Usability Testing: The Relation between Tasks and Issues

Avinash Madireddy
Human-Centered Systems (HCS)
Department of Computer and Information Science (IDA)

MASTER THESIS-TQDT30
LIU-IDA/LITH-EX-A--12/032--SE
2012-06-13
Master’s Thesis

Usability testing: the relation between tasks and issues

by

Avinash Madireddy

LIU-IDA/LITH-EX-A--12/032--SE

2012-06-13

Supervisor: Johan Åberg
Examiner: Stefan Holmlid
Upphovsrätt

Detta dokument hålls tillgängligt på Internet – eller dess framtida ersättare – från publiceringsdatum under förutsättning att inga extraordinära omständigheter uppstår.

Tillgång till dokumentet innebär tillstånd för var och en att läsa, ladda ner, skriva ut enstaka kopior för enskilt bruk och att använda det oförändrat för ickekommersiell forskning och för undervisning. Överföring av upphovsrätten vid en senare tidpunkt kan inte upphäva detta tillstånd. All annan användning av dokumentet kräver upphovsmannens medgivande. För att garantera äktheten, säkerheten och tillgängligheten finns lösningar av teknisk och administrativ art.

Upphovsmannens ideella rätt innefattar rätt att bli nämnd som upphovsman i den omfattning som god sed kräver vid användning av dokumentet på ovan beskrivna sätt samt skydd mot att dokumentet ändras eller presenteras i sådan form eller i sådant sammanhang som är kränkande för upphovsmannens litterära eller konstnärliga anseende eller egenart.

För ytterligare information om Linköping University Electronic Press se förlagets hemsida http://www.ep.liu.se/

Copyright

The publishers will keep this document online on the Internet – or its possible replacement – from the date of publication barring exceptional circumstances.

The online availability of the document implies permanent permission for anyone to read, to download, or to print out single copies for his/hers own use and to use it unchanged for non-commercial research and educational purpose. Subsequent transfers of copyright cannot revoke this permission. All other uses of the document are conditional upon the consent of the copyright owner. The publisher has taken technical and administrative measures to assure authenticity, security and accessibility.

According to intellectual property law the author has the right to be mentioned when his/her work is accessed as described above and to be protected against infringement.

For additional information about the Linköping University Electronic Press and its procedures for publication and for assurance of document integrity, please refer to its www home page: http://www.ep.liu.se/.

© Avinash Madireddy.
Abstract

The Usability of a website can be evaluated using various methods. One of the methods is usability testing, which is widely used and employed by the usability tester due to its low cost and user friendliness. The results of usability testing can be affected positively or negatively by several factors such as evaluator’s role, number of users, test environment, tasks, usability problem report, usability measures, and other factors. For more than a decade, the number of users plays a key role in usability testing. Previously done study [12], has shown that correlation exists between the number of user tasks and the number of issues found.

As an extension and follow-up of the recent studies, the current work was carried out on task design, task number, and task coverage. Two types of task designs were proposed namely, guided tasks and unguided tasks. Considering the task as a key factor, the remaining factors were also considered while employing the usability test.

In this study, two websites were chosen for performing the usability test, namely, www.mq.edu.au and www.siu.edu. The usability test was carried out with a total of 40 users. The collected usability data was analyzed with respect to comparing means, correlation between different tasks designs, and how severe the problems found were. The evaluation results showed that unguided task had a greater impact on usability testing results than a guided task had. Finally, this work concluded by recommending the usability practitioners to employ both the tasks designs in order to increase the usability of any given website.

Key words

Usability, usability testing, guided tasks, unguided tasks, quantitative analysis, and qualitative analysis, t-test.
Acknowledgements

This study was carried out at Human-Centered Systems division at the Department of Computer and Information Science at Linköping University from September 2011 to March 2012.

First and foremost, I would like to convey my sincere gratitude to my supervisor, Dr. Johan Åberg, Linköping University, for the fruitful discussions that were done throughout the study which lead me to gain in-depth knowledge in the area of usability.

My sincere thanks further goes to my friends and to all my classmates in MSc Computer Systems for their encouragement throughout my study period.

Finally I wish to thank my parents, all my family members for giving me their moral support.
# Table of Contents

1. INTRODUCTION .......................................................................................................................... 1  
   1.1 Background of Study ................................................................................................................. 1  
   1.2 Motivation ................................................................................................................................. 1  
   1.3 Goals/Research questions ......................................................................................................... 1  
   1.4 Overview .................................................................................................................................. 1  
   1.5 Disposition ............................................................................................................................... 2  

2. LITERATURE REVIEW .................................................................................................................. 4  
   2.1 What is Usability? ...................................................................................................................... 4  
   2.2 Usability Evaluation Methods .................................................................................................. 5  
      2.2.1 Usability Testing ................................................................................................................ 5  
      2.2.2 Usability Inspection ........................................................................................................... 6  
      2.2.3 Usability Inquiry ................................................................................................................ 7  
   2.3 What is Usability Testing? ......................................................................................................... 9  
      2.3.1 Factors affecting Usability Testing and its results ............................................................ 10  
   2.4 Goals of Usability Testing ....................................................................................................... 12  
   2.5 Analyzing and Reporting Metrics for Usability Issues ............................................................ 12  
      2.5.1 Frequency of Unique Issues ............................................................................................ 13  
      2.5.2 Issues by Category ........................................................................................................... 13  
   2.6 Issues in Task Design for Usability Testing ............................................................................. 13  
      2.6.1 Conclusion from the studies ............................................................................................ 14  

3. METHOD ....................................................................................................................................... 15  
   3.1 Type of study ............................................................................................................................ 15  
   3.2 Evaluator’s Role ...................................................................................................................... 15  
      3.2.1 Characteristics of an Evaluator ........................................................................................ 15  
   3.3 Participants .............................................................................................................................. 16  
      3.3.1 Participants for the pilot test ........................................................................................... 16  
      3.3.2 Participants for the main test .......................................................................................... 17  
   3.4 Test Duration ........................................................................................................................... 17  
   3.5 Targeted Websites and Tasks .................................................................................................. 17  
      3.5.1 Website Selection .............................................................................................................. 17  
      3.5.2 Task Selection ................................................................................................................... 19  
   3.6 Usability Issues ....................................................................................................................... 20  
      3.6.1 Issues by Category ............................................................................................................ 20  
      3.6.2 Issues by Tasks .................................................................................................................. 20
3.7 Comparing Means ................................................................. 20
3.8 Correlation: Relation between variables ........................................ 21
3.9 Severity Ratings .................................................................. 22
    3.9.1 Issue Severity Rating ..................................................... 22
3.10 Methodology .................................................................... 22
3.11 Test Procedure/Process .......................................................... 23
    3.11.1 Pilot test procedure ................................................... 23
    3.11.2 Main test procedure ................................................... 24
3.12 Dependent Measures ............................................................ 24
4. RESULTS ............................................................................. 25
    4.1 Overview .................................................................... 25
    4.2 Results Presentation ......................................................... 25
        4.2.1 Guided task ............................................................ 25
        4.2.2 Unguided task ......................................................... 28
        4.2.3 Comparing Means ................................................... 31
        4.2.4 Correlation .............................................................. 34
        4.2.5 Issue Severity Rating .............................................. 36
5. DISCUSSION ...................................................................... 38
    5.1 Task Completion .............................................................. 38
    5.2 Accuracy of the Tests in this study ....................................... 38
    5.3 Discussion of the Statistical Results .................................. 38
6. CONCLUSION .................................................................... 41
7. FUTUREWORK .................................................................... 42
8. REFERENCES ....................................................................... 43
9. GLOSSARY ........................................................................ 45
10. APPENDIX ........................................................................ 46
    10.1 Identified Issues ............................................................... 46
    10.2 Usability Observation Form .............................................. 49
        Introduction to Usability testing ....................................... 49
        Demographic questions ................................................... 50
        Usability testing for guided tasks ..................................... 51
        Usability testing for unguided tasks. ................................. 56
List of Figures

Figure 1: Current Issues with Usability testing ................................................................. 10
Figure 2: Quantitative model for usability ................................................................. 12
Figure 3: Adopted from www.siu.edu website ..................................................................... 18
Figure 4: Adopted from www.mq.edu.au website .............................................................. 18
Figure 5: Representation of the methodology used in this study .................................. 23
Figure 6: Frequency of participants encountered for each issue category – www.mq.edu.au 26
Figure 7: Frequency of participants encountered for each issue category – www.siu.edu........ 27
Figure 8: Percentage of total number of participants encountered each type of issues category for the tasks – www.mq.edu.au ................................................................................ 27
Figure 9: Percentage of total number of participants encountered each type of issues category for the tasks – www.siu.edu ................................................................................ 28
Figure 10: Frequency of participants encountered for each issue category – www.mq.edu.au 29
Figure 11: Frequency of participants encountered for each issue category – www.siu.edu........ 29
Figure 12: Percentage of total number of participants encountered each type of issues category for the tasks – www.mq.edu.au ................................................................. 30
Figure 13: Percentage of total number of participants encountered each type of issues category for the tasks – www.siu.edu ................................................................. 30
Figure 14: Pie-chart representation for the data represented in table 11 – www.mq.edu.au 33
Figure 15: Pie-chart representation for the data represented in table 12 – www.siu.edu ........ 34
Figure 16: Graphical representation of relationship between issues and user size – www.mq.edu.au . 35
Figure 17: Graphical representation of relationship between issues and user size – www.siu.edu........ 35

List of Tables

Table 1: List of tasks selected to perform usability testing .................................................. 19
Table 2: Severity rating based on combination of factors (impact on users experience, frequency of users experiencing a problem) – [21] ........................................................................ 22
Table 3: Overall Result for the guided task – www.mq.edu.au ........................................ 25
Table 4: Overall result for the guided task – www.siu.edu .................................................. 26
Table 5: Overall Result for unguided task – www.mq.edu.au ........................................ 28
Table 6: Overall Result for unguided task – www.siu.edu .................................................. 28
Table 7: Represents the number of unique issues encountered by each user – www.mq.edu.au ........ 31
Table 8: t-test results for the data represented in table 7 – www.mq.edu.au ..................... 31
Table 9: Represents the number of unique issues encountered by each user – www.siu.edu.......... 32
Table 10: t-test results for the data represented in table 9 – www.siu.edu ........................... 32
Table 11: Represents the number of issues comes under guided, unguided and shared groups – www.mq.edu.au.................................................................................................................. 33

Table 12: Represents the number of issues comes under guided, unguided and shared groups – www.siu.edu.................................................................................................................. 33

Table 13: The number of issues shared among Low, Medium, and High level severity between guided, unguided, and shared groups – www.mq.edu.au .................................................................................................................. 36

Table 14: The number of issues shared among Low, Medium, and High level severity between guided, unguided, and shared groups – www.siu.edu .................................................................................................................. 36

Table 15: An overview of chi-square test results for comparing the task designs in www.mq.edu.au and www.siu.edu websites. .................................................................................................................. 36

Table 16: Issues that were identified during usability evaluation of the both websites ...................... 48
1. INTRODUCTION

1.1 Background of Study

Having poor usability of a website makes the users to opt for an alternate website. The usability of a website can be increased by identifying more number of issues and altering the website with respect to the recognized issues. Evaluating the usability of a website can be done in any one of the methods stated in section 2.2. Usability testing is widely used and employed by the usability practitioners, due to its low cost and user friendliness. Results of a usability test will be influenced by several factors, which are number of users, evaluator’s role, test environment, usability problem report, usability measure, and other factors [3]. For more than a period, number of users plays a key role in evaluating usability of a website.

