
Syed Muhammad Ali Shah
Usman Sattar Alvi
This thesis is submitted to the School of Engineering at Blekinge Institute of Technology in partial fulfillment of the requirements for the degree of Master of Science in Software Engineering. The thesis is equivalent to 2 X 20 weeks of full time studies.

Contact Information:
Author(s):
Syed Muhammad Ali Shah
Address: Ronneby, Sweden
E-mail: ali_shah_uet@hotmail.com

Usman Sattar Alvi
Address: Ronneby, Sweden
E-mail: usmanalvi99@hotmail.com

External advisor(s):
Herman Afzeliu
Logica Sverige AB, P. O Box 552, SE-371 23 Karlskrona, Sweden
Phone: T: +46(0)455 32 63 73
M: +46(0)767 76 63 73

University advisor(s):
Dr Cigdem Gencel
School of Computing, BTH

School of Computing
Blekinge Institute of Technology
Box 520
SE – 372 25 Ronneby
Sweden

Internet : www.bth.se/tek
Phone : +46 457 38 50 00
Fax : + 46 457 271 25
ABSTRACT

Software testing is a key phase in software development lifecycle. Testing objectives corresponds to the discovery and detection of faults, which can be attained by utilizing manual or automated testing approaches. In this thesis, we are mainly concerned with the manual test approaches. The most commonly used manual testing approaches in the software industry are the Exploratory Testing (ET) approach and the Test Case Based Testing (TCBT) approach. TCBT is primarily used by software testers to formulize and guide their testing tasks and set the theoretical principles for testing. On the other hand ET is simultaneous learning, test design, and test execution. Software testing might benefit from an intelligent combination of these approaches of testing however there is no proof of any formal process that accommodates the usage of both test approaches in a combination.

This thesis presents a process for Mix Testing (MT) based on the strengths and weaknesses of both test approaches, identified through a systematic literature review and interviews with testers in a software organization. The new process is defined through the mapping of weaknesses of one approach to the strengths of other. Static validation of the MT process through interviews in the software organization suggested that MT has ability to resolve the problems of both test approaches to some extent. Furthermore, MT was validated by conducting an experiment in an industrial setting. The analysis of the experimentation results indicated that MT has better defect detection than TCBT and less than ET. In addition, the results of the experiments also indicate that MT provides equal functionality coverage as compared to ET and TCBT.

Keywords: Test case based testing, Exploratory testing, Mix testing, Process, Experiment
ACKNOWLEDGMENT

In the Name of Allah the Most Merciful and Beneficent

Prophet Mohammad (Peace Be Upon Him) said:

“Strive for knowledge even if you have to travel to China”

First and foremost, we would like to thank Almighty Allah, on blessing us with the strength and courage to successfully complete this thesis work. Secondly we are great full to our parents on their constant and unconditional support, which resulted in the successful conclusion of this research work.

We are very much thankful to our both supervisors Dr. Cigdem Gencel and Herman Afzelius on their encouragement, guidance and support from the very start till the very end. A special thank to Logica AB Sweden and to all the resources, directly and indirectly involved in assisting our thesis work. It was because of their constructive suggestions and valuable directions, which channeled our abilities to complete this thesis work.

Last but not the least we would also like to thank Petter Mattsson, Åsa Augustsson, Adeel Akhtar and Eva Ahlén.

In general, we thank all of our friends for their kindness and motivation during all times.

At last we dedicate our research work to the friendship of both great nations, ISLAMIC REPUBLIC OF PAKISTAN and SWEDEN.
# Table of Contents

ABSTRACT ........................................................................................................... I

ACKNOWLEDGMENT ............................................................................................ II

LIST OF FIGURES ................................................................................................. 4

LIST OF TABLE ..................................................................................................... 5

1 INTRODUCTION ............................................................................................... 6
  1.1 BACKGROUND .......................................................................................... 6
  1.2 PURPOSE ................................................................................................. 7
  1.3 PROBLEM DOMAIN ................................................................................ 7
  1.4 AIMS AND OBJECTIVES ......................................................................... 7
  1.5 RESEARCH QUESTIONS ......................................................................... 8
  1.6 RESEARCH METHODOLOGY .................................................................. 8
  1.7 RESEARCH DESIGN ............................................................................... 9
  1.8 THESIS STRUCTURE .............................................................................. 11

2 BACKGROUND: INDUSTRY PRACTICED MANUAL TESTING APPROACHES ...... 12
  2.1 EXPLORATORY TESTING ....................................................................... 12
  2.2 TEST CASE BASED TESTING .................................................................. 13

3 STRENGTHS AND WEAKNESSES OF MANUAL TEST APPROACHES .......... 15
  3.1 SYSTEMATIC LITERATURE REVIEW .......................................................... 15
    3.1.1 Planning the review ......................................................................... 15
    3.1.2 Review protocol development ......................................................... 16
    3.1.3 Conducting the review .................................................................... 20
    3.1.4 Study quality assessment ................................................................. 23
    3.1.5 Data extraction .............................................................................. 23
    3.1.6 Data synthesis .............................................................................. 23
    3.1.7 Reporting the review ...................................................................... 24
    3.1.8 Data analysis of systematic literature review results ....................... 28
  3.2 INDUSTRIAL INTERVIEWS ....................................................................... 35
    3.2.1 Purpose of interviews ...................................................................... 35
    3.2.2 Selection of interview subjects ....................................................... 36
    3.2.3 Study instruments ......................................................................... 36
    3.2.4 Interview structure ....................................................................... 36
    3.2.5 The analysis and results of the collected data by interviews .......... 38
    3.2.6 Discussion and Summary ................................................................. 42
    3.2.7 Validity threats ............................................................................ 43

4 A PROCESS FOR MT .......................................................................................... 45
  4.1 TEST PROCESS FOR TCBT (ISO/IEC 29119) ...................................... 45
    4.1.1 Test planning .................................................................................. 45
    4.1.2 Test monitoring and control .......................................................... 45
    4.1.3 Test completion Criteria ............................................................... 45
    4.1.4 Test design and implementation .................................................. 45
    4.1.5 Test execution .............................................................................. 45
    4.1.6 Test incident reporting ................................................................. 46
  4.2 SESSION BASED TEST MANAGEMENT PROCESS FOR ET ................. 46
  4.3 THE PROPOSED MIX TESTING PROCESS ........................................... 47
    4.3.1 Mapping ....................................................................................... 48
    4.3.2 Interviews ..................................................................................... 50
    4.3.3 Result ........................................................................................... 51
    4.3.4 Discussion .................................................................................... 52
    4.3.5 Process definition (Based on the result and findings) .................... 53
    4.3.6 Validity threats related to MT process ........................................ 61
5 VALIDATION OF THE PROPOSED MT PROCESS ................................................. 62
  5.1 STATIC VALIDATION ................................................................................. 62
    5.1.1 Selection of subjects ........................................................................ 62
    5.1.2 Study instrument ............................................................................ 62
    5.1.3 Feedback structure ......................................................................... 62
    5.1.4 Result ............................................................................................... 63
    5.1.5 Discussion ....................................................................................... 65
  5.2 DYNAMIC VALIDATION .......................................................................... 66
    5.2.1 Experimentation ............................................................................... 66
    5.2.2 Planning .......................................................................................... 67
    5.2.3 Experiment design ......................................................................... 69
    5.2.4 Experiment execution ...................................................................... 70
  5.3 RESULTS AND ANALYSIS ....................................................................... 71
    5.3.1 Analysis of the direct measurement ............................................... 72
    5.3.2 Discussion and summary .................................................................. 82
    5.3.3 Validity threats ............................................................................... 82

6 EPILOGUE ........................................................................................................ 84
  6.1 CONCLUSION AND DISCUSSION .......................................................... 84
    6.1.1 Answer to the research questions .................................................... 84
  6.2 FUTURE WORK .......................................................................................... 86

7 REFERENCES .................................................................................................... 87

8 APPENDIX ........................................................................................................ 89
  8.1 INTERVIEW QUESTIONNAIRES ............................................................... 89
    8.1.1 Questionnaires related to problems of ET ........................................ 89
    8.1.2 Questionnaires related to benefits of Exploratory Testing ............. 89
    8.1.3 Questionnaires related to problems of Test Case Based Testing .... 90
    8.1.4 Questionnaires related to benefits of Test Case Based Testing ....... 90
    8.1.5 Questionnaires related to MT Process definition ......................... 90
    8.1.6 Feedback of MT process ................................................................. 90
  8.2 TRANSCRIBED INTERVIEWS ................................................................... 91
    8.2.1 Interview 1: .................................................................................... 91
    8.2.2 Interview 2: .................................................................................... 93
    8.2.3 Interview 3: .................................................................................... 95
    8.2.4 Interview 4: .................................................................................... 96
    8.2.5 Interview 5: .................................................................................... 98
  8.3 CONTACT DETAILS OF STATIC EVALUATORS ........................................ 100
  8.4 EXPERIMENT INSTRUMENTS ................................................................. 100
    8.4.1 Background form ............................................................................ 100
    8.4.2 Bug Report template ...................................................................... 101
    8.4.3 Test Case template ......................................................................... 101
    8.4.4 Requirement based Test Cases for Calculator ............................... 101
    8.4.5 Mission Sheet for Mix testing ......................................................... 101
  8.5 MEASUREMENT TABLES .......................................................................... 102
    8.5.1 F-max table ..................................................................................... 102
    8.5.2 F table at (p = 0.05) ...................................................................... 102
    8.5.3 The tabulated value of t ................................................................. 104
LIST OF FIGURES

Figure1: Research Design .................................................................10
Figure2: Perceived strengths of ET from the literature..........................30
Figure3: Perceived strengths of TCBT from the literature .....................32
Figure4: Perceived weaknesses of ET from literature ..........................33
Figure5: Perceived weaknesses of TCBT from literature .......................35
Figure6: Qualitative Data Analysis (QDA) ..........................................38
Figure7: Perceived strengths of ET from interviews ............................39
Figure8: Perceived strengths of TCBT from interviews ..........................40
Figure9: Perceived weaknesses of ET from interviews .........................41
Figure10: Perceived weaknesses of TCBT from interview ....................42
Figure11: TCBT process .................................................................46
Figure12: SBTM process .................................................................47
Figure13: Process for MT ...............................................................54
Figure14: Experiment execution .......................................................70
Figure15: Box plots of defect detection for three treatments ..................73
Figure16: Box plots of functionality coverage for three treatments .............79
**LIST OF TABLE**

Table 1: Research question and relevant methodologies ................................................. 09
Table 2: Online resource utilized during systematic literature review .......................... 17
Table 3: List of selected journals, conferences, unpublished studies and books ........... 22
Table 4: List of selected articles .................................................................................. 22
Table 5: Classification of ET strengths based on similarities ...................................... 29
Table 6: Classification of TCBT strengths based on similarities .................................... 31
Table 7: Classification of ET weaknesses based on similarities .................................. 33
Table 8: Classification of TCBT weaknesses based on similarities .............................. 34
Table 9: Mapping of TCBT weaknesses to the strengths of ET ................................. 48
Table 10: Mapping of ET weaknesses to the strengths of TCBT ............................... 49
Table 11: Variables selection in the final experiment .................................................... 68
Table 12: Experiment design ...................................................................................... 69
Table 13: Detected Defects ......................................................................................... 72
Table 14: Descriptive statistics for defect detection of three treatments .................... 73
Table 15: Relevant statistics for defect detection of three treatments ......................... 75
Table 16: Values of sum of squares for defect detection ............................................ 77
Table 17: Least significance difference for defect detection ....................................... 78
Table 18: Functionality Coverage .............................................................................. 79
Table 19: Descriptive statistics for functionality coverage of three treatments ........... 79
Table 20: Relevant statistics for functionality coverage of three treatments .............. 80
Table 21: Values of sum of squares for functionality coverage .................................... 81
Table 22: Least significance difference for functionality coverage ............................. 81
Table 23: Comparison of total testing time ............................................................... 86
Table 24: Detail of static evaluators .......................................................................... 100
Table 25: Background form ...................................................................................... 100
Table 26: Bug report template .................................................................................. 101
Table 27: Test case Template .................................................................................... 101
Table 28: RBTC Template ....................................................................................... 101
Table 29: Mission Sheet .......................................................................................... 101
Table 30: F max table ............................................................................................. 102
Table 31: F table at p = 0.05 ................................................................................... 103
Table 32: Calculated values of t .............................................................................. 105
1 INTRODUCTION

This chapter provides an insight on the background of the selected research area, problem domain, aims and objectives of the thesis. Further in this chapter readers can also go through the research questions and selected research methodologies for this thesis study.

1.1 Background

Testing amounts to observing the execution of a software system to validate whether the software behaves as intended and identify potential problems [1]. Software testing can be considered as a practical activity that relates with the theoretical, technological, tools and management knowledge [2]. Testing objectives corresponds to the discovery and detection of faults, which can be attained by utilizing manual testing approaches [3].

The results from testing activities are correlated to the performance and the expertise of the human testers involved in manual testing process[3][4][5]. The testing techniques being utilized in software industry, related to manual testing, are the Exploratory Testing (ET) approach and the Test Case Based Testing (TCBT) approach [3].

TCBT is primarily used by software testers to formulize and guide their testing tasks and set the theoretical principles for testing [3][6]. In TCBT test cases are planned and designed prior to the execution of testing, which provides several benefits e.g. test awareness, test coverage, repeatability, and tracking [6][7].

However, some of the studies conducted by practitioners, keeping an industrial perspective, shows that the use of rigorous and well-documented TCBT is not very common [8][9][10]. It is also stated that utilizing and focusing only on TCBT may not reveal and uncover many important facets that can affect manual testing [3]. In [7], it is also mentioned that testers not often rely on the test cases while actually executing them. They are found in applying diverse testing strategies and techniques, which may not be specified in any of predesigned test cases [7].

Itkonen [11] stated that documenting every scenario in a test case could be very time consuming. Tester may spend more time in writing tests as compared to actually executing them. In addition, the actual effectiveness and importance of these pre-designed test cases in terms of defect detection efficiency is also unknown [11]. Agruss et al. and Andersson et al. also highlighted that, if all the pre designed test cases pass in first execution, chances of finding any new bugs by executing the same test set again are nominal [7][9]. Kaner [12] described another limitation of using predesigned test cases as;

"Something may be fundamentally wrong. If so, the program will be redesigned. Creating new test series now is risky. They may become obsolete with the next version of the program. Rather than gambling away the planning time, try some exploratory tests – whatever comes to mind."

The idea behind ET is to conduct testing without use of predesigned test cases [3]. Bach and Kaner et al. concluded that the ET approaches could be considered as quite effective in terms of revealing vital information while at the same time indicating cost-efficiency [13][14]. Another positive aspect of using ET is that, it allows the tester to freely explore the application by utilizing human intuition and experience [15][16]. In [6], the benefits of ET
are stated as; the low reliance on comprehensive documentation, rapid feedback, investigation of particular risks and testing from an end-user viewpoint.

On the other hand, the main concern with ET is the lack of effective risk management. As conducting ET, simultaneous learning and testing on mission critical applications may raise severe concerns such as threat to life and finances [15]. Itkonen et al. also highlighted some shortcomings of the ET approach as; ET does not assure the test coverage, repeatability of defects, oracle mistakes and quality of testing is not visible [3][6][8]. Agruss et al. also stated that ET is not suitable while performing acceptance testing [7].

ET is not a replacement of TCBT, as more technical defects can be found using TCBT [3]. Test cases guide the tester to pay attention on more function based areas, which may result in ignoring some of the suspicious areas of the system [3][6]. On the contrary ET makes better use of tester’s creativity and skills to discover the bugs, which TCBT may not uncover [3][6]. Agruss et al. highlighted that both approaches are complementary to each other as in many situation there are ethical and legal issues which may emphasis on having boths TCBT and ET [7]. Copeland also states in [17], that ET can be effectively utilized when the TCBT is not able to detect defects.

Software testing will benefit from an intelligent combination of various test approaches. It can be done by ideal proportions and implementation of a good test strategy [7]. This strategy will provide better defect detection efficiency and confidence for customer and management needs [16]. In [16], it is also stated that mix of TCBT and ET approaches can be very effective. In some situation, testing objectives can be better achieved from ET, while in other situation benefits may be attained by use of TCBT.

1.2 Purpose

The purpose of this thesis work is to develop and validate a process for MT. The proposed process is based on the identified strengths and weaknesses of the two commonly used manual test approaches; i.e. TCBT and ET. The MT process is validated in an industrial setting.

1.3 Problem domain

Itkonen et al. speculate that manual testing practices are not often studied, which greatly impact the effectiveness of manual testing [3]. In addition, they also mentioned in their paper the need of effective test practice for the manual testing techniques [3]. In [11], it is concluded that more related research is needed in order to get better understanding of all manual testing activities being practiced in software development companies.

Many practitioners and researchers highlight use of the MT, however literature lacks in providing the research on the effective utilization of MT [8]. Furthermore, there is no evidence of empirical research found for the utilization of MT, which raise concerns about the empirical evaluation of claimed effectiveness of it.

1.4 Aims and objectives

The aim of this study is to develop and validate the usage of TCBT and ET as a MT by defining a process for it.
Based on the weaknesses and strengths of ET and TCBT, a process is defined in which both test approaches are integrated in such a way that they can complement each other. Furthermore, the defined MT process was validated by comparing it to ET and TCBT. This comparison was made by conducting an experiment using industry professionals as test subjects.

The major objectives of this thesis study are:

- Identifying the weaknesses related to the TCBT and ET
- Identifying the strengths related to the TCBT and ET
- Defining a process for MT
- Validation of MT process

1.5 Research Questions

A statement that depicts the reason of conducting the research is known as a research question [18]. Four research questions are proposed for this thesis work.

RQ1: What are the weaknesses of TCBT and ET?

The answer to this research question highlights the potential weaknesses related to TCBT and ET keeping an industrial context in focus.

RQ2: What are strengths of TCBT and ET?

The answer to this research question highlights the strengths of using TCBT and ET keeping an industrial context in focus.

RQ3: How can a mixed process for MT be defined so that it would address the weaknesses of TCBT and ET and incorporate the strengths of these approaches?

A mixed process is defined which addresses the main issues of both testing approaches in such a way that it also incorporates the strengths of both approaches.

RQ4: How effective is the proposed MT process in comparison to individual testing approaches?

- RQ4.1: How effective is the MT in terms of defect detection as compared to TCBT and ET?
- RQ4.2: How effective is the MT in terms of functionality coverage as compared to TCBT and ET?

The effectiveness of the MT is measured based on total number of detected defects and total number functionality coverage.

1.6 Research Methodology

Creswell [18] defines research as a study that goes beyond the influences of personal ideas and experiences of an individual. A researcher’s work is primarily based on the utilization of some research methods and techniques. Creswell describes three types of methods used for research i.e. Qualitative, Quantitative and Mixed research.
In our research study, we selected both qualitative and quantitative research approaches. The answer to each research question is associated with proper selection of research methods. Two qualitative methods were used for data collection in order to answer RQ1, RQ2 and RQ3. The strengths and weaknesses were identified through a systematic literature review and by conducting interviews with testers in a software company. The inspiration behind selecting these qualitative methods is that it provides broader picture of resolution toward the identified problems. RQ4 is answered through qualitative and quantitative approaches. The static validation of the MT process was performed through interviews and the dynamic validation by conducting a controlled experiment.

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Methodology</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ1</td>
<td>Interview / Systematic Literature Review</td>
</tr>
<tr>
<td>RQ2</td>
<td>Interview / Systematic Literature Review</td>
</tr>
<tr>
<td>RQ3</td>
<td>Interview</td>
</tr>
<tr>
<td>RQ4</td>
<td>Feedback/ Experiment</td>
</tr>
</tbody>
</table>

Table 1: Research question and relevant methodologies

1.7 Research Design

The selected thesis topic is related to the development and validation of a MT process. Both qualitative and quantitative approaches of research were used in order to effectively obtain study results. The stages involved in the study process are shown in the Figure 1 below.
In order to conduct the research work industry problems were identified related to TCBT and ET by conducting initial meetings with industry professionals. It was identified that there exists no empirically validated MT. Furthermore it was also highlighted by the practitioners that there exists no formal process for using MT.

The strengths and weaknesses related to each testing approach were identified by conducting the interviews of several industry professionals and by systematic literature review. These identifications assisted in resolving the identified problems in process definition for MT. To further support and improve the process definition, interviews were conducted again with the professionals in order to incorporate their valuable suggestions and experiences in the process. Later, MT was dynamically validated by conducting an experiment on industry professionals to assess the effectiveness of MT with respect to each of the manual approaches; i.e. TCBT and ET.
1.8 Thesis Structure

Chapter 1 (Introduction): This chapter highlights about the selected problem domain, purpose of study, research aims, objectives and adopted research methodologies which will be utilized in this study.

Chapter 2 (Background on Industry Practiced Manual Testing Approaches): This chapter defines the concepts related to ET and TCBT.

Chapter 3 (Strengths and Weaknesses of Manual Test Approaches): This chapter gives details about the strengths and weaknesses associated with the use of both testing approaches i.e. ET and TCBT as specified in literature and as identified by interviewees.

Chapter 4 (A Process for MT): This chapter provides the details of currently used test processes related to ET and TCBT. Furthermore, it defines the process of MT based on the identified weaknesses and strengths of both testing approaches.

Chapter 5 (Validation of the Proposed MT Process): This chapter provides with the static validation of the proposed process based on the feedbacks of industry professionals. Further in the chapter it provides detail about the experiment design and variables that are necessary to conduct the experiment in order to validate MT process. In the end of chapter results are analyzed, interpreted, packaged and presented.

Chapter 6 (Epilogue): This chapter presents the study conclusion along with the suggestions for future work.

Chapter 7 (References)

Chapter 8 (Appendix)
2 BACKGROUND: INDUSTRY PRACTICED MANUAL TESTING APPROACHES

In this chapter, we provide an overview of the industry-practiced manual testing approaches. We describe ET and TCBT in order to provide with the basic understanding of both approaches.

2.1 Exploratory Testing

The term exploratory testing has recently gained tremendous popularity especially amongst the league of testers, consultants and practitioners [8][11]. ET is also referred to as ad hoc testing [16]. According to Software Engineering body of Knowledge (SWEBOK) [20], ad hoc testing is widely being used by the testers. Since the literal meaning of ad hoc may correspond to the sloppy and careless work, in early 1990s, a test methodologist group introduced a new term exploratory testing instead of ad hoc [16]. Testers have been practicing ET consciously or unconsciously in industry [8][21]. Furthermore, Kaner et al. highlights wide utilization of ET approach in the area of software testing [21]. Detailed elaboration of ET is provided in the below section.

The definition of ET as described in the SWEBOK is:

“Exploratory testing is defined as simultaneous learning, test design, and test execution; that is, the tests are not defined in advance in an established test plan, but are dynamically designed, executed, and modified”

Bach, also proposed a definition of ET “Exploratory testing is simultaneous learning, test design, and test execution”

Tinkham defined ET as:

“Any testing to the extent that the tester actively controls the design of the tests as those tests is performed and uses information gained while testing to design new and better tests.”

Itkonen et al. described ET as a testing approach that is well suited in finding the defects and put less stress on documenting tests. Defect detection is the key purpose of ET and documenting the outcomes of testing is of more importance than planning and writing down tests beforehand [8].

Itkonen et al. highlighted following properties of ET [8]:

- No definition of tests in advance, ET is performed without any predefined steps or instructions.
- ET is directed by the previous test results and knowledge. An ET tester can gain knowledge by the use of any available information source in order to effectively execute tests.
- ET focuses on discovering defects by pure exploration, instead of using detailed test cases.
- ET relates to simultaneous learning and at the same time executing tests on the application under test.
Effectiveness of ET is correlated with the possession of tester’s skills, knowledge and experience.

