the web pages that was tested. The number of screen areas viewed out of the total sixty four areas on the main page, the fixations and the time taken to execute each task were recorded. The averages of the results from each group of ten participants were as follows:

4.1 Results Presentation

4.1.1 Task one (Low scent task)

The website (www.yale.edu) gave a total of sixty four areas of interest. A total of twenty participants completed task one. Ten participants performed the task on the website as they concurrently thought aloud and gave feedback to the test facilitator. The other ten participants performed the task on the website without thinking aloud.

Figure 2 below shows the average number of screen areas viewed (out of the sixty four areas of interest on the website) on the main page for each group of ten participants; when using concurrent think-aloud protocol and when not using think-aloud protocol.
Figure 2: Average amount of screen areas viewed in task 1

Figure 3 below shows the average number of fixations on the main page for each group of ten participants that were recorded for task one during the concurrent think-aloud protocol method and when not using think-aloud method:

Figure 3: Average number of fixations in task 1
Figure 4 below shows the average total task time it took to execute task one for each group of ten participants when using concurrent think-aloud protocol and when not using concurrent think-aloud protocol.

After statistical analysis of the results, it was noted that higher number of fixations was recorded when performing task one while using concurrent think-aloud protocol than when not using think-aloud protocol method. The number of screen areas viewed was also higher when using concurrent think-aloud protocol than when not using concurrent think-aloud protocol while performing the task. It took more time to execute task one while using the concurrent think-aloud protocol than while not using think-aloud protocol.

4.1.2 Task two (Low scent task)

Task two was carried out on the website: www.alcatel-lucent.com. Each of the twenty participants performed task two with ten of them performing the task while concurrently thinking aloud as a method of giving task feedback to the test facilitator and the other ten performed the task without thinking aloud. The website was divided into sixty four areas of interest.

Figure 5 below displays the average number of screen areas viewed on the main page for each group of ten participants while performing task two using concurrent think-aloud protocol and while not using think-aloud protocol.
Figure 5: Average amount of screen areas viewed in task 2

Figure 6 below shows the average number of fixations recorded on the main page for each group of ten participants when task two is executed using concurrent think-aloud protocol and while not using think-aloud protocol.

Figure 6: Average number of fixations in task 2
Figure 7 below represents the average total amount of time it took to complete task two for each group of ten participants using concurrent think-aloud protocol and not using think-aloud protocol.

Based on the statistical analysis of the test results of task two, more screen areas were viewed when a participant was using concurrent think-aloud protocol during the tests as compared to when not using think-aloud protocol. Higher numbers of fixations were also recorded on average when a participant was using concurrent think-aloud protocol as opposed to when not using think-aloud protocol. It took a participant more time to execute task two as he/she concurrently thought aloud than when not thinking aloud while performing the task.

4.1.3 Task three (Low scent task)

Task three was executed on the www.ki.se website by twenty participants, each at a time. Ten of the participants were concurrently thinking aloud while performing the task while the other ten were not thinking aloud while performing the task. The website for task three was divided into sixty four areas of interest.

Figure 8 below shows the average number of screen areas (out of the total sixty four screen areas) viewed on the main webpage for each group of ten participants while performing task three using concurrent think-aloud protocol and when not using think-aloud protocol.
Figure 8: Average amount of screen areas viewed in task 3

Figure 9 below displays the average number of fixations recorded on the main page for each group of ten participants while performing task three using concurrent think-aloud protocol and while not using think-aloud protocol.

Figure 9: Average number of fixations in task 3
Figure 10 below shows the average total task time it took to execute task three for each group of ten participants while using concurrent think-aloud protocol and while not using think-aloud protocol.

From the statistical analysis done on task three, it was evident that more fixations were recorded when a participant was performing task three using concurrent think-aloud protocol than when performing the task without thinking aloud. The number of screen areas that were viewed by a participant when performing task three while concurrently thinking aloud was double the number of the screen areas viewed while performing task three without thinking aloud. It also took more time for a participant to complete task three when concurrently thinking aloud than when not thinking aloud while performing the task.