The research work done in [14], stated that focus must be from number of users to the tasks and this study revealed correlation exists between the number of user tasks and number of issues encountered. As an extension and follow up of this study, the current work concentrates on task design, task number, and task coverage.

1.2 Motivation

Usability testing is influenced positively and negatively by several factors [3]. Study done in [14], revealed correlation exists between the number of tasks and the number of issues encountered by the users. [1, 4] stated the design of task and its impact on usability testing. Selection of tasks is very critical and will be affected by several factors in usability testing [24]. Selecting the wrong set of tasks will affect the usability of a website.

Why task? Undertaking tasks plays an important role in usability testing, and is widely used technique for ensuring usability of a website. They have a tremendous impact on issues which are identified. The tasks basically determine what areas of the product are exercised and the ways in which they are exercised. Particularly with a complex product, this can have a major impact on what issues are uncovered [21]. Task design will have a serious impact on the usability test [1]. All the above considerations were taken into account and work is carried out on task design, task number, and task coverage.

1.3 Goals/Research questions

The main objective of this thesis is to contribute with knowledge about task design and its effect on the usability test results. The design of the tasks was done in to guided and unguided tasks. The thesis answers the following primary research question:

- What effects do guided and unguided tasks designs, have on the usability test results and relation between them? The effect will be looked in terms of the number of unique issues, issues by category, severity degree, and discovery rate per user.

1.4 Overview

The design of the tasks was classified in to two categories i.e., guided and unguided tasks. Both these designs were independent to each other. In common, they will have same set of tasks and same number of participants but differ with respect to two different groups of
participants. In guided tasks, users will perform the tasks with guidelines. Following example gives overview of guided task

*Task:* “find the departments at Linkoping University”

*Guidelines:*

- Browse the homepage (www.liu.se)
- Click “Department” on left hand side of the page
- Click on any one of the department links
- Come back to homepage

In Unguided tasks, users were left to behave as they would do in reality. For the above task, unguided task is as follows:

- User will be in Google home page
- Browse for Linkoping University home page
- Know more about the website and find the departments at Linkoping University

1.5 Disposition

*Chapter 1: Introduction*

This part gives the overview of the thesis goal and the area of study it deals with. It also gives a glance view of the problems regarding thesis goal and the methodology used to fulfill the research goal.

*Chapter 2: Literature review*

This part explains in detail regarding the usability, usability evaluation methods and factors affecting usability testing. In addition, discussions carried out in the related articles, and books.

*Chapter 3: Method*

This part deals with goals of the study and concepts used during this study so as to get the necessary results.

*Chapter 4: Planning the test*

This part explains how the test procedure carried out during this study for collecting the usability data.

*Chapter 5: Results*

This part presents the results drawn by analyzing the data collected during the usability testing.

*Chapter 6: Discussions*

Analyzing and discussing the results represented in chapter 5 and working towards the goal of the thesis work.
Chapter 7: Conclusion

This part explains what can be concluded from the analyzed data and relating the results to the previous studies carried out in same area.
2. LITERATURE REVIEW

2.1 What is Usability?

Usability generally refers to the method for making products and systems easier to use, and matching them more closely to the user needs and requirements. The definitions which have been used in the previous studies were derived from a number of views and following three views were related for measuring the usability [5]:

- The *product-oriented view*, measuring of usability was done in terms of ergonomic attributes of the product.
- The *user-oriented view*, measuring of usability was done in terms of the mental effort and attitude of the user.
- The *user performance view*, that usability can be measured by examining how the user interacts with the product, with particular emphasis on either:
  - Ease-of-use: how easy will be the product to use, or
  - Acceptability: whether the product will be used in the real world.

ISO standard for software qualities (ISO 1991b) was product and user-oriented view:

“A set of attributes of software which bear on the effort needed for use and on the individual assessments of such use...”

The Usability Professional Association (UPA) definition focuses more on the product development process:

“Usability is an approach to product development that incorporates direct user feedback throughout the development cycle in order to reduce costs and create products and tools that meet user needs.”

ISO 9241-11 identifies three aspect of usability, defining it as:

“The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use.”

The metrics such as user’s effectiveness, efficiency, and satisfaction were considered while evaluating the usability of a website. These metrics were stated as follows:

- **Effectiveness**: A user’s ability to successfully use a website to find information and accomplish tasks.
- **Efficiency**: The amount of effort required by the user to complete the task.
- **Satisfaction**: The degree to which the user was happy with his/her experience while performing the task.

In this work, the usability of a website was measured with respect to the user’s performance view i.e., how effectively the specified users interacts with a website to accomplish tasks.
2.2 Usability Evaluation Methods

Following sections gives a brief introduction of the methods used in evaluating the usability of a website. [13]:

2.2.1 Usability Testing

Usability testing is a technique for ensuring that the intended users of a system can carry out the proposed tasks efficiently, effectively and satisfactorily. Usability testing is carried out at pre-release of a product so that any significant issues identified can be addressed. [7] Provides the following methods which are used at different stages of the product design in usability testing:

*Coaching Method*

This method is applicable at design, code, test, and deployment stages. While employing this method, system expert sits along with the users to act as a coach. A user is allowed to ask the system related questions while performing the selected tasks and coach answers to those questions. The observer evaluates interaction between the user and the coach. The purpose of this technique is to discover the information needs of users in order to provide better training and documentation, as well as possibly redesign the interface to avoid the need for the questions, [7]. It covers the issues with respect to effectiveness and satisfaction.

*Co-discovery Learning*

This method is applicable at design, code, test, and deployment stage. Tasks that were selected will be performed by two users helping each other working towards a common goal. During the test observer encourages the users to explain their views about the product design while performing the tasks. It covers the issues with respect to effectiveness and satisfaction.

*Performance Measurement*

This technique is applicable at design, code, test, and deployment stage. Through this technique quantitative data is obtained about the test participant’s performance when they perform the tasks during usability test. The collected quantitative data is most useful in doing comparative testing, or testing against predefined benchmarks. Combining this method with a questionnaire or retrospective testing lets the usability tester to know both quantitative and qualitative data. It covers the issues with respect to effectiveness and efficiency.

*Question-asking Protocol*

This method is applicable at design, code, test, and deployment stage. While performing the tasks observer prompts participant’s with questions. It covers the issues with respect to effectiveness and satisfaction.

*Remote Testing*

This method is applicable at design, code, test, and deployment stage. In remote testing, the usability tester cannot observe the testing process directly. Remote testing will be carried out in more than one way. In one method, usability testing can be carried out at same-time but in
different places, where the tester can observe the test user’s system screen through a network and listen to tester voice through speaker telephone. Another is different-time different-place testing, where the user’s test session is started by logging in to a test session through software and the data will be stored in a database. It covers the issues with respect to effectiveness, efficiency, and satisfaction.

Retrospective Testing
This method is applicable at design, code, test, and deployment stage. While implementing this technique, user’s interaction with the computer needs to be recorded and replayed. After completing the test session recorded videotape will be observed by the usability expert and the participants. It covers the issues with respect to effectiveness, efficiency, and satisfaction.

Shadowing Method
This method is applicable at design, code, test, and deployment stage. It covers the issues with respect to effectiveness and efficiency. During the usability test, tester has an expert user sits next to him/her and explain the test user’s behavior to the tester. This method is appropriate for the test user to think aloud or talk to the tester while working on the tasks.

Teaching Method
This method is applicable at design, code, test, and deployment stage. During the test, a user will be trained by interacting with the system to gain knowledge/experience with the product. After user finishing the training, tester asks the trained user to teach the new novice participant and observes the interaction between trained user and new novice user. It covers the issues with respect to effectiveness and satisfaction.

Thinking Aloud Protocol
This method is applicable at design, code, test, and deployment stage. During this technique user is asked to vocalize his/her thoughts, feelings, and opinions while interacting with the product. It covers the issues with respect to effectiveness and satisfaction.

2.2.2 Usability Inspection
In Usability inspection approach, usability specialists, users, developers, professionals examine the usability related aspects of a user interface. It is inexpensive and quick. The following one of the methods is in use for conducting usability inspection, [18]:

Cognitive Walkthroughs
It evaluates the design of a product on how well it supports the system users in learning tasks, [23]. This type of testing is conducted by an expert in cognitive psychology and walks through the design for identifying potential problems using psychological principles.

It involves a group of evaluators inspecting a user interface by going through a set of tasks and evaluates its understandability and ease of learning. This technique is best used in the design stage of development.
**Feature Inspection**

This technique focuses on the feature set of a product. The inspectors were usually given use cases with the end result to be obtained from the use of the product. Each feature is analyzed for its availability, understandability, and other aspects of the usability. It covers the issues with respect to effectiveness.

**Heuristic Evaluation**

This method involves having a small group of usability experts for evaluating a user interface using a set of guidelines and noting the severity of each usability problem and where it exists, [18]. This is a method for quick, cheap, and easy evaluation of a user interface design. Goal of the heuristic evaluation is to find the usability problems in the design so that they can be attended to as part of an iterative design process. Heuristic evaluation involves a small set of evaluators for examining the interface of a design. Each evaluator evaluates the interface of a design and at end of their findings communication among the evaluator’s will be made to list the identified problems related to the design.

**Pluralistic Walkthrough**

At the design stage, when paper prototype is available, a group of users, developers, and human factors engineers meet together to step through a set of tasks, discussing and evaluating the usability of a system, [23]. Group walkthroughs have the advantage of providing a diverse range of skills and perspectives to bear on usability problems. As with any inspection, more people looking for problems, higher will be the probability of finding problems. Also, the interaction between the team during the walkthrough helps to resolve usability issues faster.

**Perspective-based Inspection**

The goal is to understand the feasibility, effectiveness, and scope of perspective-based usability inspection and also by inspecting from one particular perspective at a time. In this technique inspection will be done in three perspectives. Each inspection focuses on one of the three defined perspectives: novices use, expert use, and error handling.

**2.2.3 Usability Inquiry**

By using this method, usability evaluators obtain information about the user’s review i.e., what were the likes, dislikes, needs, and understanding of the system by communicating and observing the user’s interaction with the system in real work, or letting them answer to the questions verbally or in written form. Following are the commonly used inquiry methods, [19]:

**Field Observation**

This method is applicable at test, and deployment. Human factors engineers go to representative user’s workplace and observe there work to understand how the users were using the system to accomplish their tasks and the kind of mental model the users have about the system. It covers the issues related to effectiveness and satisfaction.
Focus Groups

This is a data collecting technique where about 6 to 9 users were brought together to discuss the usability issues relating to the system/website. A human factors engineer plays the role of a moderator, who needs to prepare the list of issues to be discussed beforehand and seek to gather the needed information from the discussion. This can capture spontaneous user reactions and ideas that evolve in the dynamic group process. It covers the issues related to effectiveness and satisfaction and this method is employed during test and development.