Bach, Kaner, Marick, Hendrickson, Agruss and Johnson have highlighted some common attributes of ET summarizing the results of workshop on software testing (LAWST VII) [22] as:

- Interactive
- Concurrence of cognition and execution
- Creativity
- Drive towards fast results
- De-emphasize archived testing materials

The overall structure of ET is quite easy to describe. A tester involved in a testing process, interacts with an application in order to accomplish a testing mission, which is to uncover bugs and later reports the results. The very basic external elements associated with ET [16] are as follows:

- Time
- Tester
- Product
- Test Mission
- Reporting

Effective ET testing is highly dependent on the abilities and experience of the testers, and it is considered as the martial art of the mind [16][20]. The success and failure of ET is related to the distinction of an excellent tester from an amateur one, some of the basic characteristics of a good ET tester [16] are as follows:

- Test designer
- Careful observer
- Critical thinker
- Possess diverse ideas
- Well versed in test resources

### 2.2 Test Case Based Testing

Test case based testing is a traditional method of testing in which all the right set of tests are defined and planned prior to the execution of testing in conjunction with the expected results [23]. The idea behind TCBT is to design and document test cases that cover all the inputs, outputs and other functionalities of the system to be tested [6][7][11]. According to SWEBOK, TCBT is defined as designing of the test cases to validate the correct implementation of functional specifications, which can also be referred to as conformance, correctness or functional testing [20].

Testers utilize TCBT to formulize and document their testing tasks. The creation of the test cases depend on the level of testing required to be performed. And these test cases should include the expected results [20].

According to Institute of Electrical and Electronics Engineering (IEEE) standard 829-1998 [24], TCBT comprises of documented tests containing actual values to be used as input along
with the pre anticipated outputs. A test case also categorizes the constraints, which may affect the test procedures associated with the use of specific test case [24].

ISO/IEC 29119 describes the following structure of a test case [25].

- Precondition for executing a test
- A set of test inputs (values, actions etc)
- Expected Results (Outputs, post conditions)
- Compliance with specific requirements
3 STRENGTHS AND WEAKNESSES OF MANUAL TEST APPROACHES

In this chapter, we discuss the strengths and weaknesses of industry practiced manual test approaches based on the results of a systematic literature review and industrial interviews we conducted.

3.1 Systematic Literature Review

Systematic literature review can be defined as a means by which all the available and relevant research material is identified, evaluated, and interpreted in order to answer a research question or a topic of interest [26]. The individual studies, which contribute in any way to a systematic literature review, are referred to as primary studies. Systematic literature review is considered as a secondary study [26].

There are some basic attributes associated with the systematic literature review, and their significance cannot be overlooked throughout the process of research. The three phases of systematic literature review are as follows [26]:

- Planning the review
- Conducting the review
- Reporting the review

In the first phase, the need of performing the review is identified along with the development of a review protocol. A review protocol is considered as the guidelines of searching for a complete systematic literature review process [26].

The second phase revolves around the following [26]:

- Identification of research
- Selection of primary studies
- Study quality assessment
- Data extraction and monitoring
- Data synthesis

In the last and final phase, reports are generated which can be in the form of a research report or a thesis etc. based on the results of the systematic literature review.

3.1.1 Planning the review

3.1.1.1 Identifying the need of systematic literature review:

The main aim of this systematic literature review is to gather and summarize the existing evidence and research related to the strengths and weaknesses of ET and TCBT during the period of 2000 to 2009. The main reason of conducting the systematic literature review in this specified time period was to get an overview of the latest research carried out on ET and TCBT. Another reason was that a formal process of ET was introduced in the year 2000,
which made us assume that the significant works would be published in this time frame. In addition, any gap related to the current study is suggested for further investigation.

3.1.2 Review protocol development

Review protocol is a detailed plan for conducting a systematic literature review and provides a method for selecting primary studies thereby reducing biasness [26].

3.1.2.1 Search strategy

The search strategy for this research is primarily based on online searching. The search string and the relevant resources utilized for this search are listed as below:

3.1.2.1.1 Search strings

Following search strings were used to extract the required and relevant primary studies.

1. Manual test approaches
2. Exploratory testing
3. Ad hoc testing
4. Test case based testing
5. Scripted testing
6. TCBT
7. Exploratory testing AND weakness
8. Exploratory testing AND complexities
9. Exploratory testing AND shortcomings
10. Exploratory testing AND problems
11. Exploratory testing AND issues
12. Exploratory testing AND strengths
13. Exploratory testing AND efficiency
14. Exploratory testing AND benefits
15. ET AND weakness
16. ET AND complexities
17. ET AND shortcomings
18. ET AND problems
19. ET AND issues
20. ET AND strengths
21. ET AND efficiency
22. ET AND benefits
23. Test case based testing AND weakness
24. Test case based testing AND complexities
25. Test case based testing AND shortcomings
26. Test case based testing AND problems
27. Test case based testing AND issues
28. Test case based testing AND strengths
29. Test case based testing AND benefits
30. Test case based testing AND efficiency
31. Scripted testing AND weakness
32. Scripted testing AND complexities
33. Scripted testing AND shortcomings
34. Scripted testing AND problems
35. Scripted testing AND issues
36. Scripted testing AND strengths
37. Scripted testing AND benefits
38. Scripted testing AND efficiency
39. TCBT AND weaknesses
40. TCBT AND complexities
41. TCBT AND shortcomings
42. TCBT AND problems
43. TCBT AND issues
44. TCBT AND strengths
45. TCBT AND efficiency
46. TCBT AND benefits

3.1.2.1.2 Resources utilized

The software engineering search engines that are currently available are not sufficient in supporting systematic literature reviews [27]. Hence for that reason software engineering researchers are bound to perform searches, which are more response dependent.

Brereton et al. [27] identified seven relevant sources related to software engineers:

- IEEE Xplore
- ACM Digital library:
- Google scholar (scholar.google.com)
- Citeseer library (citeseer.ist.psu.edu)
- Inspec (www.iee.org/Publish/INSPEC/)
- ScienceDirect (www.sciencedirect.com)
- EI Compendex (www.engineeringvillage2.org/Controller/Servlet/AthensService).

The online resources, which were utilized during the systematic literature review, are as follows:

<table>
<thead>
<tr>
<th>DATA SOURCE</th>
<th>DATABASES</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIGITAL LIBRARY</td>
<td>NAME OF DATABASE:</td>
</tr>
<tr>
<td></td>
<td>• IEEE XPLORER</td>
</tr>
<tr>
<td></td>
<td>• ACM DIGITAL LIBRARY</td>
</tr>
<tr>
<td></td>
<td>• SPRINGER LINK</td>
</tr>
<tr>
<td></td>
<td>• ENGINEERING VILLAGE</td>
</tr>
<tr>
<td>ONLINE SEARCH ENGINES/DATABASES</td>
<td>• GOOGLE SCHOLAR</td>
</tr>
<tr>
<td></td>
<td>• ISI</td>
</tr>
<tr>
<td></td>
<td>• SCOPUS</td>
</tr>
</tbody>
</table>

Table 2: Online resource utilized during systematic literature review
3.1.2.2 Study selection criteria

The study selection for this research was based on the following criterion of inclusion and exclusion.

3.1.2.2.1 Inclusion criteria for study

The articles and research papers that were included for investigation in this research study lie between the dates of 1st January 2000 to 31st December 2009. The suitability of these articles were judged and assessed on the basis of following inclusion criteria:

1. The articles/research papers are considered, if their full text is available and accessible.
2. The article/research papers are considered, if cross-reviewed by at least one reviewer.
3. The types of articles or research papers, which are experiments, case studies, expert reports, surveys and comparative analysis reports.
4. The article/research papers are considered, if they provide with strengths and weaknesses or any other sort of relevant information related to TCBT and ET.
5. The articles/research papers are considered, if they provide any sort of comparative analysis of both the test approaches i.e. ET and TCBT either by qualitative or quantitative means.

3.1.2.2.2 Exclusion criteria for study

The research article(s) that did not correspond to the inclusion criteria as specified above were not considered for the current study.

3.1.2.3 Study selection procedure

Following approach was followed in the selection procedure for the current study. The approach that was adopted for the study selection procedure is to first study the following sections of the research articles:

• Titles of the article
• Abstract of the article
• Conclusions of the article

If the sections above corresponded to the inclusion criteria as mentioned above, then these articles were further read and investigated in detail.

3.1.2.4 Study quality assessment checklist and procedures

The research articles selected as a primary study were evaluated on the basis of their structure i.e. introduction section, method used for carrying out research, gathered results, analysis and the conclusion section. Following checklist was prepared which guided in evaluating each section of the research article.
**Introduction Section:** Does the introduction section provide some overview of the relevant topic of interest i.e. ET and TCBT?

**Method used for carrying out research:** Is it clearly specified in the research article about the adopted research methodology? And is it suitable for our research study?

**Gathered Results:** Does the research article completely specify the results of the study? Are these results suitable in the context of our research topic? Is there any validity threats associated with the research article?

**Analysis:** How was the data evaluated and analyzed in the research article?

**Conclusion section:** How relevant is the conclusion given in the research article? And to what extent the conclusion is relevant to our research study? Whether the conclusion also discusses about the limitations and restrictions of the research study and report negative or positive results as well?

### 3.1.2.5 Strategy used for data extraction

Data from the selected primary research articles was extracted by using forms. If there was no explicit information in the primary research article related to the research topic such as study environments etc., the data was inferred on the basis of its context. This gathered data was then validated for its correctness by the internal/external supervisor. Data extracted was primarily based on general and specific information as described below:

#### 3.1.2.5.1 General information

The general information of the relevant research articles was documented as follows:

- Title of the Article
- Name of Author(s)
- Name of Conference/Journal/ Date of Publish/Presented
- Relevant Search String(s) utilized to retrieve research article
- Database used to retrieve the research article
- Date of Publication

#### 3.1.2.5.2 Specific information related to research article

**Study environment of research article:**

- Industrial
- Academia
- Consultant report
- Licentiate thesis

**Research Methodology utilized in primary study:**

- Experiment
- Case Study
- Survey
- Field observation
- Interviews
Study participants in a primary study:

- Researchers
- Industry professionals
- Students
- Total number of participants

Relevant area of research study:

- Exploratory testing (ET)
- Test case based testing (TCBT)
- Weaknesses of ET
- Strengths of ET
- Strengths of TCBT
- Weaknesses of TCBT
- Comparison of both test approaches

3.1.2.6 Synthesis of the extracted data

In the data synthesis phase, results of the selected primary studies were collected and summarized. The primary studies that are distinct from each other with respect to the outcomes and research methodologies are referred to as heterogeneous studies [26]. As the nature of the extracted data from the primary studies was mostly heterogeneous, proposition of qualitative synthesis is appropriate for this study. In the qualitative synthesis, the research articles were analyzed in detail and the relevant results were documented across the appropriate research questions. Data from each primary study was extracted by using forms in order to obtain information.

3.1.2.7 Validation of a review protocol

The review protocol is one of the most important elements of a systematic literature review. It is very important to make the validation process transparent. In [26], it has proposed that pilot searches should be carried out in order to identify primary studies by using the defined search strings as defined in review protocol. The thesis supervisor for this thesis study verified the review protocol. In addition, search strings and resources were also verified and validated by undertaking help from BTH librarian.

3.1.3 Conducting the review

The steps, which were performed in conducting this systematic literature review, are discussed in the following sub-sections.

3.1.3.1 Identification of research

The aim of the systematic literature review is to find the maximum number of studies as possible by utilizing a search strategy, which may have any relevance to the research questions of this thesis study [26]. The search strategy has been explicitly defined in the review protocol. These search strings were defined on the basis of research questions. A general approach is to break down a research question into more individual facets as follows:
• Study design
• Strengths
• Weaknesses
• Intervention
• Comparison
• Outcomes
• Context

On the basis of the above facets of a research question; more abbreviations, alternative names and synonyms can be deduced which can facilitate the search strategy. Other relevant search terms can be extracted by observing different headings in a research article or journal etc. Hence, a more sophisticated and well defined search string can be defined by using ANDs/Ors Boolean operators.

The search conducted for this research study was based on the search strings as defined in the review protocol in order to look for the relevant research material in different online and electronic resources.

• The search strategy adopted is iterative in nature.
• The search strings are verified by conducting trail searches.
• A preliminary search is carried out in order to identify the relevant literature.
• Search is carried out by trying different combinations of search strings derived from the research questions.

Search for the primary study was carried out by using digital and online libraries, but in order to be more specific in the systematic literature review, other resources were also consulted i.e. manual resources, books etc. In addition, some company articles as of satisfies (www. satisfice.com) and grey literature i.e. technical and work in progress reports etc. were also consulted.

Furthermore “Zotero” reference management tool was used to manage and keep the track of all the references for primary studies. All the details of each article were saved in Zotero.

3.1.3.2 Selection of the primary studies

The selection of the primary study had two main steps. In the first step, the title, abstract and the conclusion of the research article was studied to decide upon its relevance for the research study. In second step, inclusion and exclusion criteria was applied on the articles which have been selected in the first step. In the table below, relevant conferences and journals were selected for the primary studies. Any conflict and ambiguity in the selected articles was resolved by having a mutual discussion and by further consulting the supervisor.

<table>
<thead>
<tr>
<th>JOURNALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirical Software Engineering</td>
</tr>
<tr>
<td>Empirical Software Engineering and Measurement</td>
</tr>
<tr>
<td>ACM SIGSOFT Foundation Of Software Engineering</td>
</tr>
<tr>
<td>Software Testing, Verification and Reliability</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CONFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Future of Software Engineering, 2007</td>
</tr>
<tr>
<td>IEEE International Multi topic Conference, INMIC 2009</td>
</tr>
<tr>
<td>IEEE International Conference on Software Engineering, ICSE 2009</td>
</tr>
</tbody>
</table>
In this systematic literature review, 100 articles were scanned and 19 were selected. The list of these articles is given below in the Table 4:

### Table 3: List of selected journals, conferences, unpublished studies and books

<table>
<thead>
<tr>
<th>No.</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Software Testing Research: Achievements, Challenges, Dreams</td>
</tr>
<tr>
<td>2</td>
<td>Workflow-Based Testing Process Management of Software Project</td>
</tr>
<tr>
<td>3</td>
<td>Defect Detection Efficiency: Test Case Based vs. Exploratory Testing</td>
</tr>
<tr>
<td>4</td>
<td>How do testers do it? An exploratory study on manual testing practices</td>
</tr>
<tr>
<td>5</td>
<td>Verification and validation in industry - a qualitative survey on the state of practice</td>
</tr>
<tr>
<td>6</td>
<td>Exploratory testing: a multiple case study,” Empirical Software Engineering</td>
</tr>
<tr>
<td>7</td>
<td>Impacts of the Organizational Model on Testing: Three Industrial Cases</td>
</tr>
<tr>
<td>8</td>
<td>An empirical evaluation of the influence of human personality on exploratory software testing</td>
</tr>
<tr>
<td>9</td>
<td>Guide to the Software Engineering Body of Knowledge</td>
</tr>
<tr>
<td>10</td>
<td>IEEE Standard for Software Test Documentation</td>
</tr>
<tr>
<td>11</td>
<td>Experiments on the test case length in specification based test case generation</td>
</tr>
<tr>
<td>12</td>
<td>Maintaining and evolving GUI-directed test scripts</td>
</tr>
<tr>
<td>13</td>
<td>A preliminary survey on software testing practices in Australia</td>
</tr>
<tr>
<td>14</td>
<td>Factors affecting software testing time schedule</td>
</tr>
<tr>
<td>15</td>
<td>Experimental assessment of manual versus tool-based maintenance of GUI-directed test scripts</td>
</tr>
<tr>
<td>16</td>
<td>Test Case Prioritization for Black Box Testing</td>
</tr>
<tr>
<td>17</td>
<td>An empirical study of regression testing techniques incorporating context and lifetime factors and improved cost-benefit models</td>
</tr>
</tbody>
</table>

### 3.1.3.3 Selected articles
Pre defined search string was used to search for different articles, journals and databases. Special care was taken in not missing any relevant research article. However after conducting detailed study some research articles were rejected on the basis of non-relevant research material. For example while searching Session based test management (SBTM) process, other articles concerning user session based testing of web application were displayed, which had no relevance to the current systematic literature review and also to the topic of interest.

3.1.4 Study quality assessment

Quality assessment was performed on the selected primary study research articles as mentioned above in the review protocol Section 3.1.2.4.

3.1.5 Data extraction

In this phase, data extraction forms were designed and piloted after the finalization of review protocol and the purpose of these forms was to document and gather the extracted data from the primary studies. This assisted reader in extracting the relevant data from the primary study and reduced the chances of any biased behavior. All the extracted data was dually cross-checked in order to minimize the chances of missing any important information.

3.1.6 Data synthesis

Collecting and summarizing the results of the primary studies is referred to as data synthesis [26]. Extracted data is synthesis in such a manner that it provides the answers to the relevant research questions. The data synthesis can be descriptive and it can also be complemented by quantitative summary. There are some other forms of data synthesis such as qualitative and quantitative synthesis.

In descriptive synthesis, the extracted data of primary studies was presented in a consistent manner in order to answer the research question. The gathered results of the outcomes may be homogenous or heterogeneous. Hence, tables should be created in order present the similarities and differences between the outcomes of the primary studies in terms of study type, quality of study and sample size.

3.1.6.1 Quantitative synthesis

In quantitative synthesis the results of the studies are integrated based on the following criteria:

- Sample size intervention
- Estimated size of effect for every intervention
- Standard errors of every iteration
- Difference between mean values of each iteration
- Measuring unit used for effect measurement
3.1.6.2 Qualitative synthesis

In primary studies an article may have different language, terms and concepts having different meanings. The purpose of qualitative synthesis is to integrate results and conclusions generated by such kind of studies [26].

There are three qualitative data synthesis approaches [26]:

- Reciprocal translation
- Refutational synthesis
- Line of argument synthesis

The line of argument synthesis approach was selected as it infers the information which is more relevant to our scope of study and covers most the aspects regarding the strengths and weaknesses of each test approach. Following steps of line of argument were followed for this study:

- Analysis of the individual studies
- Analysis of the whole set of individual studies

3.1.7 Reporting the review

This is a single stage or a phase in systematic literature review. The results generated by performing systematic review are reported in this phase based on the research question. In this study, the data in the systematic literature review was extracted and gathered by the use of extraction forms. This data was then synthesized by using an appropriate synthesis approach and in the end these results were reported. In the following sub-sections we provide the results we obtained through this systematic literature review.

3.1.7.1 Strengths of ET

This section discusses about the strengths associated with the use of ET as stated in the literature.

ET is the next step, whenever test cases failed to discover bugs and it is not further known about what the next test should be. In other cases ET can be very useful because a tester may want to go beyond TCBT or the most apparent test cases. ET is considered more effective in terms of discovering defects as compared to TCBT. In [3], it is concluded that ET is more efficient in discovering fewer bogus defects than TCBT and is also quite useful in indentifying problems which are hard to detect [3][16]. Some more benefits of ET[6][12] [13][14] [15][16][21] are listed below:

- Rapid feedback on a new product or feature
- Quick learning of any new product
- Diversify the testing
- Identification of critical bugs in the shortest possible time
- Cross checking the work of another tester
• Investigation and isolation of any defect
• Low reliance on comprehensive documentation
• Cost effective testing
• Free exploration of application by tester
• Systematic utilization of tester’s skills
• Simultaneous learning and testing
• Efficient in terms of defect detection
• Adapts well to project state

Apart from the above strengths, the use of this approach can be reasonably beneficial in any of the following situations [7][16][23].

• Improvising on scripted tests
• Interpreting vague test instructions
• Product analysis and test planning
• Improving existing tests
• Writing new test scripts
• Regression testing based on old bug reports
• Identifying missing tests
• The behavior of the system cannot be predicted as project is being developed over time
• Investigating a particular risk in order to identify the need of performing TCBT in that particular area.

3.1.7.2 Strengths of TCBT

This section discusses the strengths associated with the use of TCBT as stated in the literature.

TCBT can be effectively utilized if proper test adequacy criteria are formulated. Test adequacy is considered as a strong foundation of TCBT i.e. coverage etc [11]. Test cases if properly designed represent the exact requirements of the system and hence strengthen the testing adequacy in some scenarios [28]. If a test case is properly designed it can provide better and reliable results. This ensures better performance and functional behavior of the system [29]. Furthermore, effective test case utilization may lower costs associated with productivity, testability and scheduling [29]. Conducting TCBT improves testing quality and
depicts the overall picture of perceived quality [30][31]. Proper planning and designing of
tests may provide with many other benefits besides defect detection efficiency such as test
coverage, repeatability, oracle and tracking etc [3].

Some common strengths of TCBT highlighted by researchers [3][9][11][23][29][31][32] are
listed as below:

- TCBT provides explicit oracles for validation of the expected output against actual
  output of the function.
- TCBT fits well where legal and regulatory requirements are needed to be addressed,
  and thus documenting quality management work becomes mandatory such as in
  medical industry, aerospace and other safety critical systems.
- TCBT suites well where complex relationships of a function in software are required
to be tested.
- TCBT provides detailed level of information to the tester in order to carry out
effective testing.
- Proper designing of test cases may ensure reliability, less time in maintenance and
  executing.
- Test cases positively impact customers while conducting acceptance test.
- Test cases give a better chance to analyze the system specification from diverse
  angles.
- It is relatively easy to repeat the same tests in TCBT.
- Any tester in TCBT can execute the tests as it is not strictly related to any particular
  tester.
- Quality of the test cases can be validated.
- Early estimation of software quality can be carried out.
- Test cases provide different and diverse interpretations of functional specification.
- Bugs can be easily reproduced and checked for proper bug fixing by reusing and
  repeating the test cases.
- Test cases provide with most of the test conditions to be executed along with the
  expected outcomes.
- Test case metrics can be utilized for the prediction of reliability in terms of software
  quality.

3.1.7.3 Weaknesses of ET

This section discusses the weakness associated with the use of ET as stated in the literature.
Itkonen et al., Arguss et al., and Shoaib et al. highlighted some shortcomings of ET approach in their research [6][8][7][15].

- Difficult to prioritize and select the appropriate tests, monitoring and keeping track of the progress related to testing tasks. In most of the cases it is quite hard to assess whether all new functionalities and features are tested. It is also difficult to plan and manage test coverage.

- As ET primarily relies on the capability and skills of the software testers, the quality of testing is not known. It is also quite hard to evaluate the tests as in ET no test design is created.

- Repeatability is another issue, because once a defect is located and reported back it becomes challenging to re-perform all the steps for effective verification as testers freely explore different features.

- Lack of effective risk management.

- ET does not assure oracle mistakes.

- ET is not suitable for acceptance, performance and release testing.

- ET is not an effective way of confidence builder especially when coverage, in terms of breadth and depth, are required to be demonstrated.

- A part from identifying a problem, investigating and isolating the actual cause of the problem may take longer time in ET.

- Highly situational approach.

- Traceability issues.

- Less accountable and auditable.

### 3.1.7.4 Weaknesses of TCBT

This section discusses the weaknesses associated with the use of TCBT as stated in the literature.

The testers define the test cases. However, studies conducted in the industry shows that test cases are not often rigorously documented keeping an industrial setting in context. In [9][10], it is highlighted that practitioners face and report many difficulties related to the detailed level designing of test cases. Furthermore the benefits of using TCBT are perceived to be quite less by industry professionals.

In [9], it is highlighted that there exists many problems in executing TCBT. One of the common observed problems of TCBT is the revision of test cases on the basis of changed objectives [11]. According to SWEBOK, conducting TCBT even on the simplest program can be very exhaustive, and in order to do so it could take months or years to actually execute [20]. Itkonen [11] stated that the defect detection efficiency of these pre designed and documented test cases is also not known.
Some common problems highlighted by some researchers [29][30][31][33][34][32][35] are listed as below:

- TCBT is found to be exhaustive and protracted.
- Test cases developed once are not sufficient for the entire system life cycle.
- Reusability and maintenance of the test cases can be quite expensive.
- Prioritizing test cases is considered as difficult in nature.
- Reusing, maintenance, identifications and collection of test cases takes lot of human resources.
- Reusing and changing the test cases rapidly, affect the time constraints.
- Durability of test cases is not known, as new test cases are designed every time for a new change.
- Redesigning the test cases under time constraints and pressure can lead to less sophisticated design.
- Test cases are human prone and it requires necessary skills and experience to update and understand them.
- TCBT often overruns the assigned budget and time.
- Test cases are directly derived from the test plan and if test plan is erroneous, resulting test cast cases will have no or very less effectiveness.
- TCBT is not suitable for regression testing as test cases may not state the problems, which occurred during bug fixing.
- In TCBT, testing is highly dependent on the test case, hence quality of the test cases is not known until their execution. Poorly designed test cases lack precise measure of quality metrics and success.
- Redesigned test cases are not shipped with the software, so in general they are less sophisticated then old ones.