4.1.4 Task four (High scent task)

Task four was chosen as one of the high-scent tasks. It was performed on the www.cnn.com website which was divided into sixty four areas of interest. Ten participants performed the task while concurrently thinking aloud and another ten participants performed the task without thinking aloud.
Figure 11 below shows the average number of screen areas viewed on the main webpage (out of the total sixty four areas of interest) for each group of ten participants when executing task four while concurrently thinking aloud and while not thinking aloud.

![Task 4: Average Amount of Screen Areas Viewed](image)

**Figure 11: Average amount of screen areas viewed in task 4**

Figure 12 below shows the average amount of fixations recorded on the main webpage for each group of ten participants when task four is executed while concurrently thinking aloud and while not thinking aloud.

42
Figure 12: Average number of fixations in task 4

Figure 13 below shows the average total task time it took for each group of ten participants to execute task four completely while concurrently thinking aloud and while not thinking aloud.

Figure 13: Average total task time in task 4
The statistical analysis of task four showed no difference in the number of fixations recorded when performing task four using concurrent think-aloud protocol and while not using think-aloud protocol. The number of screen areas viewed when performing task four while concurrently thinking aloud was negligibly lower than the number of screen areas viewed when performing task four while not thinking aloud. It took approximately the same time to execute task four while concurrently thinking aloud as it took to execute task four while not thinking aloud. No major differences were evident in the time records.

### 4.1.5 Task five (High scent task)

Task five was executed on the [www.studera.nu](http://www.studera.nu) website. The website was logically divided into sixty four areas of interest. Ten participants executed task five while concurrently thinking aloud and ten other participants executed task five while not thinking aloud.

Figure 14 below shows the average number of screen areas viewed on the main webpage for each group of ten participants when performing task five while concurrently thinking aloud and while not thinking aloud.

![Figure 14: Average amount of screen areas viewed in task 5](image)

Figure 15 below shows the average number of fixations recorded on the main webpage for each group of ten participants when task five is executed while concurrently thinking aloud and when task five is executed while not thinking aloud.
Figure 15: Average number of fixations in task 5

Figure 16 below shows the average total task execution time for task five for each group of ten participants when performed while concurrently thinking aloud and when performed while not thinking aloud.

Figure 16: Average total task time in task 5
The statistical analysis for task five showed that the number of fixations recorded when task five was executed using concurrent think-aloud protocol method and when not using think-aloud protocol were roughly the same. No much difference was noted in the number of fixations recorded in both cases. The number of screen areas viewed when performing task five while using concurrent think-aloud protocol was almost the same as the number of screen areas viewed when performing task five without thinking aloud. It took a participant roughly the same amount of time to execute task five while concurrently thinking aloud as it took to execute it while not thinking aloud.

### 4.1.6 Task six (High scent task)

Task six was executed on the [www.hilton.com](http://www.hilton.com) website. The website was divided into sixty-four areas of interest. A total of twenty participants executed task six, each at a time. Of the twenty participants, ten executed task six using concurrent think-aloud protocol method and the other ten executed task six without thinking aloud.

Figure 17 below shows the average number of screen areas viewed on the main webpage for each group of ten participants when executing task six using the concurrent think-aloud protocol method and when executing task six without using think-aloud protocol.

![Graph showing the average amount of screen areas viewed in task 6](image)

**Figure 17: Average amount of screen areas viewed in task 6**
Figure 18 below shows the average number of fixations recorded on the main webpage for each group of ten participants when executing task six while using think-aloud protocol method and while not using think-aloud protocol method.

![Bar Chart showing average number of fixations in task 6](chart.png)

**Figure 18: Average number of fixations in task 6**

Figure 19 below shows the average total amount of time taken to execute task six for each group of ten participants while using the concurrent think-aloud protocol method and also while not using think-aloud protocol method.
Statistical analysis of task six proved that the number of fixations that were recorded while executing task six using concurrent think-aloud protocol were roughly the same as the number of fixations that were recorded when the participant was not using think-aloud protocol. The number of screen areas viewed however proved to be a bit higher when performing task six using concurrent think-aloud protocol than when performing task six without using any think-aloud protocol. This was a bit different from the statistical analysis of task four and task five that were both high scent tasks and had each got very slight and thus negligible difference in the number of screen areas viewed while using concurrent think-aloud protocol and while not using any think-aloud protocol to perform the respective task. It took roughly the same time to execute task six while concurrently thinking aloud and while not thinking aloud during task performance.