Interviews

In this technique, human factors engineers formulate questions about the product based on the kind of usability issues. Then they interview representative users to ask them these questions in order to gather information desired. It is good at obtaining detailed information as well as information that can only be obtained from the interactive process between the interviewer and the user.

In an evaluation, the interviewer reads the questions to the user, replies verbally, and the interviewer records those responses. The methods of interviewing include unguided interviewing and guided interviewing.

Unguided interviewing methods were used during the earlier stages of usability evaluation. The objective of the investigator at this stage is to gather as much information as possible concerning the user's experience. The interviewer does not have a well-defined agenda and is not concerned with any specific aspects of the system. The primary objective is to obtain information on the procedures adopted by the users and on their expectations of the system.

Guided interviewing has a specific, predetermined agenda which includes specific questions to help guide and direct the interview. Guided interviewing will be more of an interrogation than unguided interviewing, which is closer to a conversation.

Proactive Field Study

Before designing a system, human factors engineers go to representative user's workplace and talk to them, observe there work, and ask them the questions to understand the user characteristics, the work flow, the system features, working environment etc. This technique should be used during the requirement or early design stage of the software development. This should be the first step of usability work for a project.

Questionnaires

Questionnaires have long been used to evaluate the user interfaces. Questionnaires have also been used in electronic form. For a handful of questionnaires specifically designed to assess aspects of the usability, the validity and/or reliability have been established, including some of them discussed in [21].
2.3 What is Usability Testing?

Usability testing is a technique used to improve the user interface of a product. It reveals problems with in the design of website and its functionality. The purpose of usability testing is to:

- Focus on specific aspects of the design.
- Evaluate a design based on predetermined usability criteria.
- Uncover new and previously covered problems.

The most appropriate condition for carrying out usability testing is before the pre-release of a product to address the identified issues. Usability testing is carried out at any time or at successive intervals during the development cycle of a product to make the product user friendly with its interface design. It is preferred to carry out usability testing early in the design process for finding major flaws, or whenever changes were being made. Usability testing is performed among the group of users by accomplishing the given tasks on the suggested website. During the usability test, a moderator gathers the issues encountered by the users and feedback/suggestions to improve the usability of a website. Attributes of usability testing are:

- The goal is to improve a product/website
- Participants were real users and perform real tasks
- Participants were formally observed
- The data is analyzed
- Recommendations for improvements were made

After performing the usability test, collected data can be analyzed in two ways: Quantitative and Qualitative analysis. Quantitative analysis represents the usability data numerically i.e., the number of issues faced, the number of tasks completed successfully, the number of clicks needed to perform the task, and time spent per task to complete. This quantitative data is analyzed statistically to compare the iterative designs with respect to the number of problems found. In the current study, quantitative data is collected i.e., finding the issues by observing the effectiveness of the user interaction with a website for each task design and comparing the task designs with respect to the identified issues for finding the impact of task design on usability testing.

Qualitative analysis helps to know about the users and what they were trying to accomplish. This is about the comments made by the usability test participants to some of the questions like how frequently they were using the product, was the product efficient to use, and the satisfaction rate.
2.3.1 Factors affecting Usability Testing and its results

The following are the factors that influence usability testing and its results both negatively and positively [3]. Figure 2 describes the list of factors affecting usability testing.

![Diagram of Usability Testing Factors]

**Figure 1: Current Issues with Usability testing**

2.3.1.1 Evaluator’s Role

The role of an evaluator was a critical issue in usability. In past research studies, several discussions and arguments proved that problem detection differs noticeably. Evaluators were also criticized for a lack of methodical analysis just after the test has been finished. Three strategies were also proposed to solve the evaluator’s effect [11]. Becoming a great test moderator takes four things [15]

- Understanding the basics of usability testing
- Interacting well with test participants
- Ability to establish and maintain rapport with participants
- Lots of practice

A set of rules and the things to do and not to do while interacting with test participants, [15] makes the evaluator to get the correct usability data.

2.3.1.2 Users

In usability study, users have a major impact on usability testing results. It was very important for recruiting the correct set of users for a usability study. Recruiting the users will be done based on some characteristics such as age groups, gender groups, experience in usability testing etc. The next step after selecting a relevant group of users was to set the sample size of users required. Knowing the sample size of users was discussed in a number of research papers. This was a very crucial step since size of the user’s effect in finding the usability test results. It was suggested that 5 users were enough to discover 85% of the usability problems [16]. It was found that only 35% of usability problems were identified with 5 users [14]. Based up on the type of analysis 5 and 20 users were suggested for qualitative or quantitative data respectively [17].
2.3.1.3 Tasks

Usability testing tasks refer to what users do in order to achieve a goal, but even though they are user-driven, they are an important issue and can heavily influence usability evaluation. The selection of tasks was a critical activity in usability testing as some of the tasks will have a clear end state, some are open-ended, and some are self-generated by the users. What became clear in the study was that the selection of tasks had a tremendous impact on the issues that were identified [21]. Selection of tasks helps to know which areas of the website need to be exercised and redesigned. In particular, while testing a large product it will have a major impact on identifying uncovered issues. It was suggested of moving the focus from number users to tasks based on their research carried out in [15]. Further research on task design proved that it shows a severe impact on usability testing results [2].

2.3.1.4 Usability Problem Report

At end of the usability test, tester should generate a usability problem report which effectively helps the designers and developers to make their decision with regards to the redesign stage. Reporting the list of problems may not effectively help them to solve the problems. But the detailed description along with how to deal and treat with certain problems will impact on usability test results.

2.3.1.5 Test Environment

Usability testing will be performed in a Laboratory. The physical environment of the laboratory impacts on the testing results. It may interact directly with the user or indirect interaction like video recording, video call. Conducting Usability testing in natural places shows a major impact on usability testing.

2.3.1.6 Usability Measures and Prioritizing Problems

Prior to conducting a usability test, testers should be aware of what to test and measure. There are three major ISO standards for measuring usability, which are efficiency, effectiveness and user satisfaction. This model was criticized because of it’s too abstractedness. Several models of measuring usability have been proposed and criticized. Usability measures individually such as effectiveness, efficiency, and satisfaction [11]:

- Effectiveness measures can be worked out through measuring binary task completion, accuracy, completeness, quality of outcome, and others factors
- Efficiency can be measured through measuring input rates, task completion time, mental effort, learning, use frequency
- Satisfaction measured includes standard questionnaires, preferences, satisfaction with the interface, and others.

A method was proposed in [20] for standardizing usability metrics into a single usability metric and proposed a quantitative model for usability as shown in figure 3.
2.3.1.7 Other Usability Issues
Artifact will be used for interaction between usability evaluation and the design stage. It may be a prototype or a product that was evaluated which had a huge impact on finding test results. The type of interaction may be a paper prototype, functional or semi functional prototype, or product system. The prototypes may misinterpret real system functionality. There are four challenges that can improve usability evaluation and the design stage [12]. The first challenge was the type of prototype or a running system. The Second challenge was the insufficient effort that was usually allocated to choosing tasks. The Third challenge was the usability problems report. The Fourth challenge was usability problems reports neither recommends nor suggests priority of problems severity.

2.4 Goals of Usability Testing
- To identify the usability problems.
- To make recommendations for modifying the product.
- To find out how satisfying the product was.
- Establish benchmarks for future comparisons.

2.5 Analyzing and Reporting Metrics for Usability Issues
After performing the usability test, all the identified issues need to be analyzed to derive some metrics related to usability issues. Exactly how to do this will largely depend on the type of usability questions the usability tester has. Three general questions can be answered by looking at metrics related to usability issues: [21]
- How is the overall usability of the product? This is helpful to get an overall sense of how the product did.
- Is the usability improving with the each design iteration? Focus on this question when we need to know how the usability is changing with each new design iteration.
- Where should I focus my efforts to improve the design? The answer to this question is useful when you need to decide where to focus the resources.
All of the analyses that were explained can be done with or without severity ratings. Severity ratings simply add a way to filter the issues. Sometimes it’s helpful to focus on the high-severity issues. Other times it might make more sense to treat all the usability issues equally.

### 2.5.1 Frequency of Unique Issues

The most common way to measure usability issues is to simply count the unique issues. In any design process, analyzing the frequency of unique issues is most useful in an iterative design process to know how the usability is changing with each iteration or between two designs. [21]

In this study, guided tasks and unguided tasks were compared with respect to the number of unique issues to know which task design reports more number of issues.

### 2.5.2 Issues by Category

Sometimes it’s helpful to know where to focus design improvements from a tactical perspective. Perhaps we may feel that only certain areas of the product are causing the most usability issues, such as navigation, content, terminology, and so forth [21]. In this analysis, examining each identified issue and the categorizing it into a type of issue. Then look at the frequencies of issues that fall into each category. Categorizing the issues can be done in many different ways and need to make sense to you and your audience, and use a limited number of categories, typically three to eight [21].

In previous usability research on website interface design, six usability problems categories were identified, namely Layout, Terminology, Feedback, Comprehension, Data Entry, Navigation [8, 22].

### 2.6 Issues in Task Design for Usability Testing

**Study 1**

A study carried out by [14], initiates the focus of increasing the usability test performance by shifting the focus from participants to tasks coverage. It discusses and compares the results by performing usability evaluation done among 9 teams. It also criticizes about the “magic number 5”. It concludes that task coverage will impact more than the participants on usability testing and suggested further research was needed on task coverage.

Results from this study gave basement and motivation to current work more on task coverage and some knowledge about the “magic number 5”.

**Study 2**

The studies carried out by [1, 2, 3, 4], mainly explain the factors influencing usability testing and impact of task design on the usability testing. All these papers explain the current issues of usability testing and they proposed task designs. The results from these studies indicate positive results regarding the impacts of task design on usability testing.

Design of task and results of this study was considered to carry out this thesis work.
**Study 3:**
The studies carried out by [15, 16], mainly explains the necessary reasons regarding why to test with 5 users and also the number of users needed for quantitative studies. User size of 20 was recommended to carry out the testing for quantitative study and 5 users for qualitative study.

**2.6.1 Conclusion from the studies**
The following points can be drawn from studies:

**Study 1:**
- Number of tasks had shown major impact on usability test performance than number of users and laid a basement for further research on task design, task coverage, and number of tasks.

**Study 2:**
- Proposed two types of tasks namely, guided and unguided. It showed how much a task design can influence usability testing results.
- Suggested task design was employed in this thesis work and current study was carried out regarding the number and selection criteria of tasks along with the task coverage.

**Study 3:**
- It was recommended of using 5 users for usability testing as it will be enough to detect more number of issues in a website. Still, this endorsement has been appraised by other researchers.
- Depending on the type of the usability date i.e., qualitative and quantitative this study suggested to test with 5 and 20 users respectively.
- The type of data collected during this thesis work was quantitative data and user size of 20 was set as a benchmark for current usability testing.
3. METHOD

3.1 Type of study

A well planned usability study saves time and money. The study was a quantitative usability testing to evaluate the effectiveness of a website i.e., identifying the number of issues encountered by the users in a website while performing a set of tasks. In this study, task design was categorized into guided and unguided task. In the guided task, the users had to follow a set of guidelines to accomplish the task in a website. Guidelines will help the users to follow the correct path for accomplishing the task. In the unguided task, a user will behave like the real target user of a website and accomplish the tasks with their previous knowledge or learning by themselves.