3.1.8 Data analysis of systematic literature review results

In this section, RQ1 and RQ2 are answered. In order to answer RQ1 and RQ2, most of the relevant research material that was available in different literature was studied and important points were noticed as discussed above. Later these points were labeled as strengths and weaknesses of ET and TCBT in general. Further classification of these labeled points was done assigning each to different categories such as planning, test coverage, defect detection efficiency, skills, etc.

Three primary goals were focused in thinking process, developing a sense out of each classification, observation of special patterns within a classified collection or even outside a classification in order to sketch any pattern out of it. The last goal is to discover about any
facts and figures, which may have any correspondence to the research study. The results are presented in a graphical format below in this section.

To identify the strengths and weaknesses of ET and TCBT, a comprehensive systematic literature review was carried out in order to provide partial answers to the research questions RQ1 and RQ2. In the below graphical representations the summary is provided of the most and least occurring strengths and weaknesses associated with ET and TCBT.

The strengths and weaknesses associated with ET and TCBT, which have been identified in the literature, have been classified into a number of more generic attributes.

The strengths of ET as presented in section 3.1.7.1 are classified according to the similarities amongst them. The Table 5 below shows the classifications across distinct and relevant strengths of ET.

<table>
<thead>
<tr>
<th>Attributes</th>
<th>The Strengths of ET</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Defect detection efficiency</td>
</tr>
<tr>
<td></td>
<td>1. Less bogus defects</td>
</tr>
<tr>
<td></td>
<td>2. Identification of critical bugs</td>
</tr>
<tr>
<td></td>
<td>3. Investigation and isolation of any defect</td>
</tr>
<tr>
<td></td>
<td>4. Better regression testing</td>
</tr>
<tr>
<td></td>
<td>Cost effectiveness</td>
</tr>
<tr>
<td></td>
<td>1. Rapid feedback</td>
</tr>
<tr>
<td></td>
<td>2. Quick learning</td>
</tr>
<tr>
<td></td>
<td>3. Cross checking the work</td>
</tr>
<tr>
<td></td>
<td>4. Low reliance on comprehensive documentation</td>
</tr>
<tr>
<td></td>
<td>5. Maintenance is quite easy</td>
</tr>
<tr>
<td></td>
<td>Flexibility in test analysis</td>
</tr>
<tr>
<td></td>
<td>1. Tester can freely explore</td>
</tr>
<tr>
<td></td>
<td>2. More time In actual testing of product</td>
</tr>
<tr>
<td></td>
<td>3. Simultaneous learning and testing</td>
</tr>
<tr>
<td></td>
<td>4. Improvising on scripted tests</td>
</tr>
<tr>
<td></td>
<td>5. Interpreting vague test instructions</td>
</tr>
<tr>
<td></td>
<td>6. Product analysis and test planning</td>
</tr>
<tr>
<td></td>
<td>7. Diversify testing.</td>
</tr>
<tr>
<td></td>
<td>8. Systematic utilization of testers skills</td>
</tr>
</tbody>
</table>

Table 5: Classification of ET strengths based on similarities

According to the Table 5 in the below figure total of 12 references were cited for the strengths of ET, 67% of the references highlights that ET has better defect detection efficiency. 50% of the references posed ET as cost efficient test approach and in addition it also provides better test analysis.
It was observed that the ‘defect detection efficiency’ is one of the strengths of ET which has been highlighted through the research methods like experiment, case study and field observation. In addition, it is worth noticing that out of 67% of the cited references \[3\][6][8][11][12][14][16][21], 50% were proposed by authors in referenced books which are based on their own personal experiences and observations.

The strength ‘cost effectiveness’ is highlighted in cited references \[6\][8][13][14][16][21] through the research methods of field observations and case studies. In addition, this strength is also mentioned in the referenced books.

The strength ‘flexibility in test analysis’ is highlighted in the reviewed literature. Authors in the cited references \[7\][8][12][15][16][21] performed case studies and experiments in order to propose this as strength of ET. Furthermore, the books, which have been referenced for this study, also highlight about the same strength of ET that is based on the authors’ personal experiences.

The strengths of TCBT as presented in section 3.1.7.2 are classified together according to the similarities amongst them. The Table 6 below shows the classifications across distinct and relevant strengths of TCBT.

<table>
<thead>
<tr>
<th>Attributes</th>
<th>The Strengths of TCBT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functionality coverage</td>
<td>1. Complex relationships of a function in software are required to be tested</td>
</tr>
<tr>
<td></td>
<td>2. Test cases provide with most of the test conditions</td>
</tr>
<tr>
<td></td>
<td>3. Test adequacy</td>
</tr>
</tbody>
</table>
| Availability of oracles | 1. TCBT provides detailed level of information to the tester.  
2. Ease of repeatability, to repeat the same tests in TCBT.  
3. TCBT not resource dependent in execution.  
4. Explicit oracles for validation. |
|-------------------------|--------------------------------------------------------------------------------------------------|
| Prediction of quality    | 1. Better analysis of the system specification from diverse angles.  
2. Better Planning and risk estimation  
3. TCBT improves testing quality  
4. Quality of the test cases can be validated  
5. Prediction of reliability. |
| Used for legal requirements | 1. Suitable to acceptance testing  
2. Suitable for release testing  
3. Fits well where legal and regulatory requirements are needed to be met. |

Table 6: Classification of TCBT strengths based on similarities

According to the Table 6 in the figure below total of 9 references were cited for the strengths of TCBT. 40% of the references [3][11][23][24] highlighted that TCBT has better test coverage, 30 % addresses [3][23][29] oracle as a strengths of TCBT. 40% cited references [28][29][30][31] highlighted TCBT is better in predicting overall quality of testing. 30% of the references [9][11][23] posed TCBT as an effective way of testing and fulfilling legal requirements.
Functionality coverage is one of the main strengths of TCBT, and researchers have highlighted this by performing experiments. This is also mentioned in IEEE standard for software test [24] documentation.

Another perceived strength of TCBT is the availability of test oracles which is highlighted by many researchers but as a secondary output of their research i.e. through experimentations and is also mentioned in testing books referred for this study.

The test case ability to predict the quality is also highlighted by many researchers by conducting experiment, case study and survey. This strength of TCBT has also appeared in some of the consultant reports.

Fulfillment of legal requirements is also one of the key strengths of TCBT which has been highlighted in literature. The research methodology adopted by the researchers to present this as strength of TCBT was by conducting case study. Referenced books also highlight it as strength.

The weaknesses of ET as presented in section 3.1.7.3 are classified together according to the similarities amongst them. The Table 7 below shows the classifications across distinct and relevant weaknesses of ET.

<table>
<thead>
<tr>
<th>Attributes</th>
<th>The Weaknesses of ET</th>
</tr>
</thead>
</table>
| Functionality coverage issues | 1. Plan and manage test coverage adequacy  
2. Quality of testing is not known |
| Traceability issues | 1. Difficult to prioritize and select the appropriate tests  
2. Evaluation of tests  
3. Monitoring and keeping track of the progress |
Table 7: Classification of ET weaknesses based on similarities

| Repetition of defects | 1. Repeatability of tests  
2. Re-performing of test steps for effective verification  
3. Investigating and isolating the actual cause of the problem may take longer time in ET  
| Lack of risk management | 1. ET does not assure oracle mistakes  
2. Less accountable and auditable  
3. Highly situational approach |

According to the Table 7 in Figure 4, the references are cited for the weaknesses of ET. 75% of the references [6][7][8] highlighted that ET has test coverage issues. In addition, 50% of the references [6][8] highlighted traceability in testing as an issue. 50% of cited references [6][8] highlighted repetition of defects along with lack of risk management [7][15] is an issue in ET.

Figure 4: Perceived weaknesses of ET from literature

Traceability is an issue of ET that has been highlighted in the literature through case study and field observation. The repetition of defects and lack of risk management are also perceived weaknesses of ET which have been highlighted in one of the case study and consultant report.

The weaknesses of TCBT as presented in section 3.1.7.4 are classified together according to the similarities amongst them. The Table 8 below shows the classifications across distinct and relevant weaknesses of TCBT.
<table>
<thead>
<tr>
<th>Attributes</th>
<th>The Weaknesses of TCBT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problems associated with test cases</td>
<td>1. Quality of the test cases is not known until their execution.</td>
</tr>
<tr>
<td></td>
<td>2. Test cases are not sufficient for entire system life cycle.</td>
</tr>
<tr>
<td></td>
<td>3. Prioritizing test cases is considered as difficult</td>
</tr>
<tr>
<td></td>
<td>4. Redesigning of test cases under time constraints.</td>
</tr>
<tr>
<td></td>
<td>5. Test cases are human prone and require necessary skills and expertise.</td>
</tr>
<tr>
<td>Time consuming</td>
<td>1. Reusability and maintenance quite expensive</td>
</tr>
<tr>
<td></td>
<td>2. More time in documenting tests</td>
</tr>
<tr>
<td></td>
<td>3. TCBT often overruns the assigned budget and time</td>
</tr>
<tr>
<td></td>
<td>4. TCBT is exhaustive and protracted</td>
</tr>
<tr>
<td>Lack of defect defection</td>
<td>1. TCBT is not suitable for regression testing as test cases</td>
</tr>
<tr>
<td></td>
<td>2. Defect detection efficiency not known</td>
</tr>
<tr>
<td></td>
<td>3. Durability of test cases is not known</td>
</tr>
</tbody>
</table>

Table 8: Classification of TCBT weaknesses based on similarities

According to the Table 8 in the figure below 10 references are cited for the weaknesses of TCBT, 70% of the references [9][10][11][20][29][31][35] highlighted that in TCBT mainly problems reside in test cases itself. In addition, 20% of the references [32][33] highlighted that this approach is very much time consuming and 30% of references[11][30][33] identified that this approach lacks in detecting defects.
The major weakness in TCBT is that the problems itself lies in the test cases. This is highlighted by experiments, surveys, case studies and in guide of software engineering body of knowledge. Time consuming nature of TCBT is highlighted by researchers by conducting the interviews and experiments [32][33]. The lack of defect detection ability of TCBT is mentioned in literature through experiment and survey [11][30][33].

3.2 Industrial Interviews

Interviews are particularly used for getting in-depth information and thinking about a participant’s experiences and opinions [36]. In addition, interview can be considered as an effective way of extracting and eliciting relevant research related information by interviewing a domain expert [37]. Kahn et al. describes interview as a formulized and meaningful discussion in which at least two persons are involved i.e. interviewer and interviewee [38]. Primarily, two ways of interviewing can be used i.e. face to face and telephonic/online interviews. For this study in total five semi structured interviews were conducted. The main reason of conducting semi structured interviews is, it provides with an opportunity of two way communication. Thus it can better assist in understanding the responses of interviewees, as open ended questions are asked to get maximum information on the research topic [37]. Another positive aspect of such interviews is that no predefined set of questions are strictly used as in the case of structured interviews. Hence semi structured interviews were utilized for data collection.

3.2.1 Purpose of interviews

The purpose of conducting interviews for this study was to acquire the industrial viewpoints related to the both test approaches i.e. ET and TCBT which also provides answer to the relevant research questions i.e. RQ1 and RQ2. The questions of interviews were highly focused on getting as much insight on the topic of interest as possible. The benefit of conducting interview is that it provides with personal experiences of an individual that cannot be acquired through quantitative measures. By interviewing people it provides insight into their world; their opinions, thoughts and feelings [39].
3.2.2 Selection of interview subjects

Appropriate selection of interview subjects is very important in order to get precise and useful information. People who were directly involved in software testing, specifically in ET and TCBT were categorized as the right candidates for interview. Software testers, test managers, practitioners and consultants were considered in this regard to make the research study more authentic and reliable. The subjects selected for the interview were all of same classification i.e. minimum of 10 years of software testing experience. In addition, subjects also hold a senior position in reputable organizations. An extensive background check was performed on the short listed interview subjects in order to meet the selection criteria. The experience adhered by such professionals was of great essence as they are also involved in interacting with stakeholders. By conducting interview of such people, it gave us broader aspect of the problem domain from multiple perspectives.

3.2.3 Study instruments

Four study instruments were designed to get the relevant information from the interviewees related to the research topic (see Section 3.2.4 for more details). These four study instruments primarily provide partial answers to the two research questions i.e. RQ1 and RQ2 (Appendix Section 8.1). Study instruments were mainly based on the weaknesses and strengths of ET and TCBT as identified in the literature. In order to add more relevant and follow up questions extensive discussion was carried out before the finalization of study instruments. These follow up questions were very useful in getting the emphasis on the previously answered question. The questions of the study instruments were mostly qualitative. To assure the quality of study instruments, all were cross checked by the senior researchers and industry practitioners having knowledge of ET and TCBT.

3.2.4 Interview structure

Following points [39] were fulfilled to make the interview more structured.

3.2.4.1 Thematizing

The interview questions were investigated for two purposes. The first purpose was to formulate the questionnaire based on the identified strengths and weaknesses as highlighted in the literature. The second purpose was to formulate the questionnaire based on some scenario in order to figure out weaknesses and strengths of ET and TCBT which were not mentioned in the literature, but significant in practice. This was an attempt to figure out the industry practitioners view point, and to statically validate weaknesses and strengths as indentified in the literature.

3.2.4.2 Designing of study instrument 1

Study instrument 1 was designed to get the industrial aspect on the weaknesses related to ET. The questions in the study instruments contained most of the weaknesses of ET as highlighted in the literature, in order to validate and get the opinion of industry practitioners on the perceived weaknesses. In addition, other questions were also derived to get independent views of the interviewees on weaknesses of ET, which they might have come across in practice. The questions were primarily qualitative in nature; the detailed questionnaire can be viewed in the appendix section 8.1.1.
3.2.4.3 Designing of study instrument 2

Study instrument 2 was designed to get the industrial aspect on the strengths associated with the usage of ET. The questions in the study instruments contained the perceived strengths of ET as highlighted in the literature in order to validate and get the opinion of industry practitioners on the perceived strengths. In addition, other questions were also derived to get independent views of the interviewees on the strengths that ET brings, which they might have come across in practice. The questions were primarily qualitative in nature; the detailed questionnaire can be viewed in the appendix section 8.1.2.

3.2.4.4 Designing of study instrument 3

Study instrument 3 was designed to get the industrial aspect on the weaknesses related to TCBT. The questions in the study instruments contained most of the weaknesses associated with the use of TCBT as highlighted in the literature, in order to validate and get the opinion of industry practitioners on it. In addition, other questions were also derived to get independent views of the interviewees on weaknesses of TCBT, which they might have come across in practice. The questions were primarily qualitative in nature; the detailed questionnaire can be viewed in the appendix section 8.1.3.

3.2.4.5 Designing of study instrument 4

Study instrument 4 was designed to get the industrial aspect on the strengths associated with the usage of TCBT. The questions in the study instruments contained the perceived strengths of TCBT as highlighted in the literature in order to validate and get the opinion of industry practitioner on the perceived strengths. In addition, other questions were also derived to get independent views of the interviewees on the strengths of TCBT, which they might have come across or observed in practice. The questions were primarily qualitative in nature; the detailed questionnaire can be viewed in the appendix section 8.1.4.

3.2.4.6 Interviewing

The study instruments constructed are the basis, which provides the guidelines and interview questions to be asked in the actual execution. Four interviews were conducted face to face and due to vast geographical distribution one interview was conducted online on Skype. All the interviews were scheduled by the mutual understanding and in addition interviewees were also presented with the research topic before actual commencement of interviews. Approximate duration of each interview was 60 to 90 minutes. During the interviews, notes were taken; a digital recorder was used in order to make sure that no point is missed.

3.2.4.7 Transcribing

During the interview data was collected by means of manually writing down the important points and also by recording with the consent of interviewee. In this phase the results gathered from the interviews were transcribed, so that the irrelevant materials are omitted i.e. the key points of the interview are separated from the general discussion. This was done by writing down all of the minor and major relevant details of interviews in a MS document file. All the transcribed interviews can be viewed in the Appendix Section 8.2.
3.2.4.8 Analyzing and Verifying

The transcribed output of interview was analyzed by applying Qualitative Data Analysis (QDA) methods [40] and this analyzed output was kept for verification. There are basically three phases of QDA analysis i.e. noticing, collecting, and thinking. An interaction of these indentified phases is shown in the figure below [40].

![Figure 6: Qualitative Data Analysis (QDA)](image)

The data gathered from the qualitative interviews was analyzed by adopting the above QDA model. This is a non-linear analysis model and the following characteristics are associated with it [40]:

1. The nature of QDA model is spiral and steps can be repeated.
2. QDA is recursive in nature as there may be a need to collect more data on same things.
3. Each step in QDA is a complete process.

There are three phases of QDA analysis. In the first phase, we noticed and coded all the relevant information as highlighted by the interviewees. We took notes and recorded the interviews for data collection. After that, we analyzed the recorded interviews in order to retrieve the relevant information, which were then labeled as strengths and weaknesses of each testing approach. Later, these labeled strengths and weaknesses were further sorted and classified. In the last phase, analysis was carried out by critically thinking on the classified strengths and weaknesses keeping the goal of the interview in context.

3.2.5 The analysis and results of the collected data by interviews

In the following sub-sections the strengths and weaknesses of both approaches are presented.

3.2.5.1 Strengths of ET

Figure 7 below shows the summary of the perceived strengths of ET, which have been identified by conducting five interviews from highly experienced and relevant industry professionals in this regard.
According to four of the interviewees, defect detection is higher in ET as compared to other test approaches as more time is spent on actual testing rather than any test designs, so that bugs in a system can be investigated and isolated in a much better way while performing ET. In addition testers utilize their test skills on the fly which tends to increase defect detection in particular.

ET is termed as a cost effective test approach by four of the interviewees because it provides rapid feedback and quick learning of the product and in addition it also has low reliance on comprehensive documentation. Furthermore, they highlighted that as most of the time when time is too short for testing, ET proves to be cost efficient in terms of tight schedule.

Another strength of ET as proposed by three of the interviewees is flexibility in test analysis of the product because in ET, a tester can freely explore different areas of the product and they were also of the opinion that it is a process of simultaneous learning and testing which provides better and quick test analysis.

Three interviewees were of the opinion that ET can achieve better regression testing as most of bugs are identified by not following the pre defined steps of test cases. Hence while performing regression testing ET should be considered at least as a complementary cross checking approach. As it may uncover many other bugs as testing around a fix, usually finds other problems.

Two of the interviewees were of opinion that customers are more satisfied with the use of ET as more bugs are being identified and hence bug free software are being delivered thus indirectly it is also providing satisfaction to customer.

All five interviewees highlighted a key strength of ET that is better utilization of tester’s skills as they become more responsible and creative towards testing while they are performing ET. A tester in ET consumes all his time in actual testing which increases his test participation and also polishes his testing skills. ET also facilitates testers to perform testing freely and invent fully. ET takes benefits from both domain skills of a tester and skills related to testing itself. Furthermore, ET is a habit practice for improving things, a mind set and skill set for free explorations.

According to the opinion of two interviewees another benefit associated with ET is that it has more focused documentation. It usually requires documentation after the execution like logs, test notes and videos. It does not produce the useless documentation like some sort or
predicting documents. On the other hand ET emphases on uncertainty level heuristics. ET is sort of conclusion statement. Quantity of the documentation does not help but in fact the quality.

3.2.5.2 Strengths of TCBT

Figure 8 below shows the summary of the perceived benefits of TCBT, which have been identified by conducting five interviews from highly experienced and relevant industry professionals in this regard.

![Perceived strengths of TCBT from interviews](image)

Three of the interviews highlighted test coverage as a strength of using TCBT due to the fact that documenting the test cases can provide better functionality coverage in correspondence with the requirement specification. It can also be used to secure the functionalities, which later can be presented to the customer in order to get his satisfaction. Interviewees also highlighted that achieving 100 percent test coverage is not possible.

Three of the interviews highlighted that where fulfillment of legal and regulatory requirements is mandatory it is compulsory to use TCBT in order to have documented prove of testing which may be required to be presented. Furthermore, acceptance testing is also better achieved by utilizing this test approach.

Four interviewees were of the opinion that TCBT provides better test guidance to the testers on the issues like oracle and also can make up tester’s mind for further creative testing. Test cases can only be beneficial in terms of formulizing or guiding testing tasks if they are clear and designed properly which may also increase the product quality.

Another benefit of using TCBT was pointed out by an interviewee is that bugs can be found before testing actually starts usually in requirements and functional descriptions.

3.2.5.3 Weaknesses of ET

Figure 9 below shows the summary of the perceived weaknesses of ET, which have been identified by conducting five interviews from highly experienced and relevant industry professionals in this regard.
According to three interviewees, availability of an oracle is an issue and it is correlated with the type of the application being tested. If the application is too complex, oracle becomes a serious issue. Furthermore, if a tester lacks in domain knowledge and testing skills, then it becomes quite difficult to test the application without oracles. Oracle issues in ET are often observed in the projects where time is running out and functional specifications have not been updated. Then it can become confusing to an extent on how certain functionality should be implemented and its working will be according to requirement specification.

Three of the interviewees highlighted that test coverage is a weakness of ET. The only way to deduce coverage in ET is by some indirect measures such as the total time of ET and number of bugs discovered and still chances of missing any functionality that cannot be ruled out. In addition, coverage issues also lead to other problems such as tracking and overview of testing.

One of the interviewee stated that conducting only ET on complex application alone is not suitable and it should be combined with other test approaches in order to secure critical functionality of complex and real time application.

All the interviewees highlighted that if a tester in ET is not using appropriate test skills and knowledge it can create many problems. ET is highly correlated with the tester’s abilities and hence more energy is required from testers. It was also mentioned by the interviewees that ET is more dependent on the behavior of a tester towards problem solving.

All five interviewees were of the opinion that managers and organizations are reluctant in the implementation of ET because they think they will lose control over testing. Managers and test leads in most of the organizations always require extensive documentation throughout the testing process which is not the case in ET. Traditional managers prefer to use classical test approaches.

### 3.2.5.4 Weaknesses of TCBT

Figure 10 below shows the summary of the perceived weaknesses of TCBT which have been identified by conducting five interviews from highly experienced and relevant industry professionals in this regard.
Two of the interviewees highlighted that finding defects by using TCBT may be difficult for a tester because it may be impossible for him to follow each and every step of the test case as many of the defects are identified by deviating from those explicit steps. Another problem of TCBT which corresponds to the defect detection is mentioned by all five interviewees; i.e. if a test design is not proper it can be very problematic. Test cases are considered as just one output, and problems can be inherited in them from documents such as software requirement specification and test plans which can affect defect detection. Having a proper and a well planed test design is very important to get clear test instruction in TCBT.

Three of the interviewees were of the opinion that executing, managing, and updating test cases for an application are very exhaustive and painstaking task. Furthermore, test cases in a system lifecycle once written are never sufficient for entire lifecycle and are needed to be updated continuously according to the changes in the requirement.

Three interviewees termed TCBT as a non-skilled activity because a tester may not able to make decisions based on the test cases if they are not written and designed properly. Pre defined test cases also limits the thinking and viewpoints of the tester and he may not be able to think outside the box. In addition, an interviewee highlighted that if a tester was only to follow a pre-scripted set of test steps then it does not require any testing skills.