4.2 Summary of the Results

For tasks one, two and three the results were highly statistically significant with significant differences in the total task time, task time on the main webpage, number of fixations and amount of screen areas viewed. However, no statistically significant differences were found for task four, five and six except for the amount of screen areas viewed in task six.

The data underlying all the averages for task one to six were also tested. The averages were compared with the t-test for normality to confirm whether the data was normally distributed according to the Shapiro-Wilk normality test, (Shapiro & Wilk, 1965). The findings were that all the data except the data for the number of fixations for task four passed the test of comparison with the t-test as follows:

Figure 19: Average total task time in task 6
Table 3: Shapiro-Wilk normality test results for each task

<table>
<thead>
<tr>
<th>Task</th>
<th>Total task time (t-test p-value)</th>
<th>Task time on the main page (t-test p-value)</th>
<th>Number of fixations on the main page (t-test p-value)</th>
<th>Amount of screen areas viewed on the main page (t-test p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>0.00330424</td>
<td>0.001495845</td>
<td>2.50699E-06</td>
<td>0.000132513</td>
</tr>
<tr>
<td>Task 2</td>
<td>0.000303647</td>
<td>9.78128E-07</td>
<td>1.40102E-12</td>
<td>2.1909E-09</td>
</tr>
<tr>
<td>Task 3</td>
<td>1.01406E-11</td>
<td>1.68174E-09</td>
<td>3.57639E-15</td>
<td>1.51083E-10</td>
</tr>
<tr>
<td>Task 4</td>
<td>0.954435383</td>
<td>0.99748643</td>
<td>0.970049694</td>
<td>0.325868617</td>
</tr>
<tr>
<td>Task 5</td>
<td>0.986744805</td>
<td>0.941417008</td>
<td>0.95429897</td>
<td>0.951117586</td>
</tr>
<tr>
<td>Task 6</td>
<td>0.990849816</td>
<td>0.770005246</td>
<td>0.98083532</td>
<td>0.000859965</td>
</tr>
</tbody>
</table>

Since the normality of the data is an assumption for using the t-test, a non-parametric test was also performed for this particular case. The chosen was the Mann-Whitney U test. The result was that there was no significant difference (same result as for the t-test) with U=49 and p=0.9681 for Mann-Whitney U (p-value, two-tailed).

After all the statistical analysis and clustering the results of the high information-scent and low information-scent, below were the clustered results:

![Figure 20: Task time averages (low information scent tasks)](image-url)
Figure 21: Task time averages on main page (low information scent tasks)

Figure 22: Average number of fixations on main page (low information scent tasks)
Figure 23: Average number of areas with fixation on main page (low information scent tasks)

Figure 24: Task time averages (high information scent tasks)
Figure 25: Task time averages on main page (high information scent tasks)

Figure 26: Average number of fixations on main page (high information scent tasks)
Figure 27: Average number of areas with fixations on main page (high information scent tasks)
5. DISCUSSION

The experiment that was carried out in this study demonstrated significant results. A trend of consistent results was quite evident on how the participants performed most of the tasks, which proved to be very useful during the results analysis.

5.1 Tasks Completion

As mentioned in section 3.2, the pre-study was carried out by twenty participants, the pilot study was carried out by three participants and the main study was carried out by twenty participants (ten using concurrent think-aloud protocol and ten not using think-aloud protocol). The twenty participants involved in the pre-study were different from the twenty participants involved in the main study and also different from the three participants involved in the pilot study. Each study stage therefore had its own participants chosen based on the criteria mentioned in section 3.2 above.

All the ten tasks were successfully completed by all the twenty participants during the pre-study session. Correct solution to each task was given by each of the twenty participants in the pre-study stage within a reasonable amount of time. Similarly, all the six tasks that were chosen during the pre-study (three high-scent and three low-scent tasks) were successfully completed by the other twenty participants during the main study. Ten of the main study participants performed the six tasks using concurrent think-aloud protocol and the other ten performed the six tasks while not using any think-aloud protocol. They all gave the correct solution to each of the six tasks within a reasonable amount of time. The criteria for choosing the participants (as mentioned in section 3.2) were strictly adhered to and this contributed positively to the completion of tasks at any given stage.