In the current study, an observation form was used by the moderator to note down the identified issues while user interacts with the website through a set of tasks. Observation form contains overview of the procedure and instructions need to be followed by the usability tester. It also contains the set of tasks with which tester interacts with a website. Figure 5 in Section 3.10 gives an overview of the methodology employed in the current usability testing for collecting the usability data. The independent variables in this study were testing with guided and unguided tasks. The dependent variables were the number of participants and the set of tasks selected to perform the usability test.

3.2 Evaluator’s Role

In the usability test, experience of an evaluator plays a major role while collecting the usability data. In usability testing of the current study, the author of this thesis work acted as an evaluator. Goal of the evaluator was to observe how effectively user interacts with a website and makes the user’s to express their view and issues with in a website. Usability testing was carried out only after the evaluator fulfills the following characteristics [15]:

- Having a good knowledge about the website.
- Goals and the type of date needed to be collected during usability testing.
- Prepare the test script.
- Conduct a pilot test.
- Refine the test script and materials if necessary.

3.2.1 Characteristics of an Evaluator

During the usability test, the evaluator (author of this work) acted as a neutral observer with following characteristics [15]:

- Let’s the participants speak
- Asks unbiased questions
- Neither encourages nor discourages by emotional comments
- Avoid defending the product design.
3.3 Participants

In the current usability test, two tests were performed, namely, the pilot test and the main test. The pilot test was followed by the main test. The participants for the usability test were recruited if they meet the following pre-requisites:

- Sufficient browsing knowledge
- Previous knowledge of surfing the similar website used for testing.
- Willing to participate
- Understanding and judging of tasks
- Responsiveness: must give appropriate verbal feedback.

The following steps were followed for recruiting the each participant for the current usability test:

- Each participant was approached and explained about the part of the work and the role he/she has to perform in the usability test.
- Recruiting a participant for the usability test was done if he/she meets the minimum requirements as stated in above paragraph.
- Total time required for finishing the usability test was also noted to the participant.
- Without disturbing the participant’s work usability testing was carried out either:
  - Immediately or,
  - By taking an appointment in their free time.

It was very important to note that the four participants, who participated during the pilot test, were different from the forty participants who participated during the main test. The current usability test was performed on the selected two websites, namely, www.mq.edu.au and www.siu.edu belonging to a university and the usability test participants were recruited among the students at Linköping University and all were between the ages 20 to 30.

The selected two websites falls under the same category group i.e., university website and in real world most of the users for these websites were students. So in the current usability test, an average Linköping University student can perform well by relating with his/her previous knowledge and acts like real target users. During the test, the participants were made comfortable to perform the usability test in presence of the evaluator.

3.3.1 Participants for the pilot test

The main aim of pilot testing was to find out if the proposed methodology works efficiently for gathering the accurate usability data and to identify the problems encountered with the questionnaire. A total of four participants were recruited among the students at Linkoping University and they were categorized in to two participants for each task design group, namely, guided and unguided. After finishing the pilot test, problems and improvements identified in the usability test were rectified before proceeding to the main test.
3.3.2 Participants for the main test

The total number of test participants was forty. They were recruited as the usability test participants, if they meet the minimum qualifications stated in the above section. They must behave like the real target users to perform the needed tasks on a website. Understanding of tasks and browsing experience were the key factors for a test participant in the usability test.

There were two groups of participants, namely, guided group and unguided group each consisting of twenty participants. One group of participants comes under guided group where they perform the tasks with the given guide lines and another group of participants comes under unguided group where they will perform the tasks similar to a real user i.e., without any assistance. Each participant was tested individually on a computer with same internet speed and browser. In this study, the number of participants played a major role in getting reliable results, upon which significant decision was made.

3.4 Test Duration

Each participant spent at most 40 minutes for all the tasks he/she was supposed to carry out. The evaluator carried out the testing with 4 to 5 participants per day.

3.5 Targeted Websites and Tasks

3.5.1 Website Selection

After careful study the following two websites were selected for the usability test:

- www.siu.edu
- www.mq.edu.au

Both the web-sites belong to the same category and reasons for selecting the websites were as follows:

- Both are dynamic websites.
- Meets the research requirement as it has various functions, processes, and features.
- Users (students from Linköping University) can be easily recruited for this usability evaluation.
- Recruited users can act like real target users as they were familiar with similar kind of websites.
- They are unfamiliar to the test users.
Figure 3: Adopted from www.siu.edu website

Figure 4: Adopted from www.mq.edu.au website
3.5.2 Task Selection

Structuring of tasks plays an important role in analyzing the usability data. Since some of the certain tasks reveal quantitative usability data such as number of issues identified to finish a certain task. Other tasks can expose the qualitative data such as user’s alleged ease-of-use for a given web page. In the current study, quantitative data was collected i.e., the effectiveness of a website was measured. Based on these points, the selection of tasks for the usability test was done and the chosen tasks must be:

- Reasonable and doable
- Specific and focused

The following criteria were considered while selecting tasks:

- Frequency: How often user will complete the task on a website
- Importance: The most critical tasks
- Seriousness of errors: If the tasks were done incorrectly, it was a serious problem
- Uncertainty: Having questions about the tasks.

In general, total time needed to complete a usability test on a user should be considered. If the total time was too long, many of the test users will not show interest to participate in a usability test. Considering the total time as major factor, the current usability testing was designed to complete within 30-40 minutes. A total of six tasks were selected in this study to carry out the usability testing and on an average each task will take 3-6 minutes to complete.

<table>
<thead>
<tr>
<th><a href="http://www.siu.edu">www.siu.edu</a></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Task1</td>
<td>Find the results for recently completed women's basketball events at Saluki Athletics.</td>
</tr>
<tr>
<td>Task2</td>
<td>Find about the &quot;University honors program&quot; offers at this university.</td>
</tr>
<tr>
<td>Task3</td>
<td>Find the events at <a href="http://www.siu.edu">www.siu.edu</a> for the visitors in this month.</td>
</tr>
<tr>
<td>Task4</td>
<td>Find the tuition fee for graduate studies.</td>
</tr>
<tr>
<td>Task5</td>
<td>How many libraries does this university provide to students for accessing the books?</td>
</tr>
<tr>
<td>Task6</td>
<td>Find the graduate programs offered in Art &amp; Design department.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><a href="http://www.mq.edu.au">www.mq.edu.au</a></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Task1</td>
<td>Find about the news room for the media at this university.</td>
</tr>
<tr>
<td>Task2</td>
<td>Macquarie University offers to book a slot for IELTS exam, find out the amount to be paid for this exam.</td>
</tr>
<tr>
<td>Task3</td>
<td>Find the mail-id of the person to enquire about the postgraduate scholarships.</td>
</tr>
<tr>
<td>Task4</td>
<td>Find about the student events today.</td>
</tr>
<tr>
<td>Task5</td>
<td>Find about the disciplines that are offered in research of earth sciences.</td>
</tr>
<tr>
<td>Task6</td>
<td>Find the upcoming carrier service events for students to develop their skills.</td>
</tr>
</tbody>
</table>

| **Table 1**: List of tasks selected to perform usability testing |
Table 1, represents the selection of tasks for both the websites and these set of tasks were selected by following the selection criteria stated above.

3.6 Usability Issues

Usability issues can be defined as something which prevents the user in completion of a task or make the user to deviate the task goal in a wrong way. The identified issues can be analyzed in many different ways [21], such as frequency of unique issues with respect to the design, frequency of issues per participant, Issues by category, Issues by task. In this thesis work Issues by category and Issues by task were used for analyzing the usability data.

3.6.1 Issues by Category

In this study, the issued identified during usability testing will fall into one of the following issue categories:

- **Navigation** - Navigation was Guided /Unguided s perhaps the most important element of any web design. It helps the user to reach the goal. It shows the user where they are where they can go.
- **Terminology** – Unable to understand the terminology.
- **Layout** – Inability to detect something that they need to find. Aesthetic problems, unnecessary information.
- **Feedback** – User does not receive relevant feedback or it was inconsistent with what the user expects.
- **Data entry** – Problems with entering information.
- **Cognitive/Content** – The user’s interpretations, assessments and expectations.
- **Comprehension** – Inability to understand the instructions given to them on site.

By categorizing the issues, it will be very easy to know which type of issues effects the usability of a website and the areas needed to be concentrated in the next iterative design of a website.

3.6.2 Issues by Tasks

Analyzing the issues at task level will lead to know about which task need to be prioritized. More the number of issues, higher the priority of task will be. In addition to this, the unique and common issues between tasks can be known. This analysis will help to compare the Improvements of the design iterations.

3.7 Comparing Means

Comparing means helps to know which group has more error rate or satisfactions rate compared to other group. In this work, comparison was done across two different groups of participants, namely, guided group and unguided group with respect to the number of unique issues identified. Independent samples of the data were compared between these two groups.
Comparing means can be done either by using z-test or t-test based on the attributes such as sample size and the number of samples needed to be compared. Following points need to be considered while selecting the test [21]:

- Based on the sample size
  - Choose t-test if the sample size was less than 30.
  - Choose z-test if the sample size was 30 or more.

- Based on number of samples compared
  - Choose t-test if number of compared samples was 2.
  - Choose ANOVA test if number of compared samples was 3 or more.

The samples that were compared may be dependent samples or independent samples.

**Independent samples**
Comparing the data that belongs to two different groups within a product/website was considered as independent samples.

**Dependent samples**
Comparing the data with in a group between two sets of samples was considered as dependent samples.

By considering the above requirements t-test was chosen for analyzing the data since sample size was less than 30 and two independent samples (guided and unguided groups) were compared. T-Test: Two-Sample Assuming Equal Variances i.e., assuming two variances are roughly equal was performed on these samples.

For carrying out this t-test: input value for hypothesized mean difference was set to 0 which tells that there exists no difference in means and alpha value was set to 0.05.

**3.8 Correlation: Relation between variables**
Correlation tells degree of association between the two variables. The strength of relationship can be known with the help of $R^2$ value.

The degree of association between the different variables can be known with the help of correlation coefficient. Correlation coefficient varies between -1 to +1 and the following analysis can be analyzed:

- The relation between two variables will be stronger if coefficient was nearer to -1 or +1
  - Coefficient nearer to -1 tells negative relationship. i.e., one variable grows another variable decreases.
  - Coefficient nearer to +1 tells positive relationship. i.e., one variable grows another variable increase.
- Relation will be weaker if the coefficient was nearer to 0
In the current study, the correlation between the user size and the number of issues reported for guided/unguided issues was analyzed.

3.9 Severity Ratings

3.9.1 Issue Severity Rating

Usability issues have more impact on user experience as some of the issues are severe / may exasperate users / leads to loss of data / to take wrong decisions. Rating the severity for usability issues helps to prioritize the issues to be concentrated. By prioritizing the issues developer know which need to be rectified first so burden on him can be reduced.

In [21], methods to rate the severity of issues was explained. Among them severity ratings based on a combination of factors as shown in table 2, was chosen for this study to rate the severity level for identified issues. Impact of an issue on the user experience, frequency of users experiencing a particular issue were the factors needed to be considered for rating the severity of an issue.