### 3.2.6 Discussion and Summary

The main contribution of this chapter was to present readers with the strengths and weaknesses of ET and TCBT. Systematic literature review and industrial interviews were conducted in order to get the relevant results. Interestingly, the weaknesses and strengths of both test approaches as observed in the literature were also identified to be experienced in industry to some extent. By conducting interviews of experienced and relevant industry professionals in this regard, we were able to grasp some of the aspects related to strengths and weaknesses ET and TCBT which were not observed in the literature. For example, better customer satisfaction and focused documentation are highlighted as strengths of ET by interviewees. For the strength of TCBT, the interviewees mentioned that usage of TCBT can identify the bugs before actual testing in requirement specification.

On the other hand one weakness of ET, which was repeatedly observed while conducting the interviews, was that managers and organizations are reluctant in the implementation of ET because they think they will lose control over testing. Another weakness of ET is that results...
are often difficult to interpret because results are generated based on testers own experience and intuition. Interviewees mentioned that TCBT is a non skilled activity because a tester may not be able to make decisions based on the test cases if they are not written and designed properly.

3.2.7 Validity threats

The validity threats associated with the study are described as below:

3.2.7.1 Internal validity

Internal validity inspects whether or not accurate deductions are derived from the gathered data [18]. The threats related to internal validity can be reduced by making the referral to multiple perspectives on the topic of interest.

For systematic literature review, the objective nature of classification was done according to the mutual understanding of the researchers, so its subjective nature cannot be ruled out which can be termed as threat to this study. The threat which was needed to be coped in systematic literature review is associated with publication biasness. An attempt was made to strictly follow the defined quality assessment criteria and to use well known databases. This was done in order to assess both positive and negatives findings of the primary studies, so that they can be presented in the report. By adopting this way, authors have tried to minimize the internal validity threat.

For interviews, subjective questions can lead to a poorly designed interview which will greatly affect the research artifacts. The questions of interviews were strictly designed on the basis of the problems and issues, which were relevant to the literature but needs industrial emphasis. The threat of missing any important questions related to strengths and weaknesses of TCBT and ET were minimized by mutual and extensive discussions.

3.2.7.2 External validity

External validity generalizes the outcomes in different settings [18]. External validity threats can be reduced to an extent by generalizing the study in multiple situations on a small scale.

Since the systematic literature review was conducted between the years of 2000 and 2009. There is a threat of missing any relevant strengths and weaknesses of ET which may have been published before the specified dates. In order to overcome this threat several relevant books were also consulted.

For interview, there was a threat that they may show lack of interest, this threat was coped by selecting those professionals who were interested and motivated towards the research topic. We also enlightened all the interviews before hand by sending them the research idea.

3.2.7.3 Construct validity

Construct validity evaluates the utilization of accurate definition and measures associated with the variables [18].

The constructs i.e. the search strings utilized for the systemic literature review was well defined and was reviewed by the academic supervisor. Thus it rules out any threat related to the constructs.
For interviews, the data was gathered in the form of tacit knowledge. A risk of misinterpretation is associated while making it explicit, as tacit data can have multiple interpretations. This risk was mitigated by relating the relevant data present in literature with the explicit interpretation of gathered data.

3.2.7.4 Conclusion validity

Conclusion validity also known as reliability which is related to the research study results and it makes sure that the gathered results lead to correct and reliable conclusion. It addresses about the issues which may affect the results [41].

For systematic literature review, pilot searches were performed by using the defined search strings in different databases and search engines in order to show the reliability of the systematic literature review.

ET is a new dimension in software testing; there were not many literatures available. Therefore, the results of the systematic literature review were strictly based on the available research literature.

Since both interviewers and interviewees were not native speaker of English, hence issues related to understanding on both fronts cannot be ruled out. In order to minimize the risk clear and unambiguous questions were written and asked in simple English and a recording device was also used in this regard.
4 A PROCESS FOR MT

In this chapter, we propose a process for MT considering the weaknesses and strengths of both test approaches, which we presented in Chapter 3. First, we provide a background on the test processes currently being used in the industry for ET and TCBT. Then, we describe the details of the process we propose.

Here, we use ISO/IEC 29119 software testing standard [25] for the discussions on TCBT since it is a new process which replaces a number of existing IEEE standards for software testing i.e. IEEE 829 Test Documentation and IEEE 1008 Unit Testing. Moreover, SBMT process is used for ET and the reason of selecting this process is that it is the only structured process, which is available for ET [42].

4.1 Test process for TCBT (ISO/IEC 29119)

4.1.1 Test planning

The purpose of defining a test plan is to specify the scope, testing approaches, allocation of resources and scheduling of the test activities [24]. Furthermore test plan also provides the details of the product being tested, associated risks and mitigation, features to be tested along with the resources responsible for the completion of each testing task.

4.1.2 Test monitoring and control

The purpose of test monitoring and control is to ensure that the testing tasks are carried out in line with the defined test plan and other policies, which may influence testing process [24]. Test monitoring and control can be applied at any level i.e. project or individual test phases.

4.1.3 Test completion Criteria

Test completion process ensures the availability of the useful test assets to the relevant stakeholders for later use in conjunction with the documented results of testing [24]. Test assets may include test plans, test environment infrastructure and test specifications.

4.1.4 Test design and implementation

The purpose of this phase is to develop test procedures that will be executed in the test execution phase [24]. Test cases, test conditions and execution requirements are derived as a part of this process. Each test case may contain the information related to the features to be tested; test approaches and a pass fail criteria.

4.1.5 Test execution

In this phase the test cases that are created in the test design and implementation phase are executed in order to record the results and outcomes against the executed test cases [24]. The outcomes of a successful test execution process are as follows:
4.1.6 Test incident reporting.

The purpose of this phase is to report any issues and anomalies discovered in the test execution phase, which need further actions [24]. In case of new test being executed a new incident report will be created and further managed. However in case of re-test or regression testing previous incident reports will be updated and utilized accordingly.

4.2 Session based test management process for ET

Bach [42] introduced SBTM process for structuring exploratory testing. The primary work unit of this process is a “Session” initiated by a charter; it’s neither a test case nor a bug report. A session is an uninterrupted block of pure test time allocated to the tester. At the end of a session a sheet is generated i.e. session sheet which can be reviewed by test managers and test leads. Each session is debriefed at the end. The process flow of the SBTM is shown in Figure 12 below.
Test sessions are divided into following three tasks collectively known as “TBS” metrics

- Test design and execution (scanning the product and looking for problems)
- Bug investigation and reporting
- Session setup (prerequisites addressed by the tester which makes the first two tasks possible)

4.3 The Proposed Mix Testing Process

ET and TCBT processes are currently being used distinctly for each test approach. There is a need of having a MT process, which is capable to accommodate and manage both of these test approaches in order to benefit the strengths of each and overcome some of the weaknesses. Therefore, in this section, we propose a new MT process that is an integration of the ET and SBTM test processes.

We identified and discussed the associated strengths and weakness of ET and TCBT in Section 3.1.7 and Section 3.2.5. These weaknesses and strengths were identified by conducting detailed systematic literature review and by interviewing industry professionals. After conducting the detailed analysis on the gathered data, an attempt is made in below section to map the weaknesses and strengths of both the approaches in such a manner that the strengths of one approach can address the weaknesses of other.

Our research methodology had two steps. We first used a mapping process, which is a design approach that directly captures the real world problems in order to design a solution to resolve the problems [43]. It is also used to visually depict the mapping of problems to benefits. In mapping, facts are gathered and organized for display, so that they can be questioned and improved. Any unnecessary details are masked in mapping process.
In the second step, we evaluated the outcomes of the previous step by interviewing industry professionals and utilizing their suggestions when defining the final MT process.

The details of these two steps are presented in following sections.

4.3.1 Mapping

Table 9 below shows the mapping of the weaknesses associated with TCBT to the strengths of the ET as highlighted in Section 3.1.7 and 3.2.5. For most of the identified weaknesses of TCBT solutions can be attained by utilizing the relevant strengths of the ET as demonstrated in the Table 9 below:

<table>
<thead>
<tr>
<th>Weaknesses of TCBT set as Problems</th>
<th>Solutions Based on the Strengths of ET</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0 Problem</td>
<td>Solution 1.0</td>
</tr>
<tr>
<td>1.1 TCBT is not suitable for regression testing</td>
<td>1.1 Better defect detection and isolation</td>
</tr>
<tr>
<td>1.2 Defect detection efficiency is unknown</td>
<td>1.1.1 Less bogus defects than TCBT</td>
</tr>
<tr>
<td>1.3 Durability of test cases is unknown</td>
<td>1.1.2 Identification of critical bugs</td>
</tr>
<tr>
<td>2.0 Problem</td>
<td>Solution 2.0</td>
</tr>
<tr>
<td>2.1 Reusability and maintenance is expensive</td>
<td>2.1 Cost effective</td>
</tr>
<tr>
<td>2.2 More time in documenting tests</td>
<td>2.1.1 Low reliance on comprehensive documentation</td>
</tr>
<tr>
<td>2.3 TCBT often overruns the assigned budget and time</td>
<td>2.1.2 Rapid feedback</td>
</tr>
<tr>
<td>3.0 Problem</td>
<td>Solution 3.0</td>
</tr>
<tr>
<td>3.1 Quality of the test cases is not known until their execution.</td>
<td>3.1 Better product analysis and testing.</td>
</tr>
<tr>
<td>3.2 Revision of test cases on the basis of changed objectives.</td>
<td>3.1.1 Tester freely explore</td>
</tr>
<tr>
<td>3.3 Prioritizing test cases.</td>
<td>3.1.2 More time in actual testing of product</td>
</tr>
<tr>
<td>3.4 Not based on skills of tester.</td>
<td>3.1.3 Simultaneous learning and testing.</td>
</tr>
<tr>
<td></td>
<td>3.1.4 Interpreting vague test instructions.</td>
</tr>
<tr>
<td></td>
<td>3.1.5 Skilled based testing.</td>
</tr>
</tbody>
</table>

Table 9: Mapping of TCBT weaknesses to the strengths of ET

Table 10 below shows the mapping of the weaknesses associated with ET to the strengths of the TCBT as highlighted in Section 3.1.7 and 3.2.5. For most of the identified weaknesses of ET, solutions can be attained by utilizing the relevant strengths of the TCBT as demonstrated in the Table 10 below:
<table>
<thead>
<tr>
<th>Weaknesses of ET set as Problems</th>
<th>Solutions Based on the Strengths of TCBT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1.0 Problem</strong></td>
<td></td>
</tr>
<tr>
<td>1.1 Monitoring and keeping track of the progress is difficult</td>
<td><strong>Solution 1.0</strong></td>
</tr>
<tr>
<td><strong>2.0 Problem</strong></td>
<td></td>
</tr>
<tr>
<td>2.1 Lack of effective risk management</td>
<td>1.1 Monitoring and control</td>
</tr>
<tr>
<td>2.2 Less accountable and auditable</td>
<td>1.1.1 Better tracking of test progress</td>
</tr>
<tr>
<td>2.3 Difficult to prioritize</td>
<td>1.1.2 Better monitoring of testing tasks</td>
</tr>
<tr>
<td>2.4 Company acceptance of ET</td>
<td><strong>Solution 2.0</strong></td>
</tr>
<tr>
<td><strong>3.0 Problem</strong></td>
<td></td>
</tr>
<tr>
<td>3.1 Plan and manage test coverage adequacy is difficult</td>
<td>2.1 Test planning</td>
</tr>
<tr>
<td>3.2 Quality of testing is unknown</td>
<td>2.1.1 Better Planning and risk estimation</td>
</tr>
<tr>
<td>3.3 Oracle mistakes</td>
<td>2.1.2 Risk mitigation</td>
</tr>
<tr>
<td><strong>4.0 Problem</strong></td>
<td></td>
</tr>
<tr>
<td>4.1 Repeatability is difficult in ET</td>
<td>2.1.3 Better analysis of the product</td>
</tr>
<tr>
<td>4.2 Traceability issues</td>
<td><strong>Solution 3.0</strong></td>
</tr>
<tr>
<td><strong>5.0 Problem</strong></td>
<td></td>
</tr>
<tr>
<td>5.1 ET is not suitable for acceptance and release testing.</td>
<td>3.1 Test Design</td>
</tr>
<tr>
<td>5.2 Not suitable for complex applications</td>
<td>3.1.1 Test adequacy better coverage.</td>
</tr>
<tr>
<td></td>
<td>3.1.2 TCBT improves testing quality.</td>
</tr>
<tr>
<td></td>
<td>3.1.3 Quality of the test cases can be validated.</td>
</tr>
<tr>
<td></td>
<td>3.1.4 TCBT provides solution to oracle issues.</td>
</tr>
<tr>
<td></td>
<td>3.1.5 Test cases provide with most of the test conditions.</td>
</tr>
<tr>
<td></td>
<td>3.1.6 Test guidance</td>
</tr>
<tr>
<td></td>
<td>3.1.7 Bugs found before actual testing</td>
</tr>
<tr>
<td></td>
<td><strong>Solution 4.0</strong></td>
</tr>
<tr>
<td>4.1 Test case execution</td>
<td>4.1 Test case execution</td>
</tr>
<tr>
<td></td>
<td>4.1.1 Ease of repeating the same tests.</td>
</tr>
<tr>
<td></td>
<td>4.1.2 TCBT not resource dependent in execution</td>
</tr>
<tr>
<td><strong>5.0 Problem</strong></td>
<td><strong>Solution 5.0</strong></td>
</tr>
<tr>
<td>5.1 Test completion</td>
<td>5.1 Test completion</td>
</tr>
<tr>
<td>5.2 Suitable to acceptance and release testing.</td>
<td>5.1.1 Suitable to acceptance and release testing.</td>
</tr>
<tr>
<td>5.3 TCBT fits well for legal and regulatory requirements.</td>
<td>5.1.2 TCBT fits well for legal and regulatory requirements.</td>
</tr>
</tbody>
</table>

Table 10: Mapping of ET weaknesses to the strengths of TCBT
4.3.2 Interviews

4.3.2.1 Purpose of interviews

The purpose of using interviews for this study was to evaluate the above mapping and to get the opinions and suggestions of industry professionals on preliminary defined process for MT. These suggestions acted as the inputs to further mature the MT process. The results provide the answers to RQ3. The questions of interviews were highly focused on getting as much insight on the topic of interest as possible. The benefit of conducting interview is that it provides with personal experiences of an individual that cannot be acquired through quantitative measures. By interviewing people it provides insight into their world; their opinions, thoughts and feelings [39].

4.3.2.1.1 Selection of interview subjects

Appropriate selection of interview subjects is very important in order to get precise and useful information. Professionals directly involved in software process definition, strategy development and approach utilization specifically for testing, were categorized as the right candidates for interview. Software test managers, leads and consultants were considered in this regard to make the research study more authentic and reliable. The subjects selected for the interview were of same classification i.e. managerial level. The experience adhered by such professionals is of great essence as they are also involved in interacting with stakeholders. By conducting interview of such people, it gave us broader aspect of the problem domain from multiple perspectives.

4.3.2.1.2 Study instrument

One study instrument was designed to get the relevant information from the interviewees related to the research topic. This study instrument (appendix section 8.1.5), primarily provides the answer to research question i.e. RQ3. Study instrument was mainly based on getting the suggestions and any improvement opportunities on the proposed MT process. In order to add more relevant and follow up questions extensive discussion was carried out before the finalization of study instrument, these follow up questions were very useful in getting the emphasis on the previously answered question. The questions of the study instrument were mostly qualitative. To assure the quality of study instruments all were cross checked by the internal and external supervisor.

4.3.2.1.3 Interview structure

Conducting interview was an effective and productive way of identifying the optimal solution to the process definition of implementing MT. Many expert based suggestions were acquired by the industry professionals who lead to a better and more practical process definition of MT. The following points [39] were fulfilled to make the interview more structured.

4.3.2.1.4 Thematizing:

The interview questions investigate the factors which affect the proposed process. This was done by asking questions to the industry professionals, keeping the preliminary MT process
in context. Interviewees were presented with some scenarios in order to validate or grasp improvement opportunities in the MT process. This was an attempt to figure out the industry practitioner’s viewpoint, and to statically validate the mapping for implementing the MT.

4.3.2.1.5 Designing of the study instrument:

Study instrument was designed to get an industrial aspect on the MT process. The questions in the study instrument were based to validate mapping and to get suggestions for the definition of MT process. In addition, other questions were also derived to get independent views of the interviewees on the MT process definition. The questions were primarily qualitative in nature; the detailed questionnaire can be viewed in the appendix section 8.1.5.

4.3.2.1.6 Interviewing:

Four interviews were conducted face to face. All the interviews were scheduled by the mutual understanding and in addition interviewees were also presented with the research topic before actual commencement of interviews. Approximate duration of each interview was from 30 to 45 minutes. During the interviews, notes were taken; a digital recorder was used in order to make sure no point is missed.

4.3.2.1.7 Transcribing:

During the interview data was collected by the means of manually writing down the important points and also by recording, with the consent of interviewee. In this phase the results gathered from the interviews were transcribed, so that the irrelevant materials are omitted i.e. the key points of the interview are separated from the general discussion. This was done by writing down all the minor and major relevant details of interviews in a MS document file.

4.3.2.1.8 Analyzing and verifying:

The transcribed output of interview was analyzed by applying qualitative data analysis method as explained in section 3.2.4.8.

4.3.3 Results

4.3.3.1 Feedback of interviewee 1

Interviewee 1 validated the mapping process and suggested that strengths and weaknesses of both test approaches were concise and detailed. Her concern was how in reality these strengths of each testing approach will work out on real projects and provide benefits. She added that the weaknesses of ET and TCBT are generic and that in practical environment there can be many ways to deal with such issues by other means. However, she affirmed, providing a solution inferred from strengths of both test approaches and attempting to resolve the weaknesses was quite innovative in a sense. When we presented her with process flow description she had some confusions related to the Happy Path Test Cases (HPTC) and she mentioned that it was not clear enough what HPTC actually means. Her opinion was to either provide more details about these HPTC or find any other alternative word which is
more self explanatory. She also had some reservations on de-briefing sessions because she considered that managing the test team may even take more time as de-briefing session has been introduced. Instead test leads should be involved in MT process.

4.3.3.2 Feedback of interviewee 2

Interviewee 2 mentioned that mapping the strengths of both testing approaches to the weaknesses was a good way to compare both testing approaches. He said that mapping was an ideal way of presenting the solution based on theoretical constructs but practically this mapping may not provide with 100% solution. He gave an example that while discussing the quality of testing as strength of TCBT and weakness of ET, what are the odds that derived test cases will be of quality in TCBT. He added that it is a high level presentation of strengths to weaknesses but still all strengths of both test approach may have several weaknesses that may be associated with other indirect measures. He said that HPTC should only be used as complimentary, specifically where graphical user interface testing is required and test cases are hard to codify. He was of the opinion that HPTC may provide with different meanings thus it should be further evaluated.

4.3.3.3 Feedback of interviewee 3

Interviewee 3 highlighted that mapping of strengths to weaknesses was an appropriate way of defining a process based on both test approaches i.e. ET and TCBT. He validated the mapping process and mentioned that it is quite elaborative. When we presented him with the initial process flow description he added that he is not fond of flow boxes connected to each other telling him what to do, and he was of the opinion that let the context decide which box should be used in a specific situation. In which order you use the boxes is not important if the order does not match the current context. He also recommended the introduction of free exploration in order to learn about the application i.e. before, after or during the execution of HPTC. Free exploration will provide an edge to the testers because they will be able to immediately look for any major abnormality in a very short span of time.

4.3.3.4 Feedback of interviewee 4

Interviewee 4 was of the opinion that there should be more flexibility in using any sort of test cases not only HPTC. He also suggested that these HPTCs should be made more generalized and one should not limit to HPTC only, and it should be up to testers or managers to decide upon what they need and require out of testing. He highlighted that performing ET at the beginning of testing lifecycle can provide many benefits and therefore, it should also be incorporated in the MT process. He said that HPTC could have many different meanings, hence more clarification should be provided in this regard. He pointed out that little more should be discussed about the test exit criteria. He also recommended that upon the conclusion of every de briefing session, more test missions should be drafted based on the testers report and intuitions. These newly devised test missions should become the input for further session executions.

4.3.4 Discussion

All four interviews validated the mapping of weaknesses to strengths. Since perceived strengths and weaknesses of ET and TCBT were obtained by conducting interviews and systematic literature review, all the interviewees evaluated the mapping process. However interviewees suggested some changes in the preliminary defined process of MT such as:
• Modification of HPTC
• Introduction of free exploration
• Defining more test missions based on debriefing session.
• Test exit criteria

The results of mapping and the interviews are presented in the defined process of MT in the section below:

4.3.5 Process definition (Based on the result and findings)

MT was tailored on the perceived strengths of ET and TCBT as gathered by studying the relevant literature along with conducting interviews. The suggestions from the interviews as mentioned in the Section 4.3.4 were incorporated while refining the MT process along with the mapping of weaknesses to strengths. The process of MT was also influenced by two state of the art processes: ISO/IEC 29119 for TCBT and Session SBTM for ET. Figure 13 shows the process description of the proposed MT.
In this test plan both the RBTC and ET planning would be done by specifying the scope, testing approaches, allocation of resources and scheduling associated risks, mitigation and test strategies for testing features.

The test monitoring and control ensures that the testing tasks are carried out in line with the defined test plan. This would be like an umbrella activity which run through out the test process.

The test procedures, missions RBTC, test conditions are derived which will assure the basic functionality and requirement coverage.

Session is an uninterrupted block of pure test time allocated to the mission in order to execute a mission.

Session is an uninterrupted block of pure test time allocated to the mission in order to execute a mission.

This will assures that complete testing is carried out.

Free Exploration

When every RBTC passes the execution goes for ET

RBTC are executed until every test case is passed

Session Sheet gives all the information related to testing which can assists test managers and test leads.

Session Execution

Test Mission

Session is an uninterrupted block of pure test time allocated to the mission in order to execute a mission.

Success

No

Test case

Regression

RBTC and Test mission Design

Test Planning

4.3.5.1 Overview of the MT process description

Based on the above process description, in MT initially a test plan is drafted, and this test plan becomes an input to test design phase. In the test design phase, Requirement Based Test Cases (RBTCs) [44] are designed in order to secure the high-level functionality requirements. In addition, test missions are also designed in this phase. MT process also facilitates the usage of free exploration in order to learn about the application i.e. before, after or during the execution of RBTCs in test execution phase. While performing free exploration, if any abnormality or bug in the application is detected it is reported to the development unit and relevant missions are created in the test design phase for later usage or regression testing. Hence, free exploration would assist in better product and test analysis, and also in maturing RBTCs and test missions.
During the RBTCs execution if any bug is detected, it corresponds to major functional requirement fault. Therefore, it is reported and upon its fixing, regression testing is performed by executing the same RBTC in order to verify it. In addition, relevant missions are also designed, which further look into the risk areas associated with that bug, in the later phase of testing. This test cycle keeps on rolling until all the RBTCs are successful and no further bugs are detected.

After the success of all RBTCs, test missions are executed which have already been defined and refined in the test design. Each mission is assigned with a session time based on the attributes of a mission. In the session execution a tester follows the assigned mission and if any bug is discovered it is reported for fixing. At the same time, a session sheet, containing test notes and other relevant information is also filled.

At the end of each session, a briefing session is set up in which tester and other relevant stakeholders take part and may devise more test missions based on the testers report and intuitions etc. After the fixation of the bug, more test missions are created at run time to ensure effective regression testing. Upon the completion of all the test missions, test completion report is drafted and testing process halts.

Below is the detailed explanation of each MT process activities that are recommended in order to better utilize MT process:

### 4.3.5.2 Test planning

Test planning is of iterative nature. A test plan can also be updated during the test execution process in response to the results and outcomes that may affect one or many activities in test plan and thus creating a need to re visit the activities and update them accordingly.

For example, during the course of testing some new risks are identified related to project or product. The existing risks are updated and changed accordingly. If for some reasons, the test strategies are needed to be updated during the test execution, test plan should also be updated.

#### 4.3.5.2.1 Purpose of test planning

The primary purpose of test planning in MT process is to document and communicate all the necessary and required information to all the stakeholders i.e. utilization of test approaches, techniques, identification of resources, risks and other relevant requirements considered mandatory for conducting successful testing.