5.2 Accuracy of the Tests in this Study

The results of all the WebPages were recorded for each task per participant but analysis were mainly based on the results on the main webpage because a given task might have a longer path to task completion than others whereby a participant has to visit several WebPages before completing the task as compared to another task whereby the participant might complete the task on the second webpage. It would therefore take a longer total time and even record more total fixations and more total amount of screen areas viewed when performing the task with longer task completion path even if its information scent on the main page is higher than the other tasks with shorter task completion path.

The information scent of a task during the pre-study was determined by the average number of wrong clicks on the main webpage per task. This is because a task can have high information scent on the main page but have low information scent on the subsequent web pages, which might affect the number of wrong clicks per task if the total number of wrong clicks on all the web pages of a task were to be considered. The varying information scents across the web pages would therefore have affected the results of the pre-study and the final results’ analysis and thus the pre-study and the main study were both focused on the main page of the given websites for each task.
The other reason that made this study be focused on the main webpage and not all the web pages of a given task, is the fact that a given task will always have an exact/same route on the main webpage for instance, one has to click a specific tab of a specific menu to start-off the task but in the other web pages subsequent to the main webpage, different participants might use different routes/paths to perform the task to completion. A given task might therefore give very different results (non-consistent results) for different participants who have performed the same task but through different paths/routes if the total results across all the web pages were to be considered. The main page was therefore considered to be an equal platform for judging the performance of all the participants for each given task.

Each webpage on which a given task was performed was divided into sixty four different areas of interest. The reason behind this was the fact that the more the number of areas of interest a website is divided into, the more accurate the “amount of screen areas viewed” results. Dividing a webpage into 16 areas of interest therefore proved to be smaller compared to sixty four areas of interest. Sixty four areas of interest which was the maximum available implied that each webpage was divided into eight rows and eight columns for better results. The actual structure of each webpage/website was not taken into account during the division of a webpage into sixty four areas of interest because this study was not comparing between tasks (e.g. comparison between high scent tasks and low scent tasks) but was comparing between the effects of using concurrent think-aloud protocol and not using concurrent think-aloud protocol on a given task. Not taking the structure of the webpage into account during the division would therefore not affect the results because both the tests: “using concurrent think-aloud protocol” and “not using think-aloud protocol” would be carried out on the same webpage with the same number of divisions of areas of interest (sixty four) and the same webpage structure for each task. The number of divisions of areas of interest (sixty four) and not taking the webpage structure into consideration would therefore be common when carrying out all the type of needed tests on a given webpage for a given task.

Concurrent think-aloud protocol was carried out correctly according to the main test procedure steps listed in section 3.7.2. When trying concurrent think-aloud protocol at the pilot study stage with one of the pilot study participants, it was noted that the participant was tensed and did not look comfortable performing the test in the presence of the test facilitator while giving verbal feedback at the same time. This in return, affected the performance of the participant in the given task. After this observation, it was decided that all the participants who were to participate in the main study (especially the ones using concurrent think-aloud protocol), had to be encouraged before the tests to be comfortable during the tests and to speak their minds concerning each task at hand without fear of the facilitator’s presence in the room. This made the participants feel comfortable to carry out their respective tests and give adequate and necessary verbal feedback in the presence of the test facilitator without being tensed. The chosen twenty participants for the main study met all the criteria described in section 3.2.3 and the test facilitator demonstrated to the ten participants who used concurrent think-aloud protocol during their tests, how to go about concurrent think-aloud protocol and asked each of the ten participants to perform concurrent think-aloud protocol testing using some sample tasks that were not part of the main tasks (as mentioned in section 3.7.2). This
acted as a way of mini-training to the ten participants on how to effectively carry out the concurrent think-aloud protocol so as to get reliable results; it therefore prepared the participants mentally for the concurrent think-aloud process. Once the test facilitator had ensured that a participant was carrying out the concurrent think-aloud protocol correctly and effectively, the participant then performed the main tasks in the main study using concurrent think-aloud protocol. The test facilitators therefore stuck to the protocols of the concurrent think-aloud protocol method and ensured that it was done the way it is usually done in usability testing. For this reason, the results that were got from the ten participants who tested using concurrent think-aloud protocol had a high validity just like the results of the participants who tested without using any think-aloud protocol. The concurrent think-aloud protocol was therefore done the way it is normally done and no aspect of it was overlooked.