<table>
<thead>
<tr>
<th>Few users experiencing a problem</th>
<th>Many users experiencing a problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small impact on the users experience</td>
<td>Low</td>
</tr>
<tr>
<td>Large impact on the users experience</td>
<td>Medium</td>
</tr>
</tbody>
</table>

Table 2: Severity rating based on combination of factors (impact on users experience, frequency of users experiencing a problem) – [21]

3.10 Methodology

Figure 5 represents the overview of the methodology employed in this study for collecting the usability data. The following points reflect the sequence of steps carried out in this methodology:

- A user will be selected if he/she meets the minimum requirements stated in section 3.3 for conducting the usability test.
- The user will be given a set of tasks and the group he/she belongs will be cited i.e., either guided task group or unguided task group.
- The user performs given set of task with respect to the instructions/guidelines they need to be followed depending upon the group he/she belongs to.
- The moderator observes how effectively user interacts with the website for each task and collects the necessary usability data i.e., the issues encountered by the user
3.11 Test Procedure/Process

Pilot testing was followed by the main test. Following sections reflects the procedure carried out for the pilot test and the main test.

3.11.1 Pilot test procedure

The following process was carried out for conducting the pilot test:

- In real world, most of the users to the selected websites were students. A total of 4 students were selected for carrying out this test.
- Questionnaire was given to the users which contains purpose of the test, role of the user, and set of the tasks they need to perform were stated.
- The users were asked to answer the demographic questions such as age, gender, experience.
- Four users were categorized in to two groups, namely, guided and unguided groups containing two users each.
- Users will perform the test on a computer with same browser and internet speed.
- Test was conducted individually.
- The users were asked to read and understand about the test stated in introduction and instruction parts of the questionnaire.
- After finishing the test, issues that were encountered by the users for each and every task will be noted down.
- The whole test result was analyzed and corrections were made to the methodology and questionnaire, if it doesn’t give the expected usability data.

After successful completion of pilot testing the focus must be on conducting a real test.
3.11.2 Main test procedure

In the main test, 40 users were selected for carrying out the usability testing. 40 users are categorized into two groups namely guided and unguided having 20 users each. The test process was explained as follows:

- In this test, each test participant was tested individually. For each and every participant, test introduction and instructions of the test were explained.
- Participant’s performs the usability test on a computer having an internet connection.
- Demographic information about the participants was noted regarding the age, gender and experience.
- Before starting the test, the users were intimated about the group (guided / unguided) they belong to:
  - Under Guided group, users will follow the guidelines to complete the task.
  - Under Unguided group, users have to perform tasks without any guidelines.

After fulfilling all pre-requisites for beginning the test, participants perform the given set tasks on a website.

3.12 Dependent Measures

The results were measured mainly by observing the interaction between participants and website i.e., the way they perform the given tasks effectively by fulfilling the given requirements and instructions. In addition, users were asked to give feedback or any problems they encountered while performing the task. Usability experience of an observer/moderator plays a vital role in identifying the issues i.e., more the experience more will be the number of issues identified. The issues that were identified for each task will be added to one of the issue category as discussed in the section 3.6.1.

The independent measure in this study was testing with the guided tasks and the unguided tasks. The two task designs will be performed by two different user groups respectively. The dependent measures were the number of issues shared between the two task designs.
4. RESULTS

4.1 Overview

www.mq.edu.au and www.siu.edu were the websites selected for carrying out the usability test. For each website, a set of six tasks were chosen based on the points mentioned in the section 3.5.2. Usability testing of a website was done with same set of tasks for both task designs and each task design was performed by twenty users.

A total of 40 users were participated in this usability testing. These users were divided into two groups – guided and unguided group with 20 users each. The participated users must have minimum requirements such as usage of internet, and judging of tasks. Participants having experience in usability testing was considered as an advantage. Every participant was evaluated individually on a same computer with sufficient internet speed.

Fortunately no user was excluded from the test. The results were collected in form of a questionnaire and they were analyzed to find the impact of task design, task coverage on usability testing. The questionnaire used for conducting the usability test was attached at end of the document in the appendix section.

4.2 Results Presentation

Issues that were encountered by the users have been listed at end of this document as an attachment in the appendix section. The usability data was collected for guided and unguided tasks performed by two different user groups containing 20 users for each user group. In below sections, presentation of the results was done for the websites www.mq.edu.au and www.siu.edu respectively. Following sections presents the usability testing results for evaluated websites with respect to:

- Comparing means
- Correlation
- Issue severity ratings

4.2.1 Guided task

Tables 3, 4 represents the overall usability data collected from the usability test users during the usability test for guided tasks.

<table>
<thead>
<tr>
<th>Guided task</th>
<th>Task1</th>
<th>Task2</th>
<th>Task3</th>
<th>Task4</th>
<th>Task5</th>
<th>Task6</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of Issues</td>
<td>Navigation</td>
<td>Terminology</td>
<td>Layout</td>
<td>Feedback</td>
<td>Comprehension</td>
<td>Cognitive/Content</td>
<td>Overall</td>
</tr>
<tr>
<td></td>
<td>0,00%</td>
<td>0,00%</td>
<td>0,00%</td>
<td>0,00%</td>
<td>0,00%</td>
<td>0,00%</td>
<td>0,00%</td>
</tr>
<tr>
<td>Terminology</td>
<td>55,00%</td>
<td>90,00%</td>
<td>5,00%</td>
<td>35,00%</td>
<td>75,00%</td>
<td>90,00%</td>
<td>100,00%</td>
</tr>
<tr>
<td>Layout</td>
<td>20,00%</td>
<td>0,00%</td>
<td>75,00%</td>
<td>85,00%</td>
<td>0,00%</td>
<td>25,00%</td>
<td>95,00%</td>
</tr>
<tr>
<td>Feedback</td>
<td>0,00%</td>
<td>70,00%</td>
<td>0,00%</td>
<td>30,00%</td>
<td>0,00%</td>
<td>0,00%</td>
<td>75,00%</td>
</tr>
<tr>
<td>Comprehension</td>
<td>65,00%</td>
<td>55,00%</td>
<td>30,00%</td>
<td>15,00%</td>
<td>75,00%</td>
<td>0,00%</td>
<td>90,00%</td>
</tr>
<tr>
<td>Cognitive/Content</td>
<td>85,00%</td>
<td>95,00%</td>
<td>80,00%</td>
<td>45,00%</td>
<td>90,00%</td>
<td>100,00%</td>
<td>100,00%</td>
</tr>
<tr>
<td>Overall</td>
<td>100,00%</td>
<td>100,00%</td>
<td>100,00%</td>
<td>100,00%</td>
<td>100,00%</td>
<td>100,00%</td>
<td>100,00%</td>
</tr>
</tbody>
</table>

Table 3: Overall Result for the guided task – www.mq.edu.au
4.2.1.1 Issues by Category

Figure 6 and 7 below shows the percentage of total number of users encounter for each issue category in both the websites for guided tasks.

![Guided task - www.mq.edu.au](image)

**Figure 6:** Frequency of participants encountered for each issue category – [www.mq.edu.au](http://www.mq.edu.au)
After statistical analysis of the above results, it was noted that guided task reported all kind of issues categories in both the websites. The following points can be analyzed:

- It reported all types of issues categories except navigation category in www.mq.edu.au website.
- It reported cognitive, comprehension, layout, and terminology issues majorly in www.siu.edu.

### 4.2.1.2 Issues by Task

Figures 8 and 9 shows the percentage of total number of participants encountered each issues category for each task in both the websites for guided tasks.

---

**Figure 7:** Frequency of participants encountered for each issue category – www.siu.edu

**Figure 8:** Percentage of total number of participants encountered each type of issues category for the tasks – www.mq.edu.au
Figure 9: Percentage of total number of participants encountered each type of issues category for the tasks – www.siu.edu

After statistical analysis of the above results shows most of the users had encountered all types of issues for all the tasks.

4.2.2 Unguided task

Tables 5, 6 represents the overall usability data collected from the usability test users during the usability test for unguided tasks.

<table>
<thead>
<tr>
<th>Unguided task</th>
<th>Type of Issues</th>
<th>Task1</th>
<th>Task2</th>
<th>Task3</th>
<th>Task4</th>
<th>Task5</th>
<th>Task6</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Navigation</td>
<td>15,00%</td>
<td>0,00%</td>
<td>5,00%</td>
<td>0,00%</td>
<td>5,00%</td>
<td>5,00%</td>
<td>15,00%</td>
<td></td>
</tr>
<tr>
<td>Terminology</td>
<td>10,00%</td>
<td>35,00%</td>
<td>45,00%</td>
<td>10,00%</td>
<td>70,00%</td>
<td>5,00%</td>
<td>95,00%</td>
<td></td>
</tr>
<tr>
<td>Layout</td>
<td>45,00%</td>
<td>0,00%</td>
<td>0,00%</td>
<td>35,00%</td>
<td>0,00%</td>
<td>15,00%</td>
<td>65,00%</td>
<td></td>
</tr>
<tr>
<td>Feedback</td>
<td>30,00%</td>
<td>10,00%</td>
<td>0,00%</td>
<td>0,00%</td>
<td>5,00%</td>
<td>5,00%</td>
<td>35,00%</td>
<td></td>
</tr>
<tr>
<td>Comprehension</td>
<td>10,00%</td>
<td>35,00%</td>
<td>40,00%</td>
<td>5,00%</td>
<td>50,00%</td>
<td>55,00%</td>
<td>85,00%</td>
<td></td>
</tr>
<tr>
<td>Cognitive/Content</td>
<td>70,00%</td>
<td>70,00%</td>
<td>75,00%</td>
<td>80,00%</td>
<td>75,00%</td>
<td>65,00%</td>
<td>100,00%</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>100,00%</td>
<td>90,00%</td>
<td>90,00%</td>
<td>85,00%</td>
<td>95,00%</td>
<td>95,00%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Overall Result for unguided task – www.mq.edu.au

<table>
<thead>
<tr>
<th>Unguided task</th>
<th>Type of Issues</th>
<th>Task1</th>
<th>Task2</th>
<th>Task3</th>
<th>Task4</th>
<th>Task5</th>
<th>Task6</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Navigation</td>
<td>5,00%</td>
<td>5,00%</td>
<td>25,00%</td>
<td>5,00%</td>
<td>5,00%</td>
<td>5,00%</td>
<td>45,00%</td>
<td></td>
</tr>
<tr>
<td>Terminology</td>
<td>55,00%</td>
<td>70,00%</td>
<td>0,00%</td>
<td>60,00%</td>
<td>0,00%</td>
<td>50,00%</td>
<td>95,00%</td>
<td></td>
</tr>
<tr>
<td>Layout</td>
<td>40,00%</td>
<td>0,00%</td>
<td>80,00%</td>
<td>0,00%</td>
<td>0,00%</td>
<td>70,00%</td>
<td>100,00%</td>
<td></td>
</tr>
<tr>
<td>Feedback</td>
<td>10,00%</td>
<td>0,00%</td>
<td>0,00%</td>
<td>0,00%</td>
<td>0,00%</td>
<td>0,00%</td>
<td>10,00%</td>
<td></td>
</tr>
<tr>
<td>Comprehension</td>
<td>80,00%</td>
<td>10,00%</td>
<td>50,00%</td>
<td>50,00%</td>
<td>0,00%</td>
<td>40,00%</td>
<td>100,00%</td>
<td></td>
</tr>
<tr>
<td>Cognitive/Content</td>
<td>75,00%</td>
<td>80,00%</td>
<td>0,00%</td>
<td>20,00%</td>
<td>0,00%</td>
<td>100,00%</td>
<td>100,00%</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>100,00%</td>
<td>90,00%</td>
<td>85,00%</td>
<td>65,00%</td>
<td>0,00%</td>
<td>100,00%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6: Overall Result for unguided task – www.siu.edu
4.2.2.1 Issues by Category

Figure 10 and 11 below shows the percentage of total number of users encounter for each issue category in both the websites for guided task.