#### 4.3.5.2.2 Outcomes of test planning

The outcomes of the test planning are as follows:
a) The requirements necessary for conducting testing and the context required for understanding it.

b) Risks associated with project and product are identified and further analyzed in order to classify them on an agreed level of severity.

c) Risk mitigation strategy

d) The required staff and resources needed to carry out the testing tasks are indentified and scheduling of each activity.

e) Analysis of any applicable regulatory and legal requirements or standards that may have an impact on testing.

f) Identification, planning and scheduling of those activities, which are needed to be performed.

g) Test plan

h) Test plan is published and is made accessible to all the stakeholders.

4.3.5.3 Test monitoring and control

The test monitoring and control provides assurance of whether or not the testing being performed is in line with the defined test plan. This controls and monitors every aspect within the test process i.e. tests design, test execution, test mission session, and incident reporting etc.

4.3.5.3.1 Purpose monitoring and control

The purpose is to ensure whether all the activities as specified in test plan are aligned with the actual execution of those activities.

4.3.5.3.2 Outcomes

The outcomes of successful implementation are as follows:

a) Means are setup in order to collect the suitable measures for monitoring the test progress.

b) The actual progress of test related tasks and risks are monitored against the test plan.

c) Progress of tests is reported to the relevant stakeholders.

d) Control actions are defined in order to provide remedy for the observed misalignments and these control actions are channeled to the stakeholders.

e) A decision of continuing or halting the testing is made.

f) Means of tracking and controlling the test related activities.
g) Measurement activities should be emplaced i.e. collection of test metrics, test status reporting etc. furthermore to the collect measures related to the tasks identified above.

4.3.5.4 Test design (Test missions and Requirement based test cases)

This determines the procedure of deriving the tests i.e. RBTC and test missions. Both should be documented in test specification.

4.3.5.4.1 Purpose of test design and implementation

RBTCs and test missions are derived in the test design and implementation process which are then executed in the test execution along with other necessary test execution requirements.

4.3.5.4.2 Outcomes

The outcomes of successful design and implementation are as follows:

   a) The requirements of the test item are understood based on the functional requirements.

   b) Test coverage is derived from of functional requirements.

   c) RBTC are derived in order to secure the high-level functionality requirements.

   d) Test sets are assembled based on the RBTCs prioritization.

   e) Test missions are identified to freely explore every possible scenario.

   f) Test mission sets are assembled based on the test mission prioritization.

   g) The expected result or oracle should also be determined in the RBTCs. The test mission should be derived in order to cover the criticality and importance of the functionality. The risks associated with the functionality and requirements should also be highlighted along with the mission.

   h) Test specification is drafted and approval is taken.

4.3.5.5 Test execution (Free exploration and RBTC execution)

4.3.5.5.1 Purpose

In test execution, the RBTCs from the test design are executed. Meanwhile in test execution a tester has the freedom to freely explore and learn about the test item before, after or during the execution of RBTCs.

4.3.5.5.2 Outcomes
The outcomes of successful execution are as follows:

a) RBTCs are executed.

b) The outcomes generated by the execution of the RBTCs are recorded.

c) The expected and actual results of RBTCs are compared.

d) The RBTC results are determined.

e) Results are judged on the basis of highlighted error and notes are taken of other important aspects.

f) The results generated from free exploration are recorded and notes are taken on other important aspects that need further attention.

g) Identified bugs should be recorded in the bug report or a sheet.

h) Actual results generated by the RBTCs and free exploration should be utilized to analyze the behavior of application.

i) Notes should be taken down on the aspects of the test item, which may require further attention or testing in order to derive further missions.

4.3.5.6 Test mission

This process determines the procedure of deriving the test missions, these missions should be documented in the session sheet or in a mission log file. A test mission can be defined by a test lead or a tester himself on run time in order to meet certain test criteria. Furthermore, debriefing sessions, can also lead in derivation of many test missions.

4.3.5.6.1 Purpose

Test missions that already have been derived in the test design should be executed in the session execution phase along with other necessary test execution requirements. In addition, test missions derived on the fly should be reprioritized on the basis of their criticality; hence a new test mission execution sequence should be deduced every time a new mission is added.

4.3.5.6.2 Outcomes

The outcomes of successful design and implementation are as follows:

a) The requirements of the test item are understood based on the identified test missions.

b) Test mission are identified to freely explore every possible scenario.

c) Test mission sets are assembled based on the mission prioritization.

d) New test missions are derived after each session execution iteration.
e) More test missions are derived by analyzing the output of free exploration

f) Test missions should be derived to cover the criticality and importance of the functionality.

g) The risks associated with the functionality and requirement should also be highlighted along with the test mission.

4.3.5.7 Session execution

Every test mission which is to be executed is assigned with a session time; i.e. already defined in test design phase along with the newly indentified test missions. A session is an uninterrupted block of test time, in which a tester is supposed to test according to the defined test mission. A session time may last 30 to 90 minutes.

4.3.5.7.1 Purpose

The purpose of this phase is to assign the session time to the test mission that is to be executed. Session time is assigned by the test lead or the test manager. And it should be made sure that a session remains an uninterrupted block of test time.

4.3.5.7.2 Outcomes

a) Test missions are executed in an allotted session time.

b) The outcomes generated by sessions are recorded.

c) Session execution results are determined.

d) Both results are judged on the basis of highlighted error and notes are taken of other important aspects.

e) A test session should be an uninterrupted block of allocated test execution time.

f) Test session outcome should be recorded in session sheet/ bug report.

g) Actual results generated by the test session should be utilized to analyze the behavior of application.

h) Notes should be taken down in the session sheet on the aspects of the test item which may require further attention or testing.

i) Test steps performed in the session should be documented in the session sheet.

j) Test session should be completed within the allocated time and all relevant information should be noted down in session sheet.

k) Any test technique utilized in a session should be documented in the session sheet.
4.3.5.8 Test incident reporting

This describes the mechanism of how the test results are reported. This reporting will only be entered when faults, bugs, failures or any other ambiguities related to the functionality are observed as a result of test execution.

4.3.5.8.1 Purpose

The purpose is to report the issues identified in the test execution to the relevant stakeholders in order to conduct further actions on the reported problems. If the test cases and the test sessions are executed for the first time, then new incident or bug reports will be created. If the test execution is carried out the second or the third time, then previous incident reports will be updated and utilized accordingly.

4.3.5.8.2 Outcomes

The outcomes of successful implementation are as follows:

a) Analysis is conducted on the test results and the status of newly and previously identified incidents is determined.

b) The identified incidents are documented in a new or existing bug reports.

c) The incidents new/previous/updated are channeled to the relevant stakeholders.

4.3.5.9 Debriefing

After the completion of session a debriefing session is setup between the tester and a test lead.

4.3.5.9.1 Purpose

The purpose of this debriefing session is to get the input of a tester on the test mission that was assigned to him and to discuss about his observations. A debriefing session should also provide coaching to the tester regarding further test activities that are needed to be performed. If required a debriefing session can lead to the derivation of many test missions.

4.3.5.9.2 Outcomes

a) Analysis is conducted on the session sheets and on the notes written by the tester.

b) Tester’s viewpoint is taken on the session execution cycle.

c) Further test missions are derived in the debriefing sessions.

d) Based on the session sheet evaluation of tester’s performance is determined.

e) If required, coaching is provided to the tester by the senior test leads and test managers.
f) Issues of testers are discussed and resolved etc.

4.3.5.10 Test completion

Test completion criteria are met when an agreement has been reached that the testing being performed and managed is complete.

4.3.5.10.1 Purpose of test completion criteria

The purpose of test completion criteria is to make sure that the useful test assets such as test plans, test cases, session sheets, are made available and all the results are documented, recorded and communicated to the relevant stakeholders.

4.3.5.10.2 Outcomes

The outcomes of successful implementation are as follows:

a) Test assets are archived or are sent to the relevant stakeholders and test environment is cleaned for the next test activity.

b) Relevant and required information should be extracted from any of the artifacts i.e. test plans, test results, status reports, incident/bug reports test completion reports, session sheets, where applicable.

c) Test completion report is drafted and approved.

4.3.6 Validity threats related to MT process

Since the process definition is based on the results of the systematic literature review and interviews which were conducted to get strengths and weaknesses of ET and TCBT, any overlooked concept might have not considered in the process definition. We tried to overcome this validity threat through the design of systematic review and interviews.

The weaknesses of both test approaches were mapped to the strengths. The authors did this mapping. Since this was done through a subjective thinking process, there is a probability that something might have missed in any part of mapping.

MT process was defined by getting the inputs from mapping and interviews. All those interviewees were selected who already had participated in previous interviews, which were conducted to acquire strengths and benefits. Hence the threats related to interviewers were minimized in the process definition.
5 VALIDATION OF THE PROPOSED MT PROCESS

In this chapter, we present the results of the validation for the proposed MT process by conducting static and dynamic validation. Static validation was done based on the feedbacks of the industry professionals whereas dynamic validation was carried out by conducting an experiment in an industrial context.

5.1 Static Validation

In this section, we present the static validation of the proposed MT process by taking the feedbacks of industry professionals. The following steps were followed for carrying out the static validation of MT process.

5.1.1 Selection of subjects

Appropriate selection of subjects for getting the feedback was very important in order to get precise and useful information. A major part of MT process was defined based on the recommendations of the industry professionals. For this reason same subjects, who already had worked with us during the study and possessed better understanding of the MT process were selected in order to get feedback.

5.1.2 Study instrument

One study instrument was designed to get the relevant information from the subjects related to the research topic. This study instrument (appendix section 8.1.5), partially provides the answer to research question i.e. RQ4. Study instrument was mainly based on the open-ended questions in order to get feedback on the MT process. In order to add more relevant and follow up questions extensive discussion was carried out before the finalization of study instrument. These follow up questions were very useful in getting the emphasis on the previously answered questions. The questions of the study instrument were mostly qualitative. To assure the quality of study instruments all were cross-checked by the internal and external supervisor.

5.1.3 Feedback structure

Getting the feedback was an effective and productive way of acquiring the opinions of industry professionals on the defined MT process. A number of expert based feedbacks were acquired by the industry professionals in order to statically validate MT process. The following points [39] were fulfilled to make the process of feedback more optimal.

5.1.3.1 Designing of the study instrument

Study instrument was designed in order to statically validate the MT process. The questions in the study instrument were based to validate whether MT process was sufficient in providing the solution to the weakness of both test approaches. In addition, other questions were also derived to get independent views of the interviewees on the MT process. Study instrument was designed in a manner that it facilitated subjects to write more about their
opinions on the MT process. The questions were primarily qualitative in nature; the detailed questionnaire can be viewed in appendix section 8.1.5.

5.1.3.2 Conducting of Feedback:

Four feedbacks were taken on the MT process. The feedback form was emailed to the subjects along with the process description of MT. All the subjects were requested to send back their feedbacks within the time frame of 5 working days.

5.1.3.3 Analyzing and verifying:

The feedback forms were analyzed by applying qualitative data analysis method as explained in section 3.2.4.8.

5.1.4 Result

The results of the feedbacks on MT process are presented in the section below:

5.1.4.1 Feedback of Interviewee 1

According to Annika Mansson (appendix section 8.3), MT process can prove to be an effective way of utilizing both test approaches i.e. TCBT and ET. She also stressed on the fact that the proportion that will be utilized for each test approach in MT process should also be kept in context because it is directly related to the nature of the program or platform. She added that the MT process will resolve the oracle issues to some extent as documented test cases will provide with concrete results that later can be presented to the customers and project managers. On the other hand ET will provide with more thoroughly tested product. According to her, free exploration will tend to develop interest of a tester towards testing. She had some confusions related to the introduction of RBTC. In this regard she mentioned that achieving basic functional requirements depends on the way one defines RBTC. MT process should facilitate the way one wants to define RBTC in order to get the maximum out of it. Further in this regard, she added that revision of these RBTC and test missions will be much easier to an extent and it will also facilitate rapid and less exhaustive test execution.

She also mentioned that MT process would not be suitable for testing the critical systems. She added that planning and managing the test activities would be done almost the same way every time. According to her MT process will provide better overview of what is tested and what needs to be tested. She further mentioned that this process would find more defects as it is using both test approaches i.e. TCBT and ET. She considered that MT process will provide cost effective testing and will also provide better risk management. She mentioned that MT process will work fine for most of the test projects and thus can provide better quality. She further pointed out that MT process will decrease the problems related to test case design as writing test cases for the areas which are not specified are always tricky. To know the expected result of such test cases often leads to many test specification defects but it depends on which level the test cases were written. According to her opinion, MT process will make it easier to introduce ET in an organization and will also encourage managers in doing so. She highlighted that MT process should have some exit criteria based on some test measurements.
5.1.4.2 Feedback of Interviewee 2

According to Mr. Alexander Andelkovic (appendix section 8.3), MT process will prove to be more efficient because it takes input from the both test approaches and hence will provide greater test coverage. He mentioned that MT process would also provide greater oracle coverage which will increase the awareness of correct functionality. He also mentioned that free exploration in MT process will increase quality and testability of product under test.

He added that RBTC would provide basic functional requirement coverage if used in areas normally down prioritized as user manual testing and language testing. In this regard, he further explained that RBTC would also ease test design where tricky scenarios are needed to be covered because it can be divided up into separate test missions. He highlighted that designing only RBTC may reduce the effort in test documentation; however it should be used as a complimentary technique. It can be used for graphical user interface testing or where ever it is difficult to define test cases. He also considered that because of RBTC, revision of existing test cases will become as “from not revised at all it will be a major change”. According to his opinion RBTC and test missions will tend to make test execution process rapid and less exhaustive.

He highlighted that usage of MT depends on the available time to test, and ideally ET can be used when no sufficient TCBT is available. He considered MT process effective in terms of planning and managing testing tasks because all the test time that is difficult to plan with TCBT can be better planned and covered to an extent by planning and managing it with ET tasks. He also added that MT process will provide better tracking of testing tasks because lot of tests are normally not traced and by using MT process they will become more traceable. He mentioned that MT process would provide better defect detection i.e. 30-40% more defects will be detected. According to him, finding more defects by the use of MT process means better product quality and lower future support cost, which means it, is a cost effective approach. He highlighted that this process will provide better risk management because more test areas will be covered thus increasing risk awareness.

According to his opinion, this process will provide better regression testing and is well suited for it. He pointed out that MT process might not be sufficient in determining the overall quality of testing because acceptance testing is also required. According to him, MT process will encourage manager in formal implementation of ET because in future documenting and reporting of all the test activities will be required. He added that it would be impossible to have debriefing sessions quite often, if a lot of testers are included in a test process.

5.1.4.3 Feedback of Interviewee 3

According to Petter Mattsson (appendix section 8.3), MT can be termed as an effective if compared distinctly with both ET and TCBT. He also added that when the time is too short and a strange product behavior is needed to be verified in the end of the project then ET should be preferred before TCBT. He considered introduction of RBTC in the MT process as a good way of reducing the test case size because it is extremely hard to document negative tests. In addition, RBTC will also reduce the effort in test case documentation and revision of these sorts of test cases would be much easier as compared to traditional test cases. Furthermore, while designing RBTC inputs can be taken from customers and also from less technical people in test specification reviews. They know about their desired functionality of the product but they are often not capable of judging, if a negative test case will be good at finding defects or any unwanted behavior. These should be handled by ET. However to verify e.g. a requirements specification, scripted test cases are required to have a balance with ET in order to look for critical bugs. In his opinion, if some product is needed to be
verified in a tight schedule then only ET will be preferred over MT as documenting even RBTC may take some time.

He considers that planning and managing testing tasks should not be based on which technique or approach you want to use. The plan should be more based on the current context of testing. However, he is of the opinion that having MT process as a part of your test toolbox is something worthy because it provides you with a mechanism to cover requirements and at the same time you look for defects. He said that MT might provide better defect detection and risk management if only compared with TCBT. ET on the other hand in the MT process can be utilized to discover risky areas. He also added that MT might also be cost efficient as you will be able to get rid of bunch of test cases that were created to find defects. He supports the fact of MT that these bundles of test cases should be replaced with test missions, which take less time. He also mentioned that MT would provide better view of quality as testing is being performed by utilization of both test approaches. He is also of the opinion that MT tends to provide better regression testing as RBTC will secure functionality and ET will find defects or unwanted product behavior. He further added that free exploration in MT can be considered as a beneficial tool to verify something very quickly. In some situations free exploration might be the only approach that is possible to be used according to the time you got for testing.

According to him, MT will also encourage managers in the formal implementation of ET along with TCBT as managers wants all sorts of figures and test results along with defect detection reports. This approach seems to provide all that to managers, which definitely will create room for ET.

5.1.4.4 Feedback of Interviewee 4

According to Mr. Herman Afzelius (appendix section 8.3), MT can be more efficient but real time implementation and testing will suggest how these two approaches correspond to each other in MT process e.g. doing ET in the beginning of the project can prove to be more efficient. He further added that having RBTC may provide basic requirement coverage but it depends on the testing objective and the definition of basic functionality. He was of the opinion that RBTC may reduce the effort of documentation in MT and revision of the RBTC will become easy.

He mentioned that preference to use MT also correlates with the testing objectives that need to be achieved. In his opinion, managing and planning testing tasks can be done in an effective way while following MT process as ET is based on the actual context. It is also dependent on the selected strategy, which is chosen for TCBT and ET. He highlighted that it is quite early to say about the defect detection and task tracking capabilities of this approach, however it will provide more flexibility in terms of testing. He termed this approach as efficient in terms of better risk management. He further added that MT may prove to be cost efficient but it still depends on the amount of mix that is being used, however the part of this approach associated with ET can be considered as cost efficient. It is hard to say about the quality that this approach will bring, as quality of testing depends on the measure and the way you define quality and testing in the organization. He also added that MT is suitable for all types of testing. According to him, free exploration in this approach will develop motivation and creativity amongst the test engineers if it’s managed and used properly.

5.1.5 Discussion

All the interviewees were of the opinion that MT process is well structured and has the ability to cope with the weaknesses of both test approaches. However, there were some minor concerns of interviewees, who were related to the real implementation of this process.
Since, the actual behavior of this proposed process could only be judged in detail when it will be followed for some real project. Hence no further changes were made in the process regarding these concerns. In dynamic validation, test design and test execution phases of MT process were validated due to time and scope limitations of this study.

5.2 Dynamic Validation

In this section, we present the results of the dynamic validation of the proposed MT process, which was performed by conducting an experiment in an industrial setting. The details of the experiment design, actual experiment execution and results are presented in the following sections.

5.2.1 Experimentation

The goal associated with the experimentation is to make sure that important facets related to the experiment are properly defined before the actual planning and execution [41].

5.2.1.1 Definition

The purpose of this experiment was to evaluate the MT based on ‘the total number of detected defects’ and ‘functionality coverage’ against TCBT and ET in a given constant time (t). The conclusions from the experiment were inferred by collecting and interpreting experimental data. The experiment was executed in an industrial environment and subjects were the software testers. The test cases and test missions were provided by the authors to the subjects at the beginning of the experiment.

Below are the activities of the definition phase:

5.2.1.2 Goal Definition

The goal of this empirical study was to validate the MT against ‘the total number of detected defects’ and the ‘total number of functionality coverage’. The reason behind selecting these two measures is that they can be used to evaluate the benefits of each test approach, which we identified in the literature review and from interviews. For ET, defect detection capability is claimed as a benefit, while better functionality coverage is stated as a benefit for TCBT. We aim to compare these two measures for the proposed MT as it was defined as strength for both test approaches.

5.2.1.3 Object of Study

The object of study is testing approaches.

5.2.1.4 Purpose

The purpose of the experiment is to validate MT for the ‘total number of detected defects’ and ‘functionality coverage’ by comparing it against TCBT and ET.
5.2.1.5 Perspective

The perspective is from the researcher’s point of view.

5.2.1.6 Quality Focus

The main effect studied in the experiment is the MT, as a solution to the identified problems associated with ET and TCBT approaches.

5.2.1.7 Context

This experiment is run within the context of software testers in Logica AB.

5.2.2 Planning

In this section, aspects related to the planning and preparation of the experiment is presented i.e. selection of variables, null and alternative hypothesis, experimentation context, subjects and experiment instrumentation.

5.2.2.1 Context Selection

This experiment was conducted on the industrial practitioners; i.e. software testers of Logica AB. All the subjects had similar domain knowledge and relevant skills related to testing and the test product. A fix amount of time was allocated to the subjects during the course of experimentation.

5.2.2.2 Hypothesis Formulization

According to [41], hypothesis formulization is of essence, as it provides the basis for statistically analyzing the controlled experiment. Two types of hypothesis are formulated in an experiment i.e. Null hypothesis ($H_0$) and Alternative hypothesis ($H_1$). According to null hypothesis there exist no real patterns and trends in an experimental setting; hence no difference lies when considering different treatments for one measurement. On the contrary, alternative hypothesis states about the differences lying in the patterns and trends of the experimentation results.

For validating the total number of detected defects by using ET, TCBT and MT the null and alternative hypotheses are as follows:

- $H_0$ defect detection: MT detects equal number of defects as compared to TCBT and ET
- $H_{a1}$ defect detection: MT detects greater number of defects then TCBT
- $H_{a2}$ defect detection: MT detects greater number of defects then ET

For evaluating the total number of functionality coverage of ET, TCBT and MT the null and alternative hypotheses are as follows:

- $H_0$ functionality coverage: MT provides equal number of functionality coverage as compared to TCBT and ET
5.2.2.3 Variable Selection

There exist two kinds of variables in a controlled experiment i.e. independent variable and dependent variables. The independent should be controlled in an experiment. A factor is an effect of independent variable on the dependent variable and one distinct value associated with the independent variable is known as treatment [41]. Dependent variables are studied to examine the effect of changes occurring in independent variables. They are measured by the usage of some direct and indirect measures. The main difference between a direct and indirect measure is that, in direct measurement, there is no need of referring to other measurements, however in indirect measures other measurement are also referred. Furthermore, scale and range for the selected variables i.e. dependent and independent is determined [41].

Independent Variables:

- Testing approach

Dependent Variables

The dependent variables selected for this experiment are as follows:

- Defect Detection
- Functionality coverage

In this study, there exist one independent variable as mentioned above and two dependent variables. The factor of each independent variable is shown in Table 11.

<table>
<thead>
<tr>
<th>Type</th>
<th>Variables</th>
<th>Value</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent Variables</td>
<td>Testing approach</td>
<td>ET</td>
<td>Nominal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TCBT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>MT</td>
<td></td>
</tr>
<tr>
<td>Direct Measurements</td>
<td>Total number of detected defects</td>
<td>{0,1,2 \ldots , n}, n \in \mathbb{N}</td>
<td>Interval</td>
</tr>
<tr>
<td></td>
<td>Total number of functionality coverage</td>
<td>{0,1,2 \ldots , n}, n \in \mathbb{N}</td>
<td></td>
</tr>
<tr>
<td>Dependent variables</td>
<td>Defect detection</td>
<td>{0,1,2 \ldots , n}, n \in \mathbb{N}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Functionality coverage</td>
<td>{0,1,2 \ldots , n}, n \in \mathbb{N}</td>
<td></td>
</tr>
</tbody>
</table>

Table 11: Variables selection in the final experiment
5.2.2.4 Selection of subjects

The subjects of the experiment were selected on the basis of their relevant background and experience adhered in software testing. This experiment was conducted in an industrial setting to provide with more precise and reliable results. The company selected for execution of this research experiment was Logica AB, and 6 experienced software testers of Logica AB were chosen as experimental subjects. In order to select the appropriate subjects, a background check was performed on all of the testers by giving them a questionnaire prior to the experiment execution. The background check questionnaire can be found in appendix section 8.4. The background check questionnaire was based on questions such as their knowledge and experience related to software testing. The results gathered from the background check assisted in filtering out appropriate subjects who possessed at least two years of moderate level software testing experience.

Furthermore, an initial level presentation of the MT process was given to them along with relevant contents of the experiment. In addition, background knowledge was also provided about the study. A request was made to the test manager to ensure that the testers can allocate two hours to participate in the experiment.