The level of communication between the test facilitator and the participants was good and moderate. The interaction between the test facilitator and the participants that were executing the given tasks using the concurrent think-aloud protocol; for instance was kept at a minimal level so as not to interfere with the participant’s natural thinking process. Short, encouraging feedbacks like “ok”, “mm”, and “ah” ensured that the involvement of the facilitator was minimal during the concurrent think-aloud tests and thus the participant was not frequently interrupted.

5.3 Discussion of the Statistical Results

From the statistical analysis of all the results of twenty participants, it was clear that concurrent think-aloud protocol had effects on the low scent tasks. Higher number of fixations, higher number of screen areas viewed and more time for task execution were recorded when the participants performed task one, task two and task three using concurrent think-aloud protocol than when not using any think-aloud protocol. Most of the participants seemed to look around more on the main webpage and even fix their gazes a bit longer on specific points while carrying out low scent tasks using the concurrent think-aloud protocol than when not using think-aloud protocol. The higher numbers of recorded fixations were also noted to be more widespread on the main WebPages when the participants carried out the low scent tasks using the concurrent think-aloud protocol.

On the other hand, concurrent think-aloud protocol did not seem to have substantial effects on the high-scent tasks. The performances on task four, task five and task six by the twenty participants yielded results that showed relatively the same number of fixations when concurrent think-aloud protocol was used and when no think-aloud protocol was used in performing the tasks. There were negligible differences in number of screen areas viewed when using concurrent think-aloud protocol and when not using concurrent think-aloud protocol while performing task four and task five. The number of screen areas viewed while performing task four using concurrent think-aloud protocol were mostly lesser by two than when not using any think-aloud protocol on task four by the participants. The numbers of screen areas viewed were mostly recorded to be higher by only one when task five was executed by the participants using concurrent think-aloud protocol than when not using any think-aloud protocol. Task six however yielded strange results on the number of screen areas
viewed as compared to the other high-scent tasks. Higher numbers of screen areas viewed were recorded when most of the participants performed task six using concurrent think-aloud protocol than when not using any think-aloud protocol. Given the inconsistent results that were given on the number of screen areas viewed when the participants performed task four, task five and task six, no substantial conclusion could be deduced on the effect the concurrent think-aloud protocol had on the number of screen areas viewed in regards to the high-scent tasks. Amounts of time taken to execute the high-scent tasks by the participants using concurrent think-aloud protocol and when not using concurrent think-aloud protocol were relatively the same as no substantial differences were noted. Concurrent think-aloud protocol therefore had no notable effects on the high scent tasks as compared to the effects it had on the low-scent tasks.