**Figure 10:** Frequency of participants encountered for each issue category – www.mq.edu.au

**Figure 11:** Frequency of participants encountered for each issue category – www.siu.edu
After statistical analysis of the above results, it was noted that unguided task reported all kind of issues categories in both the websites. The following points can be analyzed:

- It reported all types of issues categories majorly cognitive, comprehensive, layout, and terminology in www.mq.edu.au website.
- It reported cognitive, comprehension, layout, and terminology issues majorly in www.siu.edu website.

4.2.2.1 Issues by task

Figures 12 and 13 below shows the percentage of total number of participants encountered each issue category for each task in both the websites for unguided task.

![Figure 12: Percentage of total number of participants encountered each type of issues category for the tasks – www.mq.edu.au](image)

![Figure 13: Percentage of total number of participants encountered each type of issues category for the tasks – www.siu.edu](image)
After statistical analysis of the above results shows most of the users had encounter all types of issues for all the tasks.

### 4.2.3 Comparing Means

In this study, “t-test: two-sample assuming equal variance” was chosen for comparing means between guided and unguided task. In this t-test, two variances were assumed roughly equal and it was hypothesized that there will be no difference in mean, input value for hypothesized mean difference and alpha level was set to 0 and 0.05 respectively. Table 8 below shows the t-test results for www.mq.edu.au website to the data represented in table 7.

<table>
<thead>
<tr>
<th>User</th>
<th>No of issues - guided</th>
<th>No of issues - unguided</th>
</tr>
</thead>
<tbody>
<tr>
<td>User1</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>User2</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>User3</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>User4</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td>User5</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>User6</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>User7</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>User8</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>User9</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>User10</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>User11</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>User12</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>User13</td>
<td>14</td>
<td>22</td>
</tr>
<tr>
<td>User14</td>
<td>17</td>
<td>14</td>
</tr>
<tr>
<td>User15</td>
<td>18</td>
<td>16</td>
</tr>
<tr>
<td>User16</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>User17</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>User18</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>User19</td>
<td>13</td>
<td>22</td>
</tr>
<tr>
<td>User20</td>
<td>14</td>
<td>14</td>
</tr>
</tbody>
</table>

**Table 7:** Represents the number of unique issues encountered by each user – www.mq.edu.au

**t-Test: Two-Sample Assuming Equal Variances**

<table>
<thead>
<tr>
<th></th>
<th>No of issues - guided</th>
<th>No of issues – unguided</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>12,45</td>
<td>9,85</td>
</tr>
<tr>
<td>Variance</td>
<td>13,1026316</td>
<td>19,60789474</td>
</tr>
<tr>
<td>Observations</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Pooled Variance</td>
<td>16,3552632</td>
<td></td>
</tr>
<tr>
<td>Hypothesized Mean Difference</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>df</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>t Stat</td>
<td>2,03303371</td>
<td></td>
</tr>
<tr>
<td>P(T&lt;=t) one-tail</td>
<td>0,02453903</td>
<td></td>
</tr>
<tr>
<td>t Critical one-tail</td>
<td>1,68595446</td>
<td></td>
</tr>
<tr>
<td>P(T&lt;=t) two-tail</td>
<td>0,04907806</td>
<td></td>
</tr>
<tr>
<td>t Critical two-tail</td>
<td>2,02439416</td>
<td></td>
</tr>
</tbody>
</table>

**Table 8:** t-test results for the data represented in table 7 – www.mq.edu.au
Table 10 below shows the t-test results for www.siu.edu website to the data represented in table 9.

<table>
<thead>
<tr>
<th>User</th>
<th>No of issues - guided</th>
<th>No of issues - unguided</th>
</tr>
</thead>
<tbody>
<tr>
<td>user1</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>user2</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>user3</td>
<td>3</td>
<td>13</td>
</tr>
<tr>
<td>user4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>user5</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>user6</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>user7</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>user8</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>user9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>user10</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>user11</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>user12</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>user13</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>user14</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>user15</td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td>user16</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>user17</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>user18</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>user19</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>user20</td>
<td>6</td>
<td>12</td>
</tr>
</tbody>
</table>

Table 9: Represents the number of unique issues encountered by each user – www.siu.edu

<table>
<thead>
<tr>
<th></th>
<th>No of issues - guided</th>
<th>No of issues - unguided</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>6.8</td>
<td>10.1</td>
</tr>
<tr>
<td>Variance</td>
<td>7,431578947</td>
<td>7,989473684</td>
</tr>
<tr>
<td>Observations</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Pooled Variance</td>
<td>7,710526316</td>
<td></td>
</tr>
<tr>
<td>Hypothesized Mean</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Difference</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Df</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>t Stat</td>
<td>-3.758131116</td>
<td></td>
</tr>
<tr>
<td>P(T&lt;=t) one-tail</td>
<td>0.000287324</td>
<td></td>
</tr>
<tr>
<td>t Critical one-tail</td>
<td>1.68595446</td>
<td></td>
</tr>
<tr>
<td>P(T&lt;=t) two-tail</td>
<td>0.000574648</td>
<td></td>
</tr>
<tr>
<td>t Critical two-tail</td>
<td>2.024394164</td>
<td></td>
</tr>
</tbody>
</table>

Table 10: t-test results for the data represented in table 9 – www.siu.edu

From the tables 8, 10 the value of $p$ for both the website is less than 0.05, which indicates that there is a significant difference between the guided and unguided tasks.
Tables 11 and 12 below shows the number of issues falls under each type of issue and overall issues in total.

<table>
<thead>
<tr>
<th></th>
<th>Guided</th>
<th>Unguided</th>
<th>shared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Navigation</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Terminology</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Layout</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Feedback</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Comprehension</td>
<td>2</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Cognitive/Content</td>
<td>7</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>total</td>
<td>17</td>
<td>26</td>
<td>11</td>
</tr>
</tbody>
</table>

**Table 11:** Represents the number of issues comes under guided, unguided and shared groups – www.mq.edu.au

<table>
<thead>
<tr>
<th></th>
<th>Guided</th>
<th>Unguided</th>
<th>shared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Navigation</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Terminology</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Layout</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Feedback</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Comprehension</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Cognitive/Content</td>
<td>7</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>total</td>
<td>17</td>
<td>24</td>
<td>10</td>
</tr>
</tbody>
</table>

**Table 12:** Represents the number of issues comes under guided, unguided and shared groups – www.siu.edu

Figures 14 and 15 below shows the percentage of issues covered among guided, unguided, and shared groups.

**Figure 14:** Pie-chart representation for the data represented in table 11 – www.mq.edu.au
After statistical analysis of the above results, the following points can be drawn:

- Unguided task reports more number of issues compared to guided task.
- Unguided group reported more number of unique issues than guided group.

### 4.2.4 Correlation

In this study, correlation analysis was worked out to find the relationship between the user size and number of issues reported with respect to its user size in both the task designs. User size was varied from 1 to 20. The number of issues reported for each user size was calculated by taking the average of 20 samples.

**Relationship between the user size and number of issues:**

Figures 16 and 17 below shows the relationship between the user size and the number of issues reported for its corresponding user size in both the task designs. User size was varied from 1 to 20. By adding the trend line to the graph will help the tester for better visualization of the relation between these variables.

The measure of relationship between these variables was done with the help of $R^2$ value.
After statistical analysis of the results, the following point can be drawn:

- For both the task designs, a strong ($R^2 \approx 1$) and positive relation exists between these variables because as the user size increases, the number of issues increases.
4.2.5 Issue Severity Rating

Based on the explanation done in section 3.9.1, issues encountered in this study were rated with either low/medium/high severity. The severity ratings for all the identified issues have been listed as an attachment at end of the appendix section. Tables 13 and 14 shows the number of issues falls under low/medium/high severity issues for guided, unguided, and shared groups.

<table>
<thead>
<tr>
<th>MU</th>
<th>Guided</th>
<th>Unguided</th>
<th>Shared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>8</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Medium</td>
<td>6</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>High</td>
<td>3</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>26</td>
<td>11</td>
</tr>
</tbody>
</table>

**Table 13:** The number of issues shared among Low, Medium, and High level severity between guided, unguided, and shared groups – www.mq.edu.au.

<table>
<thead>
<tr>
<th>SIU</th>
<th>Guided</th>
<th>Unguided</th>
<th>Shared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>8</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Medium</td>
<td>4</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>High</td>
<td>5</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>24</td>
<td>10</td>
</tr>
</tbody>
</table>

**Table 14:** The number of issues shared among Low, Medium, and High level severity between guided, unguided, and shared groups – www.siu.edu.

From the tables 13, 14 the following statistical analysis can be drawn:

- Guided task reported more number of low severity issues.
- Unguided task reported more number of medium and high severity issues.

4.2.4.1 Chi-square test

The chi-square test was used to determine whether there exists a significant difference between the two task designs with respect to the number of unique issues reported for each severity level. The function, CHITEST *(actual range; expected range)* was used in Excel to perform this test. The actual range was the number of unique issues observed for each severity level. It was hypothesized that the expected frequency of issues were equal in each task design. The expected range was the total number of unique issues among both the task designs for each severity level divided by the number of groups i.e., 2.

<table>
<thead>
<tr>
<th>CHITEST</th>
<th><a href="http://www.mq.edu.au">www.mq.edu.au</a></th>
<th><a href="http://www.siu.edu">www.siu.edu</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Low severity</td>
<td>0.405380556</td>
<td>0.592980098</td>
</tr>
<tr>
<td>Medium severity</td>
<td>0.108293656</td>
<td>0.365712296</td>
</tr>
<tr>
<td>High severity</td>
<td>0.131668016</td>
<td>0.133614403</td>
</tr>
</tbody>
</table>

**Table 15:** An overview of chi-square test results for comparing the task designs in www.mq.edu.au and www.siu.edu websites.
Table 15 represents the results of the chi-square test for comparing the task designs with respect to the number of unique severity issues reported for each severity level in both the tested websites. After statistical analysis of the results, the following points can be drawn for each severity level:

Low severity: for both the websites, the chi test value was less than 0.05 which resembles that no significant difference exists between both the task designs in reporting number of low severity degree issues.

Medium severity: for both the websites, the chi test value was less than 0.05 which resembles that no significant difference exists between both the task designs in reporting number of medium severity degree issues.

High severity: for both the websites, the chi test value was less than 0.05 which resembles that no significant difference exists between both the task designs in reporting number of high severity degree issues.
5. DISCUSSION

The usability test carried out in this study revealed significant results. A trend of consistent results was quite obvious on how the participants performed the tasks for each task design, which proved to be very useful during the result analysis.