5.2.3 Experiment design

In [41], it is mentioned that three design principles are mainly used in combination for most of the experiments i.e. randomization, blocking and balancing. Since the subjects of the experiment were selected from the same company and based on the background check, all the subjects possessed very similar level of experience and knowledge. Therefore, blocking and randomization principles were not required for this experiment i.e. subjects have the same experience and are not randomly selected. This experiment has a balanced design as same numbers of subjects are used for each treatment having same background.

There are number of experiment design types which can be selected on the basis of factors and treatments. There are four types of experiment designs which are frequently utilized [41]:

- One factor with two treatments
- One factor with more than two treatments
- Two factors with two treatments
- More than two factors each with two treatments.

This experiment is designed as a ‘One factor with more than two treatments’ (see Table 12). The factor is testing approach while the treatments are ET, TCBT and MT.

<table>
<thead>
<tr>
<th></th>
<th>TCBT</th>
<th>ET</th>
<th>MT</th>
</tr>
</thead>
<tbody>
<tr>
<td>(6 subjects)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 12: Experiment design

5.2.3.1 Instrumentation

The instruments of an experiment i.e. guidelines, objects and measurement forms are supposed to be included in the planning phase of an experiment. All the required and necessary instruments were utilized by making sure that they are aligned with experiment design and data collection method [41].
The guidelines related to the successful execution of the experiment were provided to the subjects in order to give them with better understanding of the experiment related contents. In addition, guidelines also highlight important facets on how to perform the work during the course of experiment. As mentioned above, the subjects were provided with the basic knowledge of MT and other important aspects. Furthermore, a test application i.e. calculator was provided to the subjects in order to perform experimentation. This application was developed by the authors and it had self induced faults having same severity. Three versions of calculators having distinct faults were utilized for each test approach.

Forms were designed and provided to the subjects in order to collect the experiment related data for hypothesis testing, i.e. bug investigation report, session sheet and RBTC (appendix section 8.4). All the subjects were supposed to fill the templates which were provided, with the relevant data in each field, such as bugs, types of bugs and functionality tested etc.

5.2.4 Experiment execution

All the instruments required for the execution of the experiments were prepared and experiment subjects were chosen. All the subjects were bound to answer a background questionnaire form in order to verify their suitability as a subject for this experiment. The form was sent to them three days prior to the experiment execution and upon receiving the forms back, the subjects who fitted the selection criteria were chosen and grouped together. In total 6 subjects were chosen for the experimentation. The date, time and the place of the experiment were communicated to the selected subjects. Furthermore, subjects were also informed to bring their laptops. Figure 14 shows the experiment execution.

The experiment was divided into two sessions, first was the training session in which presentation was given to them on how to participate in the experiment. The contents of the presentation were what to do and what not to do in the experiment etc. In the second session, the task was given to the subjects and experiment was executed. This experiment was conducted on 29th April 2010 in Logica AB Kalmar, Sweden between the 16:30 and 18:30.

This initial session lasted for 15 minutes and extra 15 minutes were given to them in which they were requested to ask any questions in order to remove any ambiguities related to the experiment. It was also requested to all subjects not to talk during the experiment with each other and only concentrate on the experiment. The test application i.e. calculator was deployed on the laptops of each subject. Following instructions were given to the subjects for each test approach in the initial session.

For the ET part, no test cases were provided to the subjects. Therefore, they had to perform free testing and they were not bound to follow any test cases or test steps. They only needed to log the identified bugs in the bug sheet. Bug report template was also given to the subjects in order to log the identified bugs.
For MT, RBTC and test missions were provided to the subjects. They were supposed to execute test missions after completely performing the RBTC. While performing a test mission, subjects were told that they should look into the specified area as mentioned on the test sheet but they can use their ideas and follow whatever steps they want to. At the same time they should also write down some high-level test steps, which they performed during a session.

For TCBT, test cases were provided to the subjects and it was explicitly mentioned to the subjects to strictly follow the test steps as mentioned in the test cases. However if they felt a need to divert from the test steps they were asked to mention in the bug report, the reason and any related outcome.

In the second session i.e. task session started without any break. It started at 17:00 and ended at 18:30 allocating maximum 30 minutes to each testing approach. During this time period all the subjects performed the tasks as they were assigned. Upon completion of time i.e.18:30 all the relevant material that was given to them was collected including, test cases, RBTCs, test sessions and bug sheets. As an outcome all the subjects were supposed to fill the bug sheet with the indentified bugs.

5.3 Results and analysis

This section presents results of the collected data. In the end of this section, a discussion is provided based on the statistical evaluation of the data along with interpretation of results.

The collected data was analyzed by following two steps and it is based on the design type of the experiment. In the first step, evaluation of the collected data was carried out by using descriptive statistics.

A significant aspect of using statistics is that it summarizes the collected data in an understandable and clear way [41]. Descriptive statistics were utilized in order to graphically represent and describe about the interesting aspects related to the data sets which were illustrated in a box plot in order to analyze how the data is grouped together. The typical measured used for descriptive statistics were calculation of mean value, standard deviation, skewness and kurtosis.

The second step is hypothesis testing. In [41], it is mentioned that the main objective of hypothesis testing is to see whether the null hypothesis $H_0$ can be rejected by some significant level in order to draw some sort of conclusion related to the outcome of the experimentation. For this reason, the collected data was checked for its normal distribution. Afterwards appropriate statistical tests were selected for hypothesis testing. In addition, $P$-value is set to $p = 0.05$ which indicates statistical significance needed to reject a null hypothesis $H_0$.

There are many statistical tests available for hypothesis testing. Selection of the appropriate tests depends on the design, type and assumptions of the experiment [41]. The experiment design type for this experiment is one factor with more than two treatments.

In this experiment we have 3 treatments to compare (TCBT, ET and MT); it would have required 3 separate t-tests for comparing (TCBT with ET, TCBT with MT, and ET with MT). The main problems associated with these tests [45] is that, let’s suppose if we had seven treatments to compare then it would need 21 separate t-tests. Performing this number of tests would have been really time consuming and importantly it would be somewhat inherently flawed as in each t-test there is a 5% chance of conclusion being incorrect (
testing for \( p=0.05 \), hence in 21 tests it is expected that one test would provides us with a false result.

Analysis of Variance (ANOVA) provides solution to this problem, as it enables the researchers in detecting the significant difference between the treatments as a whole [45]. A single test is performed in order to look for the differences between the means at the chosen probability level.

### 5.3.1 Analysis of the direct measurement

In this section, the collected data is evaluated by descriptive statistics. Furthermore, hypothesis testing is performed for two direct measurements i.e. total number of detected defects and total number of functionality coverage. In addition, this section also presents the analysis to interpret the results.

#### 5.3.1.1 Defect detection

All those attributes of application which divert the application from the actual execution is termed as a defect [4]. All such attributes which were detected by following each test approach are presented in this section.

**5.3.1.1.1 Descriptive statistics**

The collected data on the number of detected defects by each experiment subject against each treatment i.e. ET, TCBT and MT is presented in Table 13.

<table>
<thead>
<tr>
<th>Subject</th>
<th>TCBT</th>
<th>ET</th>
<th>MT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>06</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>07</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>06</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>08</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>09</td>
<td>16</td>
<td>13</td>
</tr>
<tr>
<td>6</td>
<td>06</td>
<td>11</td>
<td>08</td>
</tr>
</tbody>
</table>

**Table 13: The Number of Detected Defects**

In descriptive statistics firstly we calculate the mean of the captured data set. In the next step standard deviation is calculated which indicates about how compactly the gathered observations are clustered around the calculated mean [46]. Following formula is used to calculate the standard deviation.

\[
\sigma = \left[ \frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \bar{x})^2 \right]^{1/2}
\]  

Furthermore, skewness is calculated to determine whether the curve is skewed more towards left or right. If the calculated value is negative it shows that the curve is skewed more towards left and if the calculated value is positive it means that it is more skewed towards right [46]. Hence symmetric data set should have skewness near to zero. All the below values are near to zero which means that it is normally distributed. The formula for calculating skewness is as follows:
Another normality test is known as kurtosis normality test which is used to measure about whether the data are flat or peaked to a normal distribution [46]. As below all the calculated kurtosis values are positive this means that all values are peak disturbed.

\[
skewness = \frac{\sum_{i=1}^{N} (Y_i - \bar{Y})^3}{(N - 1)s^3}
\]

\[
kurtosis = \frac{\sum_{i=1}^{N} (Y_i - \bar{Y})^4}{(N - 1)s^4}
\]

The Table 14 below presents the descriptive statistics of the three treatments, including mean, standard deviation, skewness and kurtosis:

<table>
<thead>
<tr>
<th>Test Approach</th>
<th>Descriptive Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>TCBT</td>
<td>7</td>
</tr>
<tr>
<td>ET</td>
<td>13.16</td>
</tr>
<tr>
<td>Mix Test</td>
<td>10.67</td>
</tr>
</tbody>
</table>

Table14: Descriptive statistics for Number of defect detection of three treatments

Figure15: Box plots of defect detection for three treatments

5.3.1.1.2 Hypothesis testing for defect detection

In order to test the hypothesis $H_0$ of defect detection, following steps of ANOVA analysis were followed as presented in [45].
Step 1:
Sum of all the defects are taken for each test approach in the very first step in order to get \( \sum x \). Furthermore sum is taken of these calculated values to obtain grand total. Afterwards the value obtained is squared to obtain \( (\sum x)^2 \). Normality test is performed to attain normal distribution of the sample data for this purpose we have to calculate the mean of each test approach i.e. \( \bar{x} \). This test is a statistical process which is used to determine whether the sample data is standard normal distribution or not. Following formula is used to calculate it:

\[
\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i
\]

(4)

In the above Equation (4) \( n \) represents the sample size, and \( x_i \) represents the value of \( i^{th} \) observation.

In the last step, we obtain the square of each data value and sum all squares in order to get \( \sum x^2 \) which is the sum of square of each value.

As all the measurements in the sample differs from the calculated mean by an amount known as deviation \( (d) \) and for every observation known as \( (x) \), the deviation from the mean is calculated as:

\[ x - \bar{x} \]

Hence:

\[ d^2 = (x - \bar{x})^2 \]

If this equation is further expanded we get:

\[ d^2 = x^2 - 2x \bar{x} + (\bar{x})^2 \]

In order to get the sum of squares of the calculated deviations, we have to sum both sides of the above equation; hence we get the following equation.

\[ \Sigma d^2 = \Sigma x^2 - 2 \Sigma x \bar{x} + \Sigma (\bar{x})^2 \]

From the above equation following very important equation is derived in order to calculate the sum of squares as \( \Sigma d^2 \).

\[
\sum d^2 = \sum x^2 - \frac{(\sum x)^2}{n}
\]

(5)

Afterwards the sample of variance is calculated which provides with the measurement of how the data is distributed across the mean and expected values. In addition, standard deviations of the sample are measured which depicts about how much the data is spread out.

Step 2:
The sample of the variance i.e. \( \sigma^2 \) is attained by summing all the calculated \( d^2 \) values and this is then divided by \( n-1 \) which is number of degree of freedom in order to get the
By doing so, we can get the standard deviation presented as $\sigma^2$ which is square root of the calculated variance.

A simple formula is used to calculate the variance of the samples (notation, $\sigma^2$), which is also known as variance in a population,

$$\frac{\sum d^2}{n - 1}$$  \hspace{1cm} (6)

**Step 3**

According to the procedure described in [45] we sum all values of $\Sigma x^2$ and address them as $A$ and later we sum all the calculated values for $\frac{\Sigma x}{n}$ and call it the sum B. In addition to that sum of all the calculated values for $\Sigma x$ is utilized to get the grand total.

<table>
<thead>
<tr>
<th>Test Approach</th>
<th>TCBT</th>
<th>ET</th>
<th>MIX</th>
<th>Row totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>06</td>
<td>12</td>
<td>10</td>
<td>28</td>
</tr>
<tr>
<td>2</td>
<td>07</td>
<td>13</td>
<td>11</td>
<td>31</td>
</tr>
<tr>
<td>3</td>
<td>06</td>
<td>15</td>
<td>12</td>
<td>33</td>
</tr>
<tr>
<td>4</td>
<td>08</td>
<td>12</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td>5</td>
<td>09</td>
<td>16</td>
<td>13</td>
<td>38</td>
</tr>
<tr>
<td>6</td>
<td>06</td>
<td>11</td>
<td>08</td>
<td>25</td>
</tr>
<tr>
<td>$\Sigma x$</td>
<td>42</td>
<td>79</td>
<td>64</td>
<td>Grand Total 185</td>
</tr>
<tr>
<td>$N$</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>$\frac{x}{x}$</td>
<td>7</td>
<td>13.16</td>
<td>10.67</td>
<td></td>
</tr>
<tr>
<td>$\Sigma x^2$</td>
<td>302</td>
<td>1059</td>
<td>698</td>
<td>(call this A) 2059</td>
</tr>
<tr>
<td>$\frac{(\Sigma x)^2}{n}$</td>
<td>294</td>
<td>1040</td>
<td>682.7</td>
<td>(call this B) 2016.83</td>
</tr>
<tr>
<td>$\Sigma d^2$</td>
<td>8</td>
<td>19</td>
<td>15.3</td>
<td>(A - B) 42.17</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>1.6</td>
<td>3.8</td>
<td>3.06</td>
<td></td>
</tr>
</tbody>
</table>

Table 15: Relevant statistics for defect detection of three treatments

**Variance Ratio**

The variance ratio donated as (F) which is the ratio of distribution of independent estimates and is calculated by dividing the largest calculated values of $\sigma^2$ by the least calculated value.
of $\sigma^2$. This obtained value is then looked into the F-max Table 30 (appendix section 8.5.1) for total number of treatments as in our data table, in conjunction with the degree of freedom which is the total number of replicates divided by number of treatments-1. If the calculated variance ratio does not exceed the $F_{\text{max}}$ value, it shows that we can safely proceed, if not the case then we need to transform the data.

$F_{\text{max}}$ test: $F = \frac{3.8}{1.6} = 2.2$. This value is lower than the value of $F_{\text{max}}$ which is 10.8 for the 3 treatments and 5 df at $P = 0.05$, hence it can be concluded that the variances are homogenous. This indicates that we can further proceed with analysis of variance. If this value would have exceeded than that of tabulated value of $F_{\text{max}}$ then we would have to transform the data.

**Step 4**

In step 4, we square the calculated grand total and it is then divided by the total number of observations. Let’s call it D, hence the formula is:

$$D = \frac{(\text{Grand total})^2}{\text{total observations}} = \frac{(185)^2}{18} = 1901.3$$

**Step 5**

In step 5, we calculate the total sum of squares which is $(S\text{ of } S) = A - D$. Hence

$$(S\text{ of } S) = A - D = 2059 - 1901.3 = 157.7$$

**Step 6**

In step 6, Between-treatments sum of square is calculated which is as follows:

$$S\text{ of } S = A - B = 2059 - 2016.83 = 42.17$$ (also known as Error sum of squares)

**Step 7**

In step 7, Residual sum of square is calculated as follows:

$$S\text{ of } S = A - B = 2059 - 2016.83 = 42.17$$ (also known as Error sum of squares)

**Step 8**

In step 8, Table 16 is constructed which shows the values of sum of squares

<table>
<thead>
<tr>
<th>Source of variance</th>
<th>Sum of squares (S of S)</th>
<th>Degrees of freedom (df)</th>
<th>Mean square ($= S\text{ of } S \div df$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between treatments</td>
<td>115.53</td>
<td>$u - 1 = 2$</td>
<td>57.765</td>
</tr>
<tr>
<td>Residual</td>
<td>42.17</td>
<td>$u(v-1) = 15$</td>
<td>2.811</td>
</tr>
<tr>
<td>Total</td>
<td>157.7</td>
<td>$(uv)-1 = 17$</td>
<td></td>
</tr>
</tbody>
</table>

Table 16: Values of sum of squares for defect detection
\( u = \) No of treatments and \( v = \) No of replicates (The total “Degree of freedom” (df) is one less than the total no of data entries)

By utilizing the mean squared in the last column of the above Table 16, variance ratio test is derived in order to get the F value which is as follows:

\[
F = \frac{\text{between treatments mean square}}{\text{Residual mean square}}
\]

\[
F = 57.765 \div 2.811 = 20.5496
\]

**Step 9**

In step 9, Table 31 (appendix section 8.5.2) of F for \( p=0.05 \) is looked for the value of \( n_1 \) which is the df between the treatments mean squares while \( n_2 \) categorized as the df of the residual mean square. If the value of F surpasses the tabulated value, it means that there exists a considerable difference between the treatments. In our case the value 20.5496 exceeds the given value of F which means that significant difference between treatments exists.

The Analysis of Variance showed that there exists difference between the treatments as a whole in the experiment. This information can be useful to some extent, however it does not demonstrate about which treatment differs from other.

Now there is a problem as, at every instant we are comparing one treatment with another i.e. comparing ET and TCBT with MT, hence we are performing equivalent t-test, and there is a fair amount of probability in making incorrect interpretation of results. Thus we are required to have some way in order to avoid from this problem.

**Method of calculating least significance difference**

Least significance difference is calculated in this method between the two means. We assume that the calculations of ANOVA are same as that of a t-test. Hence the residual mean square is considered as an estimate of \( \sigma^2 \) for all the treatments as variance for each treatment is supposed to be same in Analysis of Variance.

\( \sigma_d^2 \) in the t-test is calculated by using the following formula

\[
\sigma_d^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}
\]

\( \sigma^2 \) is considered to be as the equal for all the treatments in the analysis of variance: hence if \( n \) for all the treatment is equal, then comparison of any two means can be done by calculating the \( \sigma_d^2 \) as follows:

\[
\sigma_d^2 = 2 \times \text{residual mean square} / n
\]

The value of \( t \) can be calculate by taking the square root of \( \sigma_d^2 \) and making it \( \sigma_d \) so,

\[
t = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_d}
\]

If this was done for the two means, we can then compare the obtained value of \( t \) with the one in a t-table by utilizing the df of the residual mean square as it shows the residual variances in the overall experiment.
The simpler way of comparing any two means is, if we take the above equation and multiply both sides by $\sigma_d$

$$\text{We get } t(\sigma_d) = \bar{x}_1 - \bar{x}_2$$

Hence, any two means would have significant difference amongst them, if the difference is greater than that of "$t$ multiplied by $\sigma_d$"; this represents the least significance difference between the two means.

In the Table 17 below the means of the three treatments are given which depicts the analysis of the variance which is used to find out the least significant difference between the three treatments i.e. ET, TCBT and MT at p=0.05 (Probability level chosen against the value of t).

<table>
<thead>
<tr>
<th>Significant difference b/w two treatments</th>
<th>$t(\sigma_d) = \bar{x}_1 - \bar{x}_2$</th>
<th>Least significance difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ET-TCBT</td>
<td>13.16-7</td>
<td>6.16</td>
</tr>
<tr>
<td>ET-MT</td>
<td>13.16-10.67</td>
<td>2.49</td>
</tr>
<tr>
<td>MT-TCBT</td>
<td>10.67-7</td>
<td>3.67</td>
</tr>
</tbody>
</table>

Table17: Least significance difference for defect detection

Difference between the two treatment is considered as significant if the calculated value of t is greater than the value as given in the Table 32 (appendix section 8.5.3 ) at the probability level p=0.05. As shown in the above Table 17 the least significant difference is greater than the given value i.e. 2.45, so it can be concluded that, $H_0$ defect detection and $H_a$ defect detection are rejected. However $H_a$ defect detection is accepted i.e. MT provides greater defect detection than TCBT but less than ET.

5.3.1.2 Test Coverage

Bugs related to main functionality tells how much coverage is achieved. All the other bugs in this case are ignored which do not relate to the basic functions. By coverage it means covering the basic requirement functionality.

5.3.1.2.1 Descriptive statistics

The collected data is presented in the Table 18 i.e. total number of functionalities covered by each experiment subject against each treatment i.e. ET, TCBT and MT in order to calculate the coverage.

<table>
<thead>
<tr>
<th>Subjects</th>
<th>TCBT</th>
<th>ET</th>
<th>MT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 18: Functionality Coverage
The Table 19 below presents the descriptive statistics of the three treatments, including mean, standard deviation, skewness and kurtosis are calculated as presented in the section 5.3.1.1.1

<table>
<thead>
<tr>
<th>Test Approaches</th>
<th>Descriptive Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>TCBT</td>
<td>7.833</td>
</tr>
<tr>
<td>ET</td>
<td>6.667</td>
</tr>
<tr>
<td>MT</td>
<td>8.833</td>
</tr>
</tbody>
</table>

Table 19: Descriptive statistics for functionality coverage of three treatments

5.3.1.2.2 Hypothesis testing for functional coverage

In order to test the hypothesis $H_0$ functionality coverage, the steps of ANOVA analysis are used as presented in [45] and discussed in section 5.3.1.1.2.

<table>
<thead>
<tr>
<th>Test Approach</th>
<th>TCBT</th>
<th>ET</th>
<th>MT</th>
<th>Row totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>5</td>
<td>7</td>
<td>19</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>6</td>
<td>8</td>
<td>23</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>9</td>
<td>11</td>
<td>28</td>
</tr>
</tbody>
</table>

Figure 16: Box plot of functionality coverage for three treatments
F_{max} test: F = 2.2/0.568 = 3.87. This value is lower than the value of F_{max} (appendix section 8.5.1) which is 10.8 for the 3 treatments and 5 df at P = 0.05, hence it can be concluded that the variances are homogenous. This indicates that we can further proceed with analysis of variance. If this value would have exceeded than that of tabulated value of F_{max} then we would have to transform the data.

Now we will perform the following steps as mentioned in the section 5.3.1.1.2

**Step 4**

In step 4, we square the calculated grand total and it is then divided by the total number of observations. Let’s call it D, hence the formula is:

$$D = (\text{Grand total})^2 \div \text{total observations} = (140)^2 \div 18 = 1088.88$$

**Step 5**

In step 5, we calculate the total sum of squares which is (S of S) = A - D. Hence

$$(S \text{ of } S) = A - D = 1126 - 1088.88 = 37.12$$

**Step 6**

In step 6, Between-treatments sum of square is calculated which is as follows:

$$S \text{ of } S = B - D = 1102.83 - 1088.88 = 13.95$$
In step 8, Residual sum of square is calculated as follows:

\[
S = A - B = 1126 - 1102.83 = 23.17
\]

(also known as Error sum of squares)

The below Table 21 is constructed which shows the values of sum of squares

<table>
<thead>
<tr>
<th>Source of variance</th>
<th>Sum of squares (S of S)</th>
<th>Degrees of freedom (df)</th>
<th>Mean square (= S of S / df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between treatments</td>
<td>13.95</td>
<td>( u - 1 = 2 )</td>
<td>6.975</td>
</tr>
<tr>
<td>Residual</td>
<td>23.17</td>
<td>( u(v-1) = 15 )</td>
<td>1.544</td>
</tr>
<tr>
<td>Total</td>
<td>37.12</td>
<td>((uv)-1 = 17)</td>
<td></td>
</tr>
</tbody>
</table>

Table 21: Values of sum of squares for functionality coverage

\( u = \) No of treatments and \( v = \) No of replicates (The total df is one less than the total no of data entries)

By utilizing the mean squared in the last column of the above table, variance ratio test is derived in order to get the F value which is as follows:

\[
F = \frac{\text{between treatments mean square}}{\text{Residual mean square}} = \frac{6.975}{1.544} = 4.51
\]

**Step 10**

In step 10, Table 31 (appendix section 8.5.2) of F for (p= 0.05) is looked for the value of \( n_1 \) which is the df between the treatments mean squares while \( n_2 \) categorized as the df of the residual mean square. If the value of F surpasses the tabulated value, it means that there exists a considerable difference between the treatments. In our case the value 4.51 does not exceed the given value of F which means that there exists no significant difference between treatments.

The Analysis of Variance showed that there exists no difference between the treatments as a whole in the experiment.