5.4 Odd Result of Task Six

A deeper look at the individual fixation graphs for each of the twenty participants indicated that websites with less and scattered content prompted the participant to look around more even though the chances of making a wrong click on such a webpage is very minimal thus making it possible to record high number of “amount of screen areas viewed” and high number of fixations despite the type of information scent (high-scent or low-scent) that a task on such a website might have. This effect was mostly higher when participants use concurrent think-aloud protocol to perform a task on a more scattered website with less content than when not using think-aloud protocol. This explains the trend that was seen with task six. A look at the results from each participant for this particular task showed that most of the participants who performed task six using concurrent think-aloud protocol looked around more around the webpage and especially at the webpage’s texts. The amount of screen areas viewed were therefore many but with brief gazes on most of the areas viewed. The participants who performed task six without using think-aloud protocol on the other hand, looked around too but not as much. Their gazes in most of the areas of the screen viewed were however a bit more. This resulted to task six recording higher numbers of screen areas viewed when performing concurrent think-aloud protocol than when not using any think-aloud protocol while executing the task. The average numbers of fixations recorded were however the same in both cases since the ones that had lesser number of areas viewed, recorded higher number of fixations in most areas viewed. The general structure of task six goes ahead to explain this. The main webpage from which the results were collected from, has few scattered contents with some few, catchy texts and a big image that covers almost half the page. Most of the fixations were recorded in areas with text. The curiosity on why the task gave unexpected results despite it being the task with the highest information scent made me inquire from most of the participants after the test on how their experience with this particular website was as compared to other websites. From the answers the participants gave, it seems most participants felt that since the website has very brief and somewhat scattered content compared to the others, they could survey the whole page very quickly before starting the main task since it felt that it is easy to have a grasp of every main functionality with a quick glance. The effect was more for participants who were using concurrent think-aloud protocol to carry out the task because most of them felt that since the website did not have many
things, they could quickly wrap the whole page in a nutshell with a quick look around everything so as to get some main points to give as feedback to the facilitator concerning the website in general as well as the task. This could also be felt through the verbal feedback the participants gave. At first, most of the participants who performed task six using concurrent think-aloud protocol, gave a lot of descriptive feedback of the whole website/main webpage in general before starting to talk about the main task at hand. The participants who were not performing concurrent think-aloud protocol on task six felt the effect but it was not so much since they did not have to give any feedback but they still felt that since the website was brief, they could go through it quickly just to know what it’s all about before settling to perform the main task. This could have proved to be a little harder on the other tasks that had web pages full of content. Participants still looked around on such web pages but mostly dwelled on the main features of the webpage as they felt that they could not quickly have a look at everything when a webpage is full of texts and images. Most fixation graphs show that on the web pages of the other tasks (apart from task six), most participants looked around but did not necessarily look at the whole page from top to bottom, probably due to a lot of content but for task six, most participants glanced at almost the whole page from to bottom, probably because they felt that the content was not much and they could therefore have a look at everything at once as much as they knew where to start the task from, perhaps this explains why task six recorded the lowest number of wrong clicks but had a high number of screen areas viewed, fixations and probably task time on the main page as compared to some tasks that had lower scent than task six. Since this study was not focusing on comparison between tasks though, I mostly concentrated on the comparison between executing each of the given tasks while using concurrent think-aloud protocol and while not using any think-aloud protocol. The nature of the website therefore over- rode the effect of information scent of the task particularly on task six and the effect must have been slightly more profound on the participants who were using concurrent think-aloud protocol.

Judging from the participants’ verbal feedback during the tests and the results, the big image on the main webpage was also a contributing factor to the seemingly odd results that were recorded on task six. The image not only covered half the webpage but also kept changing with different pictures loading after every few seconds. This distracted the participants’ concentration on the task and thus few fixations were recorded on the image but with longer gazes. This in return resulted to higher amount of time participants took to execute task six.

5.5 Qualitative Aspects of Concurrent Think-aloud Protocol

Qualitative aspects of concurrent think-aloud protocol; were also evident during the tasks’ performances by the participants. When the ten participants who used concurrent think-aloud protocol performed the tasks, any satisfaction, irritation, doubts and surprises that the participant felt when executing the task, was exhibited through the verbal feedback that the participant gave. This kind of emotional response could not be felt from the participants who were not using concurrent think-aloud protocol to execute the given tasks. For instance, a lot of verbal feedback indicating a little irritation and doubts could be felt more from the participants who performed low-scent tasks using concurrent think-aloud protocol. The verbal feedback given when performing the high-scent tasks mostly contained satisfaction and
general comments about the webpage and its look especially for the web pages that had a lot of features that distracted the participant before starting the main task; e.g. advertisements and captivating texts. This can mostly be seen on task four. These kinds of feedback help the facilitator to get a more genuine feedback about the given task and the website in general from the participants. It also makes the facilitator know exactly how a participant feels while using a executing a given task on a given website.