5.1 Task Completion

As stated in section 3.1.1, the pilot study was carried out by four participants and the main study was carried out by forty participants (twenty performing guided task and twenty performing unguided task). The four participants participated in the pilot test were not involved among the forty participants participated in the main test. Each study therefore had its own participants recruited based on the criteria stated in section 3.3 which made to collect the accurate usability data.

Completing a task with success or failure was not considered as a metric. The participants try to accomplish the each given task within a certain reasonable amount of time. A Time span of 5-6 minutes was made as a benchmark for each and every task. The Observer identified the issues for each task within this amount of time.

5.2 Accuracy of the Tests in this study

For each task design, the usability data was calculated for each task per participant but analysis were mainly based on considering all the identified unique issues among tasks designs in the website. During the test, evaluator limited to a set of rules stated in section 3.2 during the usability test for collecting the accurate usability data. Before starting the usability test, evaluator encouraged the participants to make them comfortable and instructed them with the rules need to be followed during the test which made them to behave like real users and it made easy for the evaluator to collect an accurate and necessary usability data. The evaluator identified the issued by observing how effectively a user interacts with the website.

The level of communication between the evaluator and the participants was kept very less during the test. The participants were well instructed not to raise any kind of questions during the test. The participants were interrupted by the evaluator when they exceeds the benchmark time of a task with a short verbal feedback like ‘done’, ‘finish’, and ‘over’.

5.3 Discussion of the Statistical Results

From the statistical analysis of all the results from forty participants, it was clear that unguided task had a great impact on the usability test results compared to the guided task. For both the task designs, identified issues were recorded and only the unique issues with respect to the task designs were considered to find the relation between the task designs and their effect on the usability test results.

The number of unique issues encountered in each group will depend on the way user performed the tasks. In the present scenario, user performed the tasks with/without a set of guidelines. As the users will follow the guidelines for completing the task, in general view it shows that the number of unique issues encountered in a guided task will be less compared to
The unguided group reports more number of unique issues than guided task. With this implication it can’t be confirmed that unguided group will report more number of unique issues than guided group in all cases. It mainly depends on the website layout, functioning, and purpose. But in most cases, unguided group will report more unique issues than guided group. It was also observed that, for each issues category unguided group reported more number of unique issues compared to guided group.

By categorizing the issues as stated in section 3.6.1, will aid the developer or tester to know which areas of the website are causing most usability issues. By prioritizing the issue categories, the developer or tester can rectify the issues in the next iterative design of a website. Prioritization of the issue categories can be done based upon the frequency of participants encountered correspondingly i.e., more the number of participants, higher will be the priority.

In both the websites, guided task reported issues of all categories except navigation. Whereas unguided task reported issues of all types of issues categories. Issues by category will help the usability tester to know which type of issues encountered by more number of users and also which task seems to be difficult among the given tasks. It was noted that issue types of cognitive, terminology, and comprehension were encountered by maximum number of users among guided and unguided tasks for both the websites. The result of this observation mainly depends on the website and it will vary from website to website. Both this groups will help to identify issues of all the categories.

The following implication can be drawn:

- Each of the task design will report issues belonging to most of the issue categories. Usage of both the task designs will help to find all type of issues in a website.

Fixing all the identified issues immediately frustrates the developer. Not all the issues have same impact on the user experience: some have more impact than others. Rating the severity for the identified issues will help the usability practitioners or developers to know how they impact on the user experience. In the current study, severity rating for the issues was done as explained in section 3.9.1. As guided tasks will be completed successfully by the users as they were performed with a set of guidelines and the severity level for most of the identified issues will be low and medium. This group will also report the high severity level issues in a less
number compared to the number of low/medium level severity issues. For example, navigation, and layout category issues were the high severity level issues reported by guided group. In the current work, both the websites reported more number of low and medium severity level issues compared to high severity level issues.

The unguided group will report the issues of all severity levels as the users have to perform the task without no guidance and will behave like a real users. As they behave like real users most of the issues reported by them will be of medium/high level severity issues. In the current work, website www.mq.edu.au reported more number of medium severity level issues and website www.siu.edu reported more number of high severity level issues.

From the above observation, the Following implication can be drawn among the task designs in reporting the number of severity degree issues:

- Guided task reported more number of low severity issues, whereas unguided task reported more number of medium and high severity issues.

For any usability practitioner, the main goal is to identify issues of all categories for improving the usability performance of any website. From the above implication, it can stated that the usage of both the groups will help to identify the issues of all severity level issues which helps for increasing the usability of the website.

The two task designs i.e., guided and unguided tasks were compared in between two different of users with respect to the number of unique issues reported. As these were the two independent samples, “T-test: Two Samples Assuming Equal Variance” was performed to compare the number of issues reported by the guided and unguided group participants. The important piece of data in this result is p-value for the two tailed distribution. It’s very important to know which task design reported more number of issues. The p-values for both the websites are 0.04 and 0.0005, which is well below the threshold value i.e., 0.05 in both the websites. It implies:

- Statistically significant difference exists in finding the number of issues between the guided and unguided tasks.

“Chi-Square Test” was used to determine the significant difference between the guided and unguided tasks in reporting the total number of issues corresponding to the low, medium, and high severity level issues. From this test result, the following implication can be drawn:

- There are no statistically significant differences between the two task design approaches in terms of number of unique usability issues discovered (no matter what severity degree).

In addition to the above suggestions, it was proven that a positive and strong relation exists between the number of users and number of issues found for both the task designs. As the user size increases, the number of issues seemed to be increasing correspondingly.
6. CONCLUSION

From the results and discussions, the unguided task had a major impact on the usability test results compared to guided task. Usage of both the task designs will help to find all types of issues in any given website.

There exists a statistical significant difference between the guided and unguided tasks in reporting the number of unique issues.

Both the task designs reported all kinds of severity level issues. Among them, the unguided tasks design revealed more number of medium and high severity issues, whereas the guided task found more number of low severity issues.

For both the task designs, a strong and positive relation exists between the user size and the number of issues discovered correspondingly.

This study highly recommends the usability practitioner’s to consider the proposed task design while employing the usability test for improving the usability of any give website.
7. FUTUREWORK

The expectation measure of post task rating system has not been analyzed in this study. Probably it would be motivating if a future research would be carried out on it so that it would be more precise to know the effect of task design, task coverage on the usability test.

As seen in this study, the selection of tasks was done based on the factors explained in section 3.5.2. Perhaps it would be interesting if a future research would be carried out on factors that need to be considered for choosing a correct set of tasks since selecting a wrong set of tasks in a usability evaluation may lead to wrong assumptions on the usability of a website and will create issues while it is deployed in real time.
REFERENCES


9. GLOSSARY

- **Unguided task** - User tries to solve the task acting like a real user.

- **Qualitative analysis** - Examination of measurable and verifiable data.

- **Quantitative analysis** - Examination of non-measurable data.

- **Guided task** - User tries to solve the task with help of some guidelines.

- **Task** - Task is nothing but a collection of events that needs to be fulfilled for a task to be success or failure.

- **t-test** - A statistical procedure used to compare the usability data.

- **Usability** - Usability lies in the interaction of the user with the product or system and can only be accurately measured by assessing user performance, satisfaction and acceptability.

- **Usability testing** - A technique used to evaluate a product by testing it with representative users.
10. APPENDIX

10.1 Identified Issues

Following table represents the presentation of identified issues in the current usability testing and its corresponding details like group it belongs to, number of participants encountered, severity level.

<table>
<thead>
<tr>
<th>Cognitive</th>
<th>Issue</th>
<th>Group</th>
<th>Number of users encountered the issue</th>
<th>Severity level</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://www.siu.edu">www.siu.edu</a></td>
<td>Expected to get information under Academics</td>
<td>unguided</td>
<td>6</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td>Expected to get information under Colleges</td>
<td>unguided</td>
<td>9</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td>Expected to get information under Schedule</td>
<td>shared</td>
<td>11</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td>Expected to get information under Inside Athletics</td>
<td>guided</td>
<td>11</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Expected to get information under Athletics</td>
<td>guided</td>
<td>9</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Expected to get information under celebrating women’s under Athletics</td>
<td>unguided</td>
<td>12</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td>Expected to get information under Graduate</td>
<td>unguided</td>
<td>14</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Expected to get information under Undergraduate</td>
<td>unguided</td>
<td>16</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Expected to get information under About SIU</td>
<td>unguided</td>
<td>9</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Expected to get information under graduate program</td>
<td>guided</td>
<td>10</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td>Expected to get information under Research</td>
<td>shared</td>
<td>11</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Expected to get information under Agricultural Sciences</td>
<td>unguided</td>
<td>12</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Expected to get information under Applied Sciences</td>
<td>shared</td>
<td>15</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Expected to get information in colleges under Academics</td>
<td>guided</td>
<td>14</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Expected to get information under Admissions</td>
<td>guided</td>
<td>14</td>
<td>High</td>
</tr>
<tr>
<td><a href="http://www.mq.edu.au">www.mq.edu.au</a></td>
<td>Expected to get information under Events</td>
<td>unguided</td>
<td>11</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td>Expected to get information under Notices &amp; Events for student</td>
<td>shared</td>
<td>17</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Expected to get information under News &amp; Events for staff</td>
<td>unguided</td>
<td>8</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Expected to get information under “OnCampus”</td>
<td>unguided</td>
<td>5</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Expected to get information under Important dates</td>
<td>guided</td>
<td>14</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td>Expected to get information under International</td>
<td>unguided</td>
<td>5</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Expected to get information under Courses</td>
<td>shared</td>
<td>13</td>
<td>Medium</td>
</tr>
<tr>
<td>Expected to get information in New &amp; Events under “About us”</td>
<td>d</td>
<td>14</td>
<td>Medium</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------------------</td>
<td>---</td>
<td>----</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>Expected to get information under “Campus Life”</td>
<td>guide d</td>
<td>9</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Expected to get information under Courses</td>
<td>guide d</td>
<td>15</td>
<td>Medium</td>
<td></td>
</tr>
<tr>
<td>Expected to get information under Higher Degree Research</td>
<td>guide d</td>
<td>12</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Expected to get information under Centre &amp; groups</td>
<td>guide d</td>
<td>13</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Expected to get information under Support</td>
<td>guide d</td>
<td>1</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Expected to get information under Opportunities</td>
<td>guide d</td>
<td>14</td>
<td>Medium</td>
<td></td>
</tr>
<tr>
<td>Expected to get information under Study Skills</td>
<td>guide d</td>
<td>12</td>
<td>Medium</td>
<td></td>
</tr>
<tr>
<td>Expected to get information under ask.mq.edu.an</td>
<td>guide d</td>
<td>5</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Expected to get information under Business &amp; Community</td>
<td>guide d</td>
<td>9</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Expected to get information under International</td>
<td>guide d</td>
<td>14</td>
<td>Medium</td>
<td></td>
</tr>
<tr>
<td>Expected to get information under Short Courses</td>
<td>guide d</td>
<td>13</td>
<td>Medium</td>
<td></td>
</tr>
<tr>
<td>Expected to get information under Non-Award study</td>
<td>guide d</td>
<td>8</td>
<td>Low</td>
<td></td>
</tr>
</tbody>
</table>