In the Table 22 below, means are given of the three treatments which depicts that the analysis of the variance which is used to find out the least significant difference between the three treatments i.e. ET, TCBT and MT at p=0.05 (Probability level chosen against the value of t)

<table>
<thead>
<tr>
<th>Significant difference b/w two treatments</th>
<th>( t (\sigma_d) = \bar{x}_1 - \bar{x}_2 )</th>
<th>Least significance difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCBT-ET</td>
<td>7.833-6.67</td>
<td>1.1637</td>
</tr>
<tr>
<td>MT-TCBT</td>
<td>8.833-7.833</td>
<td>1</td>
</tr>
<tr>
<td>MT-ET</td>
<td>8.833-6.67</td>
<td>2.163</td>
</tr>
</tbody>
</table>

Table 22: Least significance difference for functionality coverage
Difference between the two treatment is considered as significant if the calculated value of $t$ is greater than the value as given in the Table 32 (appendix section 8.5.3) at the probability of level $p=0.05$. As shown in the above Table 22 the least significant difference is less than the given value i.e. 2.45, so it can be concluded that, $H_0$ functionality coverage cannot be rejected hence it can be concluded that there is no significant difference of functionality coverage between ET, TCBT and MT.

5.3.2 Discussion and summary

The main contribution of this chapter was static and dynamic validation of the proposed MT process. Communicating this process to the industry professionals, who then critically reviewed it and provided us with their opinions, did static validation. As an outcome of static validation many of the feedbacks suggested that MT process has the tendency to resolve the issues of both test approaches to an extent. However the actual outcome can only be seen after its practical and customized implementation in a real environment.

In the dynamic validation, two of the aspects of MT process were validated; i.e. defect detection and functionality coverage. The results suggested that MT has better defect detection than TCBT but less than ET. On the other hand, while conducting analysis on the functionality coverage it was revealed that there is no significant value upon which null hypothesis could have been rejected. Hence all the three test approaches resulted in having almost same functionality coverage.

Furthermore, the subjects who participated in the experiment were of the opinion that following each and every explicit test step while performing TCBT was quite annoying and time consuming. However while performing ET; subjects highlighted that they experienced difficulty in tracking their own test steps. MT was a new experience for the testers, because they mentioned that adding structured ET was quite facilitating because test missions focused their testing abilities on certain functionalities. In addition, they also had the freedom to make their own test steps based on their own intuition. Discussing about RBTCs with the subjects their opinion was that it was very easy to follow simple and short test steps in quick intervals.

It is worth mentioning that the application, which was used for the experiment, was a simple calculator and had only 11 functionalities, which might have influenced the results. Therefore, a similar experiment should be conducted in the future on more complex and bigger applications having more functionality to be able to generalize the results.

5.3.3 Validity threats

Below mentioned are the validity threats related to the experiment.

5.3.3.1 Internal validity

There was a threat that subjects might influence the experiment in a negative manner, as during the course of the experimentation their behavior and commitment should be a key factor. This was not a threat for this case as the experiment was non-beneficiary.

The threat related to the instrumentation of the experiment was minimized by providing all the instruments beforehand to the subjects.

Maturation is the threat which is a problem in this study, since the experiment subjects evaluated all approaches on the same occasion without any delay. This means that the testers
remembered the way of testing while performing MT, this effect was minimized by the use of different versions of test application.

The threat of mortality was reduced by choosing subjects on the basis of relevant background and knowledge possessed by the subjects related to the software testing.

5.3.3.2 Construct validity

The threat of evaluation apprehension is reduced by informing all the subjects that their performance in the experiment would not be used for any sort of performance assessment.

Hypothesis guessing is a risk that experimental subject tries to figure out i.e. the intention and outcome of this experiment. This thing can lead to wrong and interrupted conclusion. This threat was minimized by not opening the purpose of experiment to the experimental subjects.

The application which was used for the experiment was a simple calculator and had 11 functionalities, which might have affected the experimentation results. We could not make an experiment with a larger and more complex application due to the limited experiment execution time that was available in the company.

5.3.3.3 External validity

The threat related to external validity is the limitation in generalizing the results of the experiment. This experiment was performed with 6 testers in industry, and only one application i.e. calculator was used for this purpose. Hence, this could be considered as vital threat to external validity.

5.3.3.4 Conclusion validity

The main threat regarding the conclusion validity is the quality of the data collected during the experimentation. The subjects were expected to deliver the experimentation data, hence risk of delivering fake or incorrect data was ruled out by providing them with reporting templates. All the subjects were bound to use these templates in order to report the experimentation data.

As the nature of experiment was voluntary and was conducted on industry professionals, so it has quite low statistical power due to the low sample size. The statistical power can be increased in future by increasing the population size of experiment subjects.
6 EPILOGUE

6.1 Conclusion and discussion

In this thesis, MT process was developed and validated. In the first step, the weaknesses and strengths of manual test approaches i.e. TCBT and ET were identified. This was done by first conducting a systematic literature review, which was then followed by semi-structured interviews in order to obtain strengths and weaknesses as observed by industry professionals.

The results showed that there exists a gap amongst the perceived strengths and weaknesses of ET and TCBT keeping industrial and the theoretical perspectives in context. It was also observed that there was no formal process of MT being used in the industry in order to get benefit from the both approaches. For this reason, in the second step these indentified weaknesses and strengths of both test approaches were mapped to each other to define a MT process that incorporates the strengths of both ET and TCBT. This process definition was also based on the inputs and suggestions of industry professionals which were obtained by conducting interviews. In the third step, this process was statically validated by getting the feedback from the industry professionals. Furthermore, MT was dynamically validated by conducting an experiment in an industrial setting to compare its effectiveness in terms of defect detection and functionality coverage against ET and TCBT.

However, one should not only focus and limit MT in terms of only defect detection and functionality coverage. For this study these two measures were selected because of the limited scope and time. The idea was to make MT repeatable by thoroughly defining a process, so that it facilitates both TCBT and ET and treat them as one combined approach. Hence, if any organization tends to take advantage of these two approaches in combination; MT process can provide assistance in this regard.

This research study covers the gap of having no formal MT process by introducing a validated process of MT. The results gained by static validation showed that MT provides solution to the weaknesses of both test approaches to some extent. Whereas the dynamic validation showed that MT has better defect detection than TCBT but less than ET; however MT has equal functionality coverage as compared to both test approaches. In addition, MT process provides a mechanism for companies that are willing to get advantages by utilizing both test approaches. This process also facilitates those organizations which are attempting to implement ET and desires to reduce their test bed size to a considerable extent.

6.1.1 Answer to the research questions

In this section, results are mapped to the relevant research questions in order to present and verify its completeness. Based on the results and outcomes of the research study, each research question along with its answer is explained below.

RQ1: What are the weaknesses related to the TCBT and ET?

Section 3.1.8 and section 3.2.5 present the weaknesses related to TCBT and ET which were obtained through systematic literature review and by conducting interviews. In these sections many weaknesses have been identified, however the most important and precise weaknesses are mentioned. The perceived weaknesses of TCBT are less defect detection, problems in test case design, time consuming and non-skilled activity. The perceived weaknesses of ET are functional coverage, traceability issues, repetition of defects, risk management issues, not
suitable for complex application, oracle issues, difficult to interpret test results, companies are often reluctant to accept ET and effectiveness of ET is highly correlated with the tester’s skills.

RQ2: What are strengths related to the TCBT and ET?

Section 3.1.8 and section 3.2.5 present the strengths related to TCBT and ET which were obtained through systematic literature review and by conducting interviews. In these sections many strengths have been identified, however the most important and precise strengths are mentioned. The perceived strengths of TCBT are functional coverage, provide oracles, prediction of quality, test guidance, identification of bugs before actual testing and accommodation of legal requirements. The perceived strengths of ET are defect detection efficiency, cost effectiveness, flexibility in test analysis, good for regression testing, customer satisfaction, focused documentation and better utilization of tester skills.

RQ3: How can a mixed process for MT be defined so that, it would address the weaknesses of TCBT and ET and incorporate the strengths of these approaches?

The RQ3 was answered by conducting analysis on the answers of RQ1 and RQ2 in conjunction with the suggestions of industry professional by conducting interviews. The answer is presented in detail in section 4.3.5

RQ 4: How effective is the proposed MT process in comparison to individual testing approaches?

Static validation of the MT process suggested that MT has ability to resolve the weaknesses of both test approaches to some extent such as issues related to test oracles, functional coverage, tracking of testing tasks, risk management, cost effectiveness, regression testing, test design, time consumption and it will also encourage managers in formal implementation of ET. The answer is presented in detail in section 5.1.4

- RQ4.1: How effective the MT will be in terms of defect detection as compared to TCBT and ET?

The results generated by hypothesis testing and statistical analysis illustrate that, for defect detection, the specified null hypothesis $H_0$ defect detection and one alternative hypothesis $H_{a2}$ defect detection are rejected, however the other alternative hypothesis $H_{a1}$ defect detection is accepted i.e. MT has greater defect detection than TCBT.

- RQ4.2: How effective the MT will be in terms of functionality coverage as compared to TCBT and ET?

The results generated by hypothesis testing and statistical analysis illustrates that, for functionality coverage evaluation, the specified null hypothesis $H_0$ functionality coverage is accepted, however the other two alternative hypothesis $H_{a1}$ functionality coverage and $H_{a2}$ functionality coverage are rejected i.e. MT provides equal number of functionality coverage as compared to TCBT and ET.

As presented above MT has less defect detection ability than ET but better than TCBT. On the other hand MT has equal functionality coverage compared to both test approaches. Furthermore, the total time spent on designing the test cases for TCBT and MT was also taken in account. If we add the test case design time in the test execution time than it further decreases the worth of TCBT. By adding the total testing time of TCBT i.e. (Test case design
time + Test case execution time) shows that this time is double than that of MT (RBTC design and execution time). Hence based on this observation it can be further concluded that MT is better than TCBT. The Table 23 presents the details of total time consumed by each testing approach.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Design Time</th>
<th>Execution Time</th>
<th>Total Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCBT</td>
<td>1.5 Hours</td>
<td>0.5 Hours</td>
<td>2 Hours</td>
</tr>
<tr>
<td>MT</td>
<td>0.5 Hours</td>
<td>0.5 Hours</td>
<td>1 Hours</td>
</tr>
<tr>
<td>ET</td>
<td>0 Hours</td>
<td>0.5 Hours</td>
<td>0.5 Hours</td>
</tr>
</tbody>
</table>

Table 23: Comparison of total testing time

Furthermore, MT utilized test missions to test other scenarios which provided with better defect detection than TCBT and in addition, it also saved the time in documenting test cases prior to execution. Finally, the study indicates that MT has equal functional coverage compared to both test approaches and provides better defect detection than TCBT.

### 6.2 Future Work

In this section, we provide future research possibilities. The time period of the systematic review should be further broadened in order to get more data on the relevant topic. The factors which affect the effectiveness of ET should also be studied in detail keeping cognitive sciences in context. Domains like beta and unit testing can also be studied in order to get broader aspect of those attributes that may be useful in understanding ET. MT, TCBT and ET should also be compared by keeping bug severity and type in context so that each testing approach can be further studied in this regard.

The proposed process of MT also needs to be further validated in an industrial setting for a real project, so that the actual outcomes of this can be analyzed. In addition, large scale suggestions and opinions of industry professionals should also be taken in account on MT process in order to further mature this process for practical usage.
7 REFERENCES


8 APPENDIX

8.1 Interview Questionnaires

8.1.1 Questionnaires related to problems of ET

Q1. Have you come across with oracle issues while performing ET? (If yes then how you cope with it)
Q2. Have you experienced test coverage issues while performing ET? (If yes then how you cope with it)
Q3. In what scenarios do you think ET is not preferred on other testing approaches?
Q4. Do you think it is difficult to prioritize tests in ET?
Q5. Have you experienced any issues related to planning and managing of testing tasks in ET?
Q6. Do you consider tracking of tasks as an issue in ET?
Q7. How you define the quality of testing keeping ET in context?
Q8. Do you think ET is sufficient in determining the quality of testing? (Please specify)
Q9. Have you experienced any problems while performing regression testing in ET? (Please specify)
Q10. Do you think it is difficult to perform the test reviews?
Q11. Have you come across with any test repeatability of issues in ET, how you cope with it? (Please specify)
Q12. Please list down any problems related to ET which you have experienced?
Q13. How you think that free explorations affects the testing?
Q14. Do you think manager are reluctant in formal implementation of ET approach for testing? (Please explain)

8.1.2 Questionnaires related to benefits of Exploratory Testing

Q1. What factors you think of which can make ET more beneficial?
Q2. Do you think that ET is more efficient in defect detection as compared to other test approaches? (Please specify the reasons)
Q3. Do you think ET is an effective approach for investigation and isolation of defects?
Q4. Do you think that ET is a cost effective approach? (Please specify)
Q5. Do you find free exploration of application as an advantage?
Q6. Do you think better regression testing can be achieved by utilizing ET? (Please specify how)
Q7. Do you think ET is helpful in investigation of particular risks area of software?
Q8. Please list down the specific factors which caused the need of introducing ET?
Q9. What benefits you observed after introducing ET in your company?
Q10. Do you think your customers are satisfied by the use of ET?
Q11. Please list down any perceived benefits of ET?

8.1.3 Questionnaires related to problems of Test Case Based Testing

Q1. Do you think that designing of test cases is time consuming and an expensive activity? (If yes specify)
Q2. Have you come across with issues while revising existing test cases?
Q3. In your experience have you found test case execution as an exhaustive process?
Q4. Do you think TCBT is suitable for the regression testing?
Q5. Do you think pre designed test cases are sufficient for the entire system life cycle?
Q6. Do you think that test cases are durable (If not please specify)
Q7. Do you think that redesigned test cases are sophisticated then old one (if yes please specify)
Q8. Please list down any problems related to TCBT that you have come across?

8.1.4 Questionnaires related to benefits of Test Case Based Testing

Q1. Do you consider TCBT as an effective way of detecting and discovering faults?
Q2. Do you think TCBT is an effective way for formulizing and guiding the testing tasks?
Q3. What benefits you think of, which can be achieved while designing and planning the tests before the actual test execution?
Q4. Do you think test coverage is better achieved while using TCBT?
Q5. Do you think that better tracking would be achieved through TCBT?
Q6. Do you think TCBT improves the overall quality of testing?
Q7. Do you think TCBT provides with better and reliable test results?
Q8. Do you think that TCBT is beneficial where regulatory and legal requirements are needed to be fulfilled?
Q9. Do you think that TCBT can predict reliability of software? (Please specify)
Q10. Please list down any benefits related to TCBT which you have experienced?

8.1.5 Questionnaires related to MT Process definition

Q1. How you find mapping of weaknesses to strengths of both test approach?
Q2. Do you think mapping process will lead to a better definition of a MT process?
Q3. Do you consider that weaknesses are correctly mapped to the strengths?
Q4. Do you think MT process based on these mappings have the tendency to solve the problems of both test approaches?
Q5. What are your suggestions on the preliminary defined process of MT?
Q6. Do you think that the preliminary defined MT process solves the weaknesses of both test approaches?
Q7: How this MT process can be further enhanced in order to meet industrial needs?
Q8: Do you think this MT process covers the weaknesses of both test approaches?
Q9: Do you consider that this MT process will have some weaknesses associated with it?
Q10: Please mention about the drawbacks associated with preliminary MT process definition?

8.1.6 Feedback of MT process

Q1. Do you think that MT will be more efficient then TCBT and ET if used separately?
Q2. Do you think that the proposed MT process will address the oracle issue as it was observed while performing ET only?
Q3. Do you think introduction of RBTC will provide basic functional requirement coverage to an extent?
Q4. In what scenarios do you think MT process will not be preferred on other testing approaches and why?
Q5. Do you think that in MT process you can plan and manage testing task more effectively?
Q6. Do you think that MT process will provide better tracking of testing tasks?
Q7. Do you think that MT process will provide better defect detection?
Q8. Do you think that MT will provide better risk management?
Q9. Do you think that MT will be termed to be a cost effective approach?
Q10. Do you think MT process is sufficient in determining the quality of testing?
Q11. Do you think that MT process is suitable for performing regression testing?
Q12. Do you think that RBTC and test mission will reduce the problems related to test design?
Q13. How you think that free explorations in MT process will affect the testing?
Q14. Do you think mix test process will encourage manager in formal implementation of ET along with TCBT?
Q15. Do you think that designing only RBTC reduces the effort in test documentation?
Q16. Do you think that revision of test cases will become easy after designing RBTC?
Q17. Do you think that RBTC and test missions will tend to make test execution process rapid and less exhaustive?
Q18. Please list down any problems or suggestion you have for MT process?

8.2 Transcribed Interviews

8.2.1 Interview 1:

8.2.1.1 Interviewee

Name: Herman Afzelius  
Contact information: www.logica.se  
Email address: herman.afzelius@logica.com  
Phone: +46 76 7766373  
Date: March 5, 2010  
Start and End time: 3pm to 5pm  
Location: Ronneby, Sweden

The contacted person for interview was Mr. Herman who currently is working for Logica AB as Test manager for last 2 years. He also provided his services in Microsoft and other reputable organizations and hence has sound knowledge and experience of different test approaches being followed in industry. During his stay at UIQ technologies, he as a test manager introduced ET approach and gained the practical experience as a tester and as a manger at the same time. The interviewee was an extremely valuable resource because he was well versed on the topic of research area and proved as very beneficial.

8.2.1.2 Transcribed interview

The interviewee indentified many benefits of the ET. He was posed with many questions which were based on the benefits of ET as mentioned in the literature. He highlighted that if more structured approach for ET, it can be more beneficial. Interviewee didn’t considered ET as more efficient in detecting defects as compared to other test approaches. He instead highlighted that it is more cost efficient test approach and consumes less time in producing results which makes this approach more preferable. Furthermore interviewee considered free exploration as a better learning curve for a tester but not as any major benefit of ET because he believes in having a more structured approach as mentioned earlier. In the interview when he was asked if better regression testing can be achieved by ET his response was quite positive. He further explained in this regard that most of bugs not all are identified outside the pre defined steps of test cases, hence while performing regression testing ET should be considered at least as a complementary cross checking approach.

While discussing more about the benefits of the ET he added that ET can come handy in investigation of particular risks areas of software. He shared some of the important base
factors of introducing ET in UIQ Technology which were previous test approach was taking too much time, test cases were not efficient in detecting defects, pile of test cases were creeping up all the time, it became really hard to manage them. The benefits that he observed after introducing ET in UIQ testers were identifying more bugs in less time and they became more responsible and creative towards testing. In addition, he also pointed out that customers were more satisfied as more bugs were being identified and hence bug free software’s were being delivered. He added that not only ET is suitable but other test approaches may also provide customer satisfaction.

The second part of the interview was related to the problems of TCBT, in this regards interviewee highlighted about the problems that he coped in UIQ Technology, while performing TCBT. He said that if a test design is not proper it can be very problematic, and test cases are the just one output and problems can be inherited from any documents such as software requirement specification and test plans etc. This can be minimized to an extent by continuously communicating and meeting with all the relevant stakeholders i.e. developers, managers, testers etc. A proper design can cut down the costs of designing and re designing the test cases. He also mentioned in the interview that in UIQ they had more than 10,000 test cases for one application and revising some or all of them were very exhaustive and painstaking task. He added that test case strategy should also be taken in account as some test cases might be high-level, other may be detailed or unit level. When we asked him whether test cases were sufficient for entire test lifecycle, his response was mixed in this regard and he considered it as very unlikely and he told that executing test cases was an exhaustive and time consuming process and chances of finding defects were quite slim.

In the third quarter of interview we posed questions related to the benefits of using TCBT approach, he responded that Test cases are good way of knowing the product from multiple scenarios. He added that quality in TCBT can only be achieved if test cases are designed properly but still measuring the quality of test cases prior to actual execution is quite difficult, hence to an extent quality is directly proportional to design of test cases. He supported that TCBT provides better test coverage but it does not mean it provides better defect detection as a byproduct. He told in the interview that where fulfillment of legal and regulatory requirements is mandatory TCBT can be quite useful. In the last quarter problems of ET were discussed with the interviewee, when we posed a question related to oracle issues in ET, his response was that this issue is correlated with the type of application one is testing if it’s too complex, oracle then is an issue of ET. However if application is simple or a tester is well experienced in domain than common sense of tester plays an important role while performing ET. In addition he also mentioned that this issue can be dealt with to an extent if documentation, manuals etc are studied prior to the testing.

While discussing about the coverage issues in ET, he responded that 100 percent test coverage is not possible. However coverage can be deduced by some indirect measures such as the total time of ET and number of bugs discovered, if no problems are identified further it means more or else we have covered the functionality of a system. Another problem of ET that was highlighted by the interviewee was conducting ET on complex application alone is not suitable and it should be combined with other test approaches. He also highlighted that if a tester in ET is not using appropriate test skills and knowledge it can create many problems because ET is highly correlated with the tester’s abilities and hence more energy is required from testers. When we asked about the problems associated with free form ET, interviewee responded that in such scenario managing, planning, tracking and prioritizing the tests becomes quite difficult. For this reason a structured approach is preferred. In the last we asked him to define quality of testing and ET, and he said in this regard that quality has many definitions some consider detecting maximum number of defects as quality while other considers coverage as quality and it just goes on, hence combing ET and quality together and defining something is very difficult.
8.2.2 Interview 2:

8.2.2.1 Interviewee

Name: Petter Mattsson  
Contact information: www.logica.se 
Email address: petter.mattsson@logica.com  
Phone: +46(0)767 97 63 13  
Date: March 11, 2010  
Start and End time: 5pm to 6:30 pm  
Location: Ronneby, Sweden

The contacted person for interview was Mr Petter Mattsson who currently is working for Telenor AB as consultant for last 2 years. He has been working in many reputable organizations and hence has sound knowledge and experience of different test approaches being followed in industry. His main specialties includes Test Management, Quality Assurance, Project Management, Software Testing, Team leading, Coaching, Education, Exploratory Testing, Session-Based Test Management, Test Process Improvement, Context Driven Testing, Risk Based Testing. He also spoke at different test conferences like:

- SAST Q1 (Swedish Association for Software Testing), 2008  
- NFI “Testforum”, appointed as "Best speaker of the conference”, 2008  
- Blekinge Institute of Technology - "Test Management", 2006

During his stay at UIQ technologies, he as a test manager introduced ET approach and gained the practical experience as a tester and as a manager at the same time. The interviewee was an extremely valuable resource because he was well versed on the topic of research area and proved as very beneficial.

8.2.2.1.2 Transcribed interview

In the first quarter of interview we discussed and inquire about the benefits of ET, interviewee highlighted many benefits during the course of interview. Hence pointed out that, it is very difficult to categorize any particular factors which can make ET more effective. In some scenarios more experience related to product is beneficent and others tester’s skill can be useful in ET. However he highlighted that ET can provide more benefits if domain experts are used. While discussing about the defect detection efficiency of ET he added in this regard that defect detection is higher in ET as compared to other test approaches as more time is spent on actual testing rather than any test designs etc. Furthermore he also added that bugs in a system can be investigated in a much better way while performing ET and in addition ET is a cost effective approach as it produces results in less amount of time and spends more time on testing rather on any other activities. In his opinion, free exploration is good approach only for getting the domain knowledge of a system or to look for bugs in a very strict time constraint scenarios, where test cases or missions cannot be utilized. However using free exploration for the sole purpose of testing is not suitable and he told us that more structure approach will be preferred. Interviewee told that better regression testing can also be achieved by ET by but it is still correlated with the tester who discovered that bug. Furthermore he added, in order to get good regression testing developers can also be involved in the process of bug fixing and communicating with developers, test managers to get better know how of impact analysis. If some other tester is doing regression testing of a bug which was discovered by some other it can be a bit tricky.
and difficult to do so. Some more benefits were highlighted by the interviewee as ET has better defect detection and it also reduces the effort and time of documenting each and every test thus it provides motivation to testers by allowing them to perform testing freely and invent fully. While discussing with him about the customer satisfaction and ET he told us that as ET is finding more defects thus indirectly it is providing satisfaction to customer, however since in ET not much is documented hence details of ET can cause negative impact on customer satisfaction. He listed some of the benefits as follows:

- Cost effective
- More actual testing
- Better defect detection
- Tester creativity
- Motivating factor for a tester
- Better utilization of tester skills.

In the second quarter of the interview problems of the TCBT were discussed with the interviewee. He highlighted many issues of TCBT like very time consuming and costly test approach and furthermore managing and updating the test cases is also a hectic and exhaustive procedure. He supported his opinions by mentioning about the problems he cope at UIQ while managing, updating, and finding the links between the tests cases which were more than 10,000 in number. He listed these main issues of the TCBT:

- Time consuming
- Exhaustive
- Difficult to manage
- Less defect detection
- No or less utilization of testers skills

In the third quarter of the interview we discussed about the perceived benefits of TCBT as mentioned in the literature. In this regard he told in the interview that to some extent test cases can provide guidance to the tester on issues like oracle and can make up his mind for further creative testing. Furthermore he added that test cases can provide better requirement/functional coverage and it can also be used to secure the functionalities which later can be presented to the customer in order to get his satisfaction. He emphasized on the fact that quality of the testing can be improved if TCBT is used with other approaches as well because relay on only one approach can hinder testing. He rounded up by saying that not many benefits can be achieved as actual quality of test cases is not known. It is because there is a possibility that a test designer may have interpreted the requirements in a different manner and in real that requirement was different. Discovering bugs/faults by following strictly steps of test cases may not reveal any more defects.

In the last quarter of the interview we discussed about the problems associated with the use of ET. In this regard we discussed about the oracle issues, his response was that as ET is correlated to the testing skills and domain knowledge that a tester possess, hence a good tester who has the domain knowledge can cope with oracle issues but it can be quite difficult if a tester lacks the domain knowledge. While discussing about the test coverage in ET he responded that in free exploration chances of missing any test functionality are high however while following a more structured approach towards ET better coverage can be achieved. Further he added in free exploration managing and prioritizing tests is also an issue but it also can be resolved to an extent if SBTM or some other structured approach is preferred. According to him more or less ET can be used in every scenario of testing i.e. unit test, system test. While discussing about the quality of testing in ET he pointed out that as more critical bugs can be detected by the use of ET hence quality in terms of defect detections is
sufficient because a tester spends more time on actual testing. Some of the problems of ET listed by the interviewee are as follows:

- Less coverage.
- Not auditable in case of free exploration.
- Less customer satisfaction as very less documentation.
- Results of ET testing are difficult to interpret by a customer.

In the last we inquired about the why managers are reluctant in the implementation of ET and his response was that because they think they will lose control over testing and they always require extensive documentation all along in the testing process which is not the case in ET. Traditional managers prefer to use classical test approaches.

8.2.3 Interview 3:

8.2.3.1 Interviewee

Name: Michael Bolton  
Contact information: www.developsense.com  
Email address: michael@developsense.com  
Phone: +1 (416) 992-8378  
Date: March 21, 2010  
Start and End time: 11am to 12:30 pm  
Location: Trondheim, Norway

The contacted person for interview was Mr. Michael Bolton who is currently the owner of DevelopSense and provides consulting, training, coaching, and other services in software testing for companies, teams, and individuals world-wide. In particular, he teaches Rapid Software Testing and Rapid Software Testing for Managers, courses with James Bach.

From 2006 through 2009, he was Program Chair of TASSQ, The Toronto Association of System and Software Quality. In 2008, he was the Conference Chair for the Conference for the Association for Software Testing which is the shining achievements of him.

8.2.3.1.2 Transcribed interview

In the interview we attempted to follow a particular pattern but interviewee was not comfortable with that and he asked us to have a discussion rather than following any list of questions. Hence according to him the benefits and problems of both the test approaches were discussed simultaneously according to him, ET should be a habit practice for improving things, a mind set and skill set for free explorations. TCBT is unable to make decisions; we never know what is right. The test should be chosen based on the risky interesting useful, fun (control or not). There are A and B for the testing i.e. A follows that ET is a free explorations and B follows that scripted items are sub set of the explorations. In ET you can’t use automation. Write down some list of things and do some medium and some process. Exploratory testing would be a loop in which there are no of decisions for executions. ET is also termed as opportunity testing. Testing entirely handle with scripts is like ice and water is mix of both ET and TCBT while steam is a free form ET. ET is much more engaging activity, a cost efficient and a careful activity. ET has more focused documentation. It usually requires documentation after the execution like logs, test notes and videos. It not produces the useless documentation like some sort or predicting documents.
ET emphasizes on uncertainty level heuristics. ET is sort of conclusion statement. Customers are satisfied from the suggestion at the end of test life cycle. Quantity not helps but the quality. Customer’s matters should be decided based on understanding and discovery. It can be done by consistent concentration between testers and client. While discussing the coverage issue in ET he highlighted that every where the basic norm of logic is can’t be done. Real basic coverage is not done in real meaning and considering that TCBT provides full or better coverage is an absurd argument. He then presented us with an example of just a simple program of hitting only two buttons and producing the same results and asked us, can you document every test step in order to achieve maximum coverage. The oracle issue in ET was discussed with him and his response in this regard was that this has different perception from literature and practice. Result of function is everything, tool is limiting the output. Everything is apparent to you based on it you have to decide it yourself. Output is different from the outcome and a tester should be able to understand the clear difference between the both. Output can be considered as $2+2=4$ but if it takes 2 hours to calculate its outcome, so many other aspects are considered while discussing about the oracle. The test cases not often find the problems and are completely uncomfortable to observe the actuality. If you do not have skills or domain experience then you need oracles or outputs. Tester interacts with product and learns and interacting is guidance.

8.2.4 Interview 4:

8.2.4.1.1 Interviewee

Name: Andelkovic Alexander
Contact information: www.maquet.com
Email address: Alexande.Andelkovic@maquet.com
Date: March 29, 2010
Start and End time: 4 pm to 5 pm
Location: Solna, Sweden

The contacted person for interview was Andelkovic Alexander who currently is working for Maquet Critical Care AB as a Test engineer for last 6 years. He has been working in many reputable organizations like Microsoft Mobile Internet Business Unit for more than 4 years and Sendit AB for more than 2 years. He was awarded by the best speaker award in EURO STAR conference 2009 of software testing.

8.2.4.1.2 Transcribed interview

In the first quarter of the interview, the questions related to the benefits of ET were posed to the interviewee and in this regards he highlighted that ET can be very beneficial as more time is spent on testing and if test resources are utilized in a better way. Further while discussing about the defect detection efficiency of ET, his response was that it’s a more of a complementary test approach and it can enhance the defect detection efficiency if it is used with other test approaches. He also pointed out that if time of testing is limited and a product is need to be delivered, ET can be very beneficial in terms of investigating and isolating particular risk areas of the product. In addition, he also mentioned that ET can be considered as a cost efficient approach but it should be used with other test approaches. While discussing about the advantages of free explorations he said that it is good to have free exploration during unit testing and in stabilization period to get a quick overview of product quality. He added that performing ET while doing regression testing can uncover many other bugs as testing around a fix usually finds other problems. Further in the interview we asked that whether using of ET in order to look for bugs in the risk based areas of the system can
be more beneficial, his reply was mixed and he said that only using ET is not sufficient however it can provide you with the overview. He also discussed about the reasons of implementing ET in his organization which were limited time to do testing and a need of having a cost effective complementary test approach/method. Later he highlighted that after implementation of ET more time was actually being spent on testing which resulted in identification of more bugs and as an end result increasing customer satisfaction. Interviewee listed down some of the benefits of ET:

- Increased quality
- Better test coverage
- Increased test participation

In the second quarter of the interview, problems related to the TCBT were discussed with the interviewee and different questions were posed to him in this regard. He mentioned that one cannot specifically say that TCBT is an expensive or time consuming activity because it all depends on what level of quality is desired and quality is not a cheap but rather an expensive activity. While discussing about the revision of existing test cases he told in the interview that it is very important to write clear and unambiguous test instruction to avoid any wastage of testing time on wrong functionality. Hence having a proper and a well planed test design is very important to get clear test instruction in TCBT. He also added that sometimes repetition of the test cases become really exhaustive process and in order to avoid this situation those test cases which are often repeated should be automated or rewritten. Furthermore we asked whether or not regression testing can be better achieved through TCBT, and his response in this perspective was that if a test regression suite is developed from TCBT pool it can provide better and effective regression testing. When we discussed about the life of test cases in a system lifecycle he said that test cases once written are never sufficient for entire lifecycle and are needed to be updated according to the requirement changes, which can ensure test durability. He also pointed out that redesigned test cases are often more sophisticated that the old ones because due to the fact that it’s possible to remove test steps that did not work or in other words emphasize on the parts that are more productive or important in a test cases. Interviewee told that it is very important to keep the test cases updated all the time otherwise they tend to become useless in terms of reusability.

In the third quarter of the interview we discussed about the perceived benefits of the TCBT approach. In this regard interviewee told us that TCBT is good approach but it should be complimented with the ET as well. He also highlighted that test cases can only be beneficial in terms of formulizing or guiding testing tasks if they are clear and are written well. Another benefit of using TCBT was pointed out by the interviewee was that bugs can be found before testing actually starts usually in requirements and functional descriptions. When we discussed about the test coverage, test tracking and the overall quality of testing using TCBT, his response was that only using this test approach may not be sufficient enough in addressing any one of them. He further added that in order to meet regulatory and legal requirements it is mandatory to use TCBT and it is not possible without it. He listed some of the benefits which he experienced as follows:

- Better for user acceptance tests
- Increased product quality
- Increased requirement coverage

In the last quarter of the interview we asked questions related to the problems of ET. In this regard he responded to our questions. He mentioned that oracle issues in ET are often in projects where time is running out and functional specifications have not been updated. Then it can become confusing to an extent on how certain functionality should be implemented and what its working will be according to requirement specification. While discussing about
the coverage and tracking issues in ET, he mentioned that without an ET management tool it becomes very difficult to get an overview of test coverage and tracking. Hence for that reason they have developed a tool SBT to keep the track of ET performed task and the only parameter possible they track across is time. Furthermore he does not consider ET as a separate test approach but rather as a complementary test method and it should be used in many test scenarios as possible. He pointed out that available time of testing is the major factor in prioritizing ET related tasks and usually risk related test scenarios have higher test priority. In his opinion it is very important to know what has been tested and with ET it’s not much difficult to track all that has been done. When we asked about the quality of testing keeping ET in context, his response was that with the passage of ET test time bug ratio should go down at the end of the test period which can predict quality of testing to an extent. In addition he also mentioned that ET alone is not sufficient in deterring the overall quality of testing as it is a complementary approach and usually it is a combination of ET and other test approaches(requirement testing, environment, stress, UI and so on) which depicts the overall quality of testing. According to his experience performing regression testing in ET and determining if old bug has been fixed or not can sometimes be confusing if ET repro scenario is not available to examine and understand how earlier bug was found. While discussing about the test reviews in ET, he told that it’s not difficult to perform any such reviews in ET as long as functional specification is available. He further added that because different testers may test same ET scenarios so rarely test repeatability problems occur. The main issue that he pointed out was the acceptance of ET in any company or organization and it is very important to communicate the benefits of ET in order to get acceptance in any organization, otherwise perception of ET may have many unrealistic drawbacks such as control over test coverage and time problems etc

8.2.5 Interview 5:

8.2.5.1.1 Interviewee

Name: James Bach
Contact information: www.satisfice.com
Email address: james@satisfice.com
Date: April 2, 2010
Start and End time: 3 pm to 6 pm
Location: Lund, Sweden

The contacted person for interview was James Bach who is a famous scholar of exploratory testing and usually known as “Bad boy of Software Testing.” He has a rich software testing experience. Now a days he is helping individuals and organizations to plan SQA, change control, and testing processes that allow them to understand and control the risks of product failure. Most of his experience is with market-driven Silicon Valley software companies like Apple Computer and Borland, so he claimed that the techniques he has gathered and developed are designed for use under conditions of compressed schedules, high rates of change, component-based technology, and poor specification.

In the interview we attempted to follow a particular pattern, but interviewee insisted in having a detailed discussion about the relevant topic. Hence this interview lasted for 3 hours in which interviewee highlighted about several dimensions of ET and TCBT. He mentioned that it is very difficult to apart ET and TCBT from each other because when a test designer is in the process of designing test cases he uses his exploratory nature to document the test steps. He further added that even while performing TCBT, it is very difficult for a tester to follow each and every step of the test case. Hence chances of finding defects by following those explicit test steps are very less and in addition it is also very time consuming to document each step. Further he added that it is very likely that a test designer didn’t have much experience in writing test nor had the domain knowledge and in contrast he wrote bundles of immature test cases. He concluded in this regard that skills of a tester plays a very vital role in any form of testing and if a tester was only to follow a pre-scripted set of test steps then monkeys could also have done it in better way.

While discussing about the test coverage, he added that there are many types of test coverage and documenting each and every single one of them in order to satisfy the argument of TCBT providing better coverage is absurd because it will just take always to only document test coverage and actually performing it “God knows how much time”. He also added that it’s a great misconception that ET and TCBT are considered as two different test techniques, but in real they are two approaches and they tend to work better if both are combined and have a balance in them. He compared TCBT with a straight line which has a start and an end; on the other hand he depicted ET as a circle having reviews and lots of discussion which acts as an input to ET. He pointed out that a straight line i.e. TCBT can have multiple small circles i.e. ET and a circle i.e. ET can have multiple straight lines i.e. TCBT. He also mentioned that while meeting some legal and regulatory requirements test cases can be utilized in an effective manner. He stressed on the fact that ET is not a technique rather an approach which is strictly correlated with the cognitive skills of the tester.

He mentioned that ET is more dependent on the behavior of a tester towards problems solving. He also mentioned that in ET a tester learn about an application and creates multiple scenarios to test and also write it, thus creating an effective test case at the same time. He added that there are two mind sets of a tester while performing ET one is to de focus and other is to have super focus. An ET tester according to the interviewee can de focus his thinking process if he is not able to find bugs or problems on some other aspects of the system in order to get a fresh and better start or in other words to create a super focus. He pointed out that when ET starts it may look like unstructured but as the testing further advances ET gets even more structured than any other tester approach but then again it relays greatly on the abilities of the tester. He added that training ET testers is an important task because less skilled tester can hamper the true essence of ET.

While discussing about the defect detection efficiency of ET, his response in this regard was that in ET defect detection is primarily associated with the abilities of a tester and in TCBT it’s dependent on the test case design. For this reason it’s quite difficult to distinguish between both of them as which has better defect detection unless and until the ingredients of the both approaches are not studied, however he also told that as more time of actual testing is done in ET hence it can be considered as having more defect detection in terms of testing time.
8.3 Contact Details of Static evaluators

<table>
<thead>
<tr>
<th>Name</th>
<th>Designation</th>
<th>Company</th>
<th>Contact Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annika Mansson</td>
<td>Test Manager</td>
<td><a href="http://www.toolaware.com">www.toolaware.com</a></td>
<td><a href="mailto:info@toolaware.com">info@toolaware.com</a></td>
</tr>
<tr>
<td>Herman Afzelius</td>
<td>Project Manager</td>
<td><a href="http://www.logica.se">www.logica.se</a></td>
<td><a href="mailto:herman.afzelius@logica.com">herman.afzelius@logica.com</a></td>
</tr>
<tr>
<td>Petter Mattsson</td>
<td>Test Manager</td>
<td><a href="http://www.logica.se">www.logica.se</a></td>
<td><a href="mailto:petter.mattsson@logica.com">petter.mattsson@logica.com</a></td>
</tr>
<tr>
<td>Andelkovic Alexander</td>
<td>Test Manager</td>
<td><a href="http://www.maquet.com">www.maquet.com</a></td>
<td><a href="mailto:Alexande.Andelkovic@maquet.com">Alexande.Andelkovic@maquet.com</a></td>
</tr>
</tbody>
</table>

Table 24: Detail of static evaluators

8.4 Experiment Instruments

8.4.1 Background form

<table>
<thead>
<tr>
<th>Background Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tester Name:</td>
</tr>
<tr>
<td>Department:</td>
</tr>
<tr>
<td>Have you gained the knowledge of software testing?</td>
</tr>
<tr>
<td>How would you assess your knowledge of Software Testing?</td>
</tr>
<tr>
<td>How long have you been involved in Software Testing?</td>
</tr>
<tr>
<td>Do you know about exploratory Testing (ET)?</td>
</tr>
</tbody>
</table>

Table 25: Background form
8.4.2 Bug Report template

**Bug report:**

<table>
<thead>
<tr>
<th>Functionality</th>
<th>Bug Description</th>
<th>Severity</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 26: Bug report template

8.4.3 Test Case template

**Tester Name:**

<table>
<thead>
<tr>
<th>Module</th>
<th>Test case description</th>
<th>Test Steps</th>
<th>Test Status P/F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Step</td>
<td>Expected</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 27: Test case Template

8.4.4 Requirement based Test Cases for Calculator

**Tester Name:**

<table>
<thead>
<tr>
<th>RBTC ID</th>
<th>RBTC description</th>
<th>Test steps</th>
<th>Expected outcome</th>
<th>Actual Outcome</th>
<th>Test Status</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 28: RBTC Template

8.4.5 Mission Sheet for Mix testing

**Tester Name:**

<table>
<thead>
<tr>
<th>Mission No</th>
<th>Mission Description</th>
<th>Notes if any</th>
</tr>
</thead>
</table>

Table 29: Mission Sheet
## 8.5 Measurement Tables

### 8.5.1 F-max table

<table>
<thead>
<tr>
<th>Degrees of Freedom-1</th>
<th>Number of treatments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>39.0</td>
</tr>
<tr>
<td>3</td>
<td>15.4</td>
</tr>
<tr>
<td>4</td>
<td>9.6</td>
</tr>
<tr>
<td>5</td>
<td>7.2</td>
</tr>
<tr>
<td>6</td>
<td>5.82</td>
</tr>
<tr>
<td>7</td>
<td>99</td>
</tr>
<tr>
<td>8</td>
<td>4.43</td>
</tr>
<tr>
<td>9</td>
<td>4.03</td>
</tr>
<tr>
<td>10</td>
<td>3.72</td>
</tr>
<tr>
<td>12</td>
<td>3.28</td>
</tr>
<tr>
<td>15</td>
<td>2.86</td>
</tr>
<tr>
<td>20</td>
<td>2.46</td>
</tr>
<tr>
<td>30</td>
<td>2.07</td>
</tr>
<tr>
<td>60</td>
<td>1.67</td>
</tr>
<tr>
<td>?</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 30: F max table

### 8.5.2 F table at (p = 0.05)

<table>
<thead>
<tr>
<th>D.of Freedom, v (for replicates)</th>
<th>Degrees of Freedom, u (for treatments)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>161 200 216 225 230 234 239 244 254</td>
</tr>
<tr>
<td>2</td>
<td>18.5 19.0 19.2 19.3 19.3 19.3 19.4 19.4 19.5</td>
</tr>
<tr>
<td>3</td>
<td>10.1 9.6 9.3 9.1 9.0 8.9 8.8 8.7 8.5</td>
</tr>
<tr>
<td>4</td>
<td>7.7 6.9 6.6 6.4 6.3 6.2 6.0 5.9 5.6</td>
</tr>
<tr>
<td>5</td>
<td>6.6 5.8 5.4 5.2 5.1 5.0 4.8 4.7 4.4</td>
</tr>
<tr>
<td></td>
<td>6.0</td>
</tr>
<tr>
<td>----</td>
<td>-----</td>
</tr>
<tr>
<td>7</td>
<td>5.6</td>
</tr>
<tr>
<td>8</td>
<td>5.3</td>
</tr>
<tr>
<td>9</td>
<td>5.1</td>
</tr>
<tr>
<td>10</td>
<td>5.0</td>
</tr>
<tr>
<td>11</td>
<td>4.8</td>
</tr>
<tr>
<td>12</td>
<td>4.8</td>
</tr>
<tr>
<td>13</td>
<td>4.7</td>
</tr>
<tr>
<td>14</td>
<td>4.6</td>
</tr>
<tr>
<td>15</td>
<td>4.5</td>
</tr>
<tr>
<td>16</td>
<td>4.5</td>
</tr>
<tr>
<td>17</td>
<td>4.5</td>
</tr>
<tr>
<td>18</td>
<td>4.4</td>
</tr>
<tr>
<td>19</td>
<td>4.4</td>
</tr>
<tr>
<td>20</td>
<td>4.4</td>
</tr>
<tr>
<td>21</td>
<td>4.3</td>
</tr>
<tr>
<td>22</td>
<td>4.3</td>
</tr>
<tr>
<td>23</td>
<td>4.3</td>
</tr>
<tr>
<td>24</td>
<td>4.3</td>
</tr>
<tr>
<td>25</td>
<td>4.2</td>
</tr>
<tr>
<td>26</td>
<td>4.2</td>
</tr>
<tr>
<td>27</td>
<td>4.2</td>
</tr>
<tr>
<td>28</td>
<td>4.2</td>
</tr>
<tr>
<td>29</td>
<td>4.2</td>
</tr>
<tr>
<td>30</td>
<td>4.2</td>
</tr>
<tr>
<td>40</td>
<td>4.1</td>
</tr>
<tr>
<td>60</td>
<td>4.0</td>
</tr>
<tr>
<td>120</td>
<td>3.9</td>
</tr>
<tr>
<td>}</td>
<td>3.8</td>
</tr>
</tbody>
</table>

Table 31: F table at (p = 0.05)
8.5.3 The tabulated value of $t$

<table>
<thead>
<tr>
<th>Degrees of Freedom</th>
<th>Probability, $p$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.1</td>
</tr>
<tr>
<td>1</td>
<td>6.31</td>
</tr>
<tr>
<td>2</td>
<td>2.92</td>
</tr>
<tr>
<td>3</td>
<td>2.35</td>
</tr>
<tr>
<td>4</td>
<td>2.13</td>
</tr>
<tr>
<td>5</td>
<td>2.02</td>
</tr>
<tr>
<td>6</td>
<td>1.94</td>
</tr>
<tr>
<td>7</td>
<td>1.89</td>
</tr>
<tr>
<td>8</td>
<td>1.86</td>
</tr>
<tr>
<td>9</td>
<td>1.83</td>
</tr>
<tr>
<td>10</td>
<td>1.81</td>
</tr>
<tr>
<td>11</td>
<td>1.80</td>
</tr>
<tr>
<td>12</td>
<td>1.78</td>
</tr>
<tr>
<td>13</td>
<td>1.77</td>
</tr>
<tr>
<td>14</td>
<td>1.76</td>
</tr>
<tr>
<td>15</td>
<td>1.75</td>
</tr>
<tr>
<td>16</td>
<td>1.75</td>
</tr>
<tr>
<td>17</td>
<td>1.74</td>
</tr>
<tr>
<td>18</td>
<td>1.73</td>
</tr>
<tr>
<td>19</td>
<td>1.73</td>
</tr>
<tr>
<td>20</td>
<td>1.72</td>
</tr>
<tr>
<td>21</td>
<td>1.72</td>
</tr>
<tr>
<td>22</td>
<td>1.72</td>
</tr>
<tr>
<td>23</td>
<td>1.71</td>
</tr>
<tr>
<td>24</td>
<td>1.71</td>
</tr>
<tr>
<td>25</td>
<td>1.71</td>
</tr>
<tr>
<td>26</td>
<td>1.71</td>
</tr>
<tr>
<td>27</td>
<td>1.70</td>
</tr>
<tr>
<td>28</td>
<td>1.70</td>
</tr>
<tr>
<td>29</td>
<td>1.70</td>
</tr>
<tr>
<td>30</td>
<td>1.70</td>
</tr>
<tr>
<td>40</td>
<td>1.68</td>
</tr>
<tr>
<td>60</td>
<td>1.67</td>
</tr>
<tr>
<td></td>
<td>1.66</td>
</tr>
<tr>
<td>-----</td>
<td>------</td>
</tr>
<tr>
<td>120</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.65</td>
</tr>
</tbody>
</table>

Table 32: Calculated value of t