Handling two things at the same time however did not seem very easy. Participants who executed the tasks using concurrent think-aloud protocol had to execute the given task and verbalize their thoughts. This made a participant using concurrent think-aloud protocol to shift some of the attention he/she could have given fully to a task, to verbalizing his/her own thoughts. This in return resulted to deviation of the participants from optimal working and was especially evident when the participants using concurrent think-aloud protocol performed the low-scent tasks. From the fixation graphs and heat maps, more dwell time was noted mostly on the main features of the main webpage for instance menus, tabs and headings when a participant was performing a low scent task using concurrent think-aloud protocol. A participant would ‘pause’ on a feature a bit as he/she gives his/her verbal feedback while still trying to figure out the right tool on the webpage to click so as to perform the given task. Performance of high-scent tasks using concurrent think-aloud protocol however did not depict much of this and most of the participants recorded normal performance as their counterparts who did not use any think-aloud protocol while executing the tasks.

5.6 The Surveying Behavior

There was no logical way of quantifying the surveying behavior parameters (number of fixations, number of screen areas viewed and the surveying behavior duration) using the eye-tracking methodology. The amount of surveying behavior of each participant on any given task could not be measured in terms of the given surveying behavior parameters, which included number of fixations during the surveying behavior, amount of screen areas viewed during the surveying behavior and the surveying behavior duration. This was due to the fact that from the moment that a participant started a given test, it was not possible to differentiate when a participant was just looking around by glimpse of the whole webpage so as to acclimatize himself/herself with what the webpage is generally all about and when a participant was actually looking around the webpage so as to find the first step of the task. It was therefore not possible to determine whether the surveying behavior affects low-scent tasks or high-scent tasks based on any numerical parameters.
6. CONCLUSION

It was proved from this study that no difference was noted on the number of fixations recorded and no substantial difference could be noted on the number of screen areas viewed when the high scent tasks are executed using the concurrent think-aloud protocol method and while not using think-aloud protocol method. Statistically significant differences were however noted when low-scent tasks were executed using concurrent think-aloud protocol and when low scent tasks were executed without using any think-aloud protocol. Using concurrent think-aloud protocol when performing low-scent tasks resulted to higher numbers of fixations, higher amount of screen areas viewed and higher task time on the main webpage being recorded. This implied a negative impact that the concurrent think-aloud protocol has on low-scent tasks.

This study therefore implies to the usability practitioners that concurrent think-aloud protocol affects the way a given participant performs a low-scent task but has no effects on the participant when performing a high-scent task. The results that indicated higher number of fixations, higher amount of screen areas viewed and higher task time on the main page when concurrent think-aloud protocol is used when executing low scent tasks than when not using any think-aloud protocol at all should be interpreted by the usability practitioners that concurrent think-aloud protocol is not a good method to be used when carrying out usability testing on low-scent tasks since it has effects on the participants, which in return affects the low-scent tasks. The participants tend to look around more and spend more time in performing the task as compared to when not using think-aloud protocol.

Usability practitioners can however maintain the use of concurrent think-aloud protocol when testing high-scent tasks as no notable effects of this method on high-scent tasks was evident. Concurrent think-aloud protocol still remains to be a good usability testing method due to the fact that the verbal feedback that the participant gives as he/she carries out the task may exhibit any surprises, irritation, doubts, satisfaction or other feelings that may arise during the task performance process. This helps the facilitator to get a more genuine feedback on how the participant feels when performing a given task. The participant is not able to hide or fabricate his/her feelings concerning how easy/difficult it is to use a given application. The verbal output given by the participant during the task performance is more in a reactive manner. Other method could however be considered when testing low-scent tasks since the concurrent think-aloud protocol has negative effects of low-scent tasks.

As mentioned in the discussion section, surveying behavior parameters could not be quantified using the eye-tracking methodology and it was therefore difficult to determine whether surveying behavior occurred only when executing either the low scent tasks or the high scent tasks using concurrent think-aloud protocol.

In general, it has been concluded from this study that concurrent think-aloud has effects on low scent tasks and not on high scent tasks. This is due to the higher number of fixations, higher amount of screen areas viewed and longer task times that were recorded when participants performed low scent tasks using concurrent think-aloud protocol than when not
using any think-aloud protocol. Concurrent think-aloud protocol is therefore a good method of usability testing. It is highly recommended and would be more effective when used while executing high-scent tasks since it would not have major effects on high-scent tasks. Alternative usability testing methods should however be considered for low-scent tasks since concurrent think-aloud protocol has negative effects on low-scent tasks as seen from this study.
7. FUTURE WORK

As seen in this study, there was no logical way to quantify the surveying behavior in numerical terms for comparison. Perhaps it would be interesting if a future research would be carried out on ways to quantify the surveying behavior based on its parameters like surveying behavior duration, number of fixations recorded during the surveying behavior and amount of screen areas viewed during the surveying behavior.

As much as concurrent think-aloud protocol did not prove to be a favorable method of usability testing while executing low scent tasks, think-aloud protocol in general is still considered to be a good method of usability testing due to the fact that the facilitators not only get the verbal feedback from the test participants but also their emotional feedbacks. Retrospective think-aloud protocol would therefore be considered as a good alternative while executing low-scent tasks. This however can only be proved by further research on whether retrospective think-aloud protocol usability testing method would have effects on low-scent tasks.
8. REFERENCES


Experience Solutions Ltd. (den 07 January 2010). Experience Solutions; making websites easy to use. Hämtat från Experience Solutions; making websites easy to use: http://www.experiencesolutions.co.uk/what-is-usability-testing.html den 14 March 2011


9. GLOSSARY

- **Calibration** – The process of checking and determining eye-position of a participant on the screen in comparison with a pre-defined standard.

- **Concurrent** – Happening at the same time.

- **Eye-tracking** – The process of measuring the motion of the eye or the point of gaze relative to the participant’s head.

- **High scent task** – A task whose path consists of labels that strongly relates to the destination information.

- **Information scent** – The strength of a navigational path of a task.

- **Low scent task** – A task whose path consists of labels that do not relate to the destination information.

- **Participant** – An individual that executes or performs the task at hand. A representative of the target user.

- **Susceptible** – Easily affected by

- **Task** – A piece of work to be done

- **Think-aloud** – To talk one’s thoughts out

- **Usability** – The extent to which a product e.g. an application can be used by a targeted user to achieve a specific goal.

- **Usability testing** – the technique used to evaluate and determine the ease of using a given product
10. APPENDIX

8.1 List of Figures

Below are samples of some of the results got from the participants during the main study:

Figure 28: Amount of screen areas viewed on task one (with concurrent think-aloud protocol)
Figure 29: Number fixations on task 1 (with concurrent think-aloud protocol)

Figure 30: Amount of screen areas viewed on task 1 (without think-aloud protocol)
Figure 31: Number of fixations on task 1 (without think-aloud protocol)
Figure 32: Amount of screen areas viewed on task 2 (with concurrent think-aloud protocol)
Figure 33: Number of fixations on task 2 (with concurrent think-aloud protocol)
Figure 34: Amount of screen areas viewed on task 2 (without think-aloud protocol)
Figure 35: Number of fixations on task 2 (without think-aloud protocol)
Figure 36: Amount of screen areas viewed on task 3 (with concurrent think-aloud protocol)
Figure 37: Number of fixations on task 3 (with concurrent think-aloud protocol)
Figure 38: Amount of screen areas viewed on task three (without think-aloud protocol)
Figure 39: Number of fixations on task 3 (without think-aloud protocol)
Figure 40: Amount of screen areas viewed on task 4 (with concurrent think-aloud protocol)
Figure 41: Number of fixations on task four (with concurrent think-aloud protocol)
Figure 42: Amount of screen areas viewed on task four (without think-aloud protocol)
Figure 43: Number of fixations on task four (without think-aloud protocol)
Figure 44: Amount of screen areas viewed on task five (with concurrent think-aloud protocol)
Figure 45: Number of fixations on task five (with concurrent think-aloud protocol)
Figure 46: Amount of screen areas viewed on task five (without think-aloud protocol)
Figure 47: Number of fixations on task five (without think-aloud protocol)
Figure 48: Amount of screen areas viewed on task six (with concurrent think-aloud protocol)
Figure 49: Number of fixations on task six (with concurrent think-aloud protocol)

Figure 50: Amount of screen areas viewed on task six (without think-aloud protocol)
Figure 51: Number of fixations on task six (without think-aloud protocol)