Comprehensive:

| not sure whether to click Schedule/Sports                      | unguide d | 3 | Low |
| not sure whether to click Sports / Inside Athletics            | guide d | 7 | Low |
| not sure whether to click Schedule/Inside Athletics            | guide d | 9 | Low |
| not sure whether to click All news/SIU events                  | guide d | 5 | Low |
| Searched whether parts of the webpage                          | guide d | 4 | Low |
| not sure whether to click Admissions/Academics                 | guide d | 11 | High |
| not sure whether to click Graduate/Cos & Finance               | guide d | 14 | High |
| not sure whether to click Agricultural Sciences/Applied Sciences | guide d | 7 | Low |
| not sure whether to click graduate/undergraduate               | guide d | 11 | Medium |

<p>| not sure whether to click offices &amp; units / news &amp; events       | guide d | 14 | Medium |
| not sure whether to click Courses / International               | guide d | 11 | Medium |</p>
<table>
<thead>
<tr>
<th>Issue</th>
<th>Website</th>
<th>Difficulty</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>not sure whether to click English language / IELTS</td>
<td>guided</td>
<td>9</td>
<td>Medium</td>
</tr>
<tr>
<td>not sure whether to click Student page / Main Page</td>
<td>unguided</td>
<td>13</td>
<td>High</td>
</tr>
<tr>
<td>not sure whether to click Campuslife / News &amp; Events</td>
<td>guided</td>
<td>9</td>
<td>Medium</td>
</tr>
<tr>
<td>not sure whether to click Courses / Research</td>
<td>unguided</td>
<td>13</td>
<td>High</td>
</tr>
<tr>
<td>not sure whether to click Career Development / Study Skills</td>
<td>unguided</td>
<td>14</td>
<td>High</td>
</tr>
<tr>
<td>Terminology</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://www.siu.edu">www.siu.edu</a></td>
<td>unable to understand Schedule/Results</td>
<td>unguided</td>
<td>14</td>
</tr>
<tr>
<td>unable to understand Inside Athletics</td>
<td>guided</td>
<td>11</td>
<td>Low</td>
</tr>
<tr>
<td>unable to understand Cost &amp; Finance AID</td>
<td>guided</td>
<td>12</td>
<td>Low</td>
</tr>
<tr>
<td>unable to understand Arts &amp; Design department</td>
<td>guided</td>
<td>12</td>
<td>Medium</td>
</tr>
<tr>
<td><a href="http://www.mq.edu.au">www.mq.edu.au</a></td>
<td>unable to understand difference between career development and study skills</td>
<td>unguided</td>
<td>11</td>
</tr>
<tr>
<td>unable to understand high degree research</td>
<td>guided</td>
<td>15</td>
<td>Medium</td>
</tr>
<tr>
<td>unable to understand world leading research</td>
<td>unguided</td>
<td>12</td>
<td>Medium</td>
</tr>
<tr>
<td>unable to understand Centre &amp; groups</td>
<td>unguided</td>
<td>15</td>
<td>High</td>
</tr>
<tr>
<td>Layout</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://www.siu.edu">www.siu.edu</a></td>
<td>unwanted link Schedule/Result</td>
<td>unguided</td>
<td>4</td>
</tr>
<tr>
<td>Inability to detect ‘visitors’ on home page</td>
<td>guided</td>
<td>12</td>
<td>High</td>
</tr>
<tr>
<td>Wrong placement of ‘visitors’</td>
<td>shared</td>
<td>11</td>
<td>High</td>
</tr>
<tr>
<td>unable to identify ‘Colleges’ due to color</td>
<td>guided</td>
<td>8</td>
<td>High</td>
</tr>
<tr>
<td><a href="http://www.mq.edu.au">www.mq.edu.au</a></td>
<td>no separate ‘tab/link' for news</td>
<td>shared</td>
<td>6</td>
</tr>
<tr>
<td>most of them searched on left of the webpage</td>
<td>shared</td>
<td>11</td>
<td>Low</td>
</tr>
<tr>
<td>most of them searched on bottom of the webpage</td>
<td>guided</td>
<td>7</td>
<td>Medium</td>
</tr>
<tr>
<td>unable to distinguish student page/main page/staff page</td>
<td>unguided</td>
<td>15</td>
<td>High</td>
</tr>
<tr>
<td>Navigation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://www.siu.edu">www.siu.edu</a></td>
<td>no link for siu events on home page</td>
<td>unguided</td>
<td>12</td>
</tr>
</tbody>
</table>

Table 16: Issues that were identified during usability evaluation of the both websites
10.2 Usability Observation Form

Introduction to Usability testing

The test that you will perform is part of a Master thesis at Human-Centered System, Linköping University. After this test, results are used to find the correlation between tasks and number of issues.

Participation in the test is completely voluntary and faced issues must be reported. All information collected will remain anonymous. The consolidated results of the study will be published at Linköping University.

The test is performed on a computer provided with an internet connection. In general, you have to perform usability testing on two web sites. The test consists of a set of tasks to be performed on a given website. The user who performs the test will fall into any one of the categories. First category is guided users, they are provided with a set of guidelines to complete the task. Secondly unguided tasks, here users have to complete the tasks by their own knowledge, no guidelines will be provide.

What you have to report? Either task success or failure must be reported. If failure, the issues you faced or the reasons that made them to go wrong must be reported. The time you taken to finish a task are also taken into consideration.

In some part of the questions you have to rate from 1 to 7. Here 1 implies hard and 7 implies easy.

The results you provide are so valuable.

Thank you for participating in the study!
Demographic questions

Gender (F/M):

How old are you:

Do you have any Usability testing experience?:

Rate your usage of internet between 1 and 7:
Usability testing for guided tasks

Instructions:
- You have to perform the tasks on the websites.
- Follow the guideline for each task to perform.
- For each task you have to rate before (expectation) and after (experience) performing the task.
- You have to rate the task between 1 and 7. Rating 1 is Hard ------- Rating 7 is Easy.
- You have to report the issues faced for each task.
- Don’t use the search option while completing the task.

Definitions:
- Expectation rate: How easy or difficult do you expect this task to be?
- Experience rate: How easy or difficult was task to do.

Usability testing 1

Perform the following tasks on Southern Illinois University website.

Task 1: Find the results for recently completed women’s basketball events at Saluki Athletics.

Guidelines:
   a. Go to home page.
   b. Click on Athletics.

Expectation rating (1-7):
Experience rating (1-7):
Issues:

Task 2: Find about the "University honors program" offers at this university.

Guidelines:
   a. Go to home page
   b. Go to academics

Expectation rating (1-7):
Experience rating (1-7):
Issues:
**Task 3:** Find the events at www.siu.edu for the visitors in this month.

Guidelines:
- a. Browse the home page
- b. Go to ‘Visitors’ on home page.
- c. Come back to home page.

Expectation rating (1-7):
Experience rating (1-7):
Issues:

**Task 4:** Find the tuition fee for Graduate studies.

Guidelines:
- a. Browse the home page
- b. Click on academics
- c. Come back to home page.

Expectation rating (1-7):
Experience rating (1-7):
Issues:

**Task 5:** How many of libraries does Southern Illinois University provide to students for accessing the books?

Guidelines:
- a. Browse the home page
- b. Click on Libraries

Expectation rating (1-7):
Experience rating (1-7):
Issues:
**Task 6:** Find Graduate programs offered in Arts & Design department.

Guidelines:
- a. Go to home page.
- b. Click on Colleges
- c. Go to Liberal Arts under Colleges.

Expectation rating (1-7):
Experience rating (1-7):
Issues:

**Usability testing 2**

Perform the following tasks on Macquarie University website.

**Task 1:** Find about the news room for the media at this university.

Guidelines:
- a. Go to the homepage.
- b. Go to the “About us” page.
- c. Check for the task to do.

Expectation rating (1-7):
Experience rating (1-7):
Issues:

**Task 2:** Macquarie University offers to book a slot for IELTS exam, Find out the amount to be paid for this exam.

Guidelines:
- a. Go to the homepage.
- b. Go to the courses page.
- c. Perform the task.

Expectation rating (1-7):
Experience rating (1-7):
Issues:
**Task 3:** Find the mail-id of the person to enquire about the postgraduate scholarships.

Guidelines:
- a. Go to the homepage.
- b. Click on Future students.
- c. Perform the task.

Expectation rating (1-7): 
Experience rating (1-7): 
Issues:

---

**Task 4:** Find about the student events today.

Guidelines:
- a. Go to the homepage.
- b. Go to student webpage
- c. Perform the task.

Expectation rating (1-7): 
Experience rating (1-7): 
Issues:

---

**Task 5:** Find about the disciplines that are offered in research of earth sciences.

Guidelines:
- a. Go to the homepage.
- b. Click on “research”
- c. Perform the task.

Expectation rating (1-7): 
Experience rating (1-7): 
Time to finish:
Task 6: Find the upcoming carrier service events for students to develop their skills.

Guidelines:
   a. Browse the homepage.
   b. Go to Students page.
   c. Check under Support tab.

Expectation rating (1-7): Time to finish:
Experience rating (1-7):
Issues:
Usability testing for unguided tasks.

Instructions:
- You have to perform the mentioned tasks on the websites.
- For each task you have to rate before (expectation) and after (experience) performing the task.
- You have to rate the task between 1 and 7. Rating 1 is Hard ------Rating 7 is Easy.
- You have to report the issues faced for each task.
- Don’t use the search option while completing the task.

Definitions:
- Expectation rate: After reading the task how much you can rate between 1 and 7.
- Experience rate: After performing the task how much you can rate between 1 and 7.

Usability testing 1

Perform the following tasks on Southern Illinois University website.

Task 1: Find the results for recently completed women's basketball events at Saluki Athletics.

Expectation rating (1-7):
Experience rating (1-7):
Issues:

Task 2: Find about the "University honors program" offers at this university.

Expectation rating (1-7):
Experience rating (1-7):
Issues:
**Task 3:** Find the events at Southern Illinois University for the visitors in this month.

Expectation rating (1-7):  
Experience rating (1-7):  
Issues:

---

**Task 4:** Find the tuition fee for Graduate studies.

Expectation rating (1-7):  
Experience rating (1-7):  
Issues:

---

**Task 5:** How many of libraries does Southern Illinois University provide to students for accessing the books?

Expectation rating (1-7):  
Experience rating (1-7):  
Issues:

---

**Task 6:** Find Graduate programs offered in Arts & Design department.

Expectation rating (1-7):  
Experience rating (1-7):  
Issues:
Usability testing 2
Perform the following tasks on Macquarie University website.

**Task 1:** Find about the news room for the media at this university.

Expectation rating (1-7):
Experience rating (1-7):
Issues:

**Task 2:** Macquarie University offers to book a slot for IELTS exam, Find out the amount to be paid for this exam.

Expectation rating (1-7):
Experience rating (1-7):
Issues:

**Task 3:** Find the mail-id of the person to enquire about the postgraduate scholarships.

Expectation rating (1-7):
Experience rating (1-7):
Issues:
**Task 4:** Find about the student events today.

Expectation rating (1-7): 
Experience rating (1-7): 
Issues:

**Task 5:** Find about the disciplines that are offered in research of earth sciences.

Expectation rating (1-7): 
Experience rating (1-7): 
Issues:

**Task 6:** Find the upcoming carrier service events for students to develop their skills.

Expectation rating (1-7): 
Experience rating (1-7): 
Issues: