

Received March 15, 2018, accepted May 6, 2018. Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2018.2834957

# Levels of Exploration in Exploratory Testing: From Freestyle to Fully Scripted

AHMAD NAUMAN GHAZI<sup>1</sup>, KAI PETERSEN<sup>1</sup>, ELIZABETH BJARNASON<sup>2</sup>,  
AND PER RUNESON<sup>2</sup>, (Member, IEEE)

<sup>1</sup>Department of Software Engineering, Blekinge Institute of Technology, 371 79 Karlskrona, Sweden

<sup>2</sup>Department of Computer Science, Lund University, 221 00 Lund, Sweden

Corresponding author: Ahmad Nauman Ghazi (nauman.ghazi@bth.se)

This work was partly funded by the EASE Industrial Excellence Center for Embedded Applications Software Engineering, (<http://ease.cs.lth.se>).

**ABSTRACT** Exploratory testing (ET) is a powerful and efficient way of testing software by integrating design, execution, and analysis of tests during a testing session. ET is often contrasted with scripted testing and seen as a choice of either exploratory testing or not. In contrast, we pose that exploratory testing can be of varying degrees of exploration from fully exploratory to fully scripted. In line with this, we propose a scale for the degree of exploration and define five levels. In our classification, these levels of exploration correspond to the way test charters are defined. We have evaluated this classification through focus groups at four companies and identified factors that influence the choice of exploration level. The results show that the proposed levels of exploration are influenced by different factors such as ease to reproduce defects, better learning, and verification of requirements and that the levels can be used as a guide to structure test charters. Our study also indicates that applying a combination of exploration levels can be beneficial in achieving effective testing.

**INDEX TERMS** Exploratory testing, test charter, test mission, session-based test management, levels of exploration, exploratory testing classification, software testing.

## I. INTRODUCTION

Advocates of exploratory testing (ET) stress the benefits of providing the tester with freedom to act based on his/her skills, paired with the reduced effort for test script design and maintenance. ET can be very effective in detecting critical defects [1]. We have found that exploratory testing can be more effective in practice than traditional software testing approaches, such as scripted testing [1], [2]. ET supports testers in learning about the system while testing [1], [3]. The ET approach also enables a tester to explore areas of the software that were overlooked while designing test cases based on system requirements [4]. However, ET does come with some shortcomings and challenges. In particular, ET can be performed in many different ways, and thus there is no one-way of training someone to be an exploratory tester. Also, exploratory testing tends to be considered an ad-hoc way of testing and some argue that defects detected using ET are difficult to reproduce [5].

The benefits of exploratory testing are discussed both within industry and academia, but only little work relates to how to perform this kind of testing [6]. Bach introduced a technique named Session Based Test Management (SBTM)

[7] that provides a basic structure and guidelines for ET using test missions. In the context of SBTM, a test mission is an objective to provide focus on what to test or what problems to identify within a test session [7]. SBTM provides a strong focus on designing test charters to scope exploration to the test missions assigned to exploratory testers. A test charter provides a clear goal and scopes the test session, and can be seen as a high level test plan [8] thus the level of detail provided in the test charter influences the degree of exploration in the testing.

However, little guidance exists on how to define test charters in order to achieve different, or combine various degrees of exploration. Though, there is a need in the industry to have support for choosing the “right” degree of exploration see e.g. [9]. In order to make an informed decision there is a need to define what is meant by “degree of exploration.”

We pose that testing can be performed at varying degrees of exploration from freestyle ET to fully scripted, and propose a scale for the degree of exploration defined by five distinct levels of exploration. In this paper, we present a classification consisting of five levels of exploratory testing (ET) ranging from *free style testing* to *fully scripted*

testing. We exemplify these five levels of exploration with test charter types that were defined based on studying existing test charters in industry. In a previous research study [8], we provided a checklist of contents to support test charter design in exploratory testing. The focus of that research was to support practitioners in designing test charters depending on the context of the test mission and the system under test. We have extended this research to explore different levels of exploration and how they map to the contents of the test charters, e.g. test goals, test steps, etc.

We evaluated our classification through focus groups at four companies, Sony Mobile Communications, Axis Communications, Ericsson, and Softhouse Consulting. In addition to validating the levels of exploration, these focus groups provided insight into factors that influence the choice of one or more exploration levels to be used in exploratory testing.

The remainder of this paper is structured as follows. Section II presents related work on exploratory testing and test charter design. Section III presents the research methodology and case descriptions. The results are presented in Section IV, including the classification (Section IV-A) and evaluation (Section IV-B). Section V provides the conclusions from the research and directions for the future work.

## II. RELATED WORK

Exploratory testing (ET) is a way to harness the skills, knowledge and creativity of a software tester. The tester explores the system while testing, and uses the knowledge gained to decide how to continue the exploration. The design, execution, and analysis of tests take place in an integrated fashion [10]. The experience and skills of the tester plays a vital role in ET and influences the outcome of the testing [11], [12]. ET displays multiple benefits, such as testing efficiency and effectiveness [1], [6], a goal-focused approach to testing, ease of use, flexibility in test design and execution, and providing an interesting and engaging task for the tester [6].

Shah *et al.* [5] conducted a systematic review of the literature on ET, and found that the strengths of ET are often the weaknesses of scripted testing and vice versa. They conclude that ET and scripted testing should be used together to address the weaknesses of one process with the strength of the other. While Shah *et al.* [5] do not consider different types of ET, other research highlights the existence of different levels or levels of exploration [4]. Bach [13] shows that there exists a continuum of exploration between fully scripted and freestyle ET but does not classify this continuum of exploration with distinct levels.

We complement the existing work by defining different levels of exploration and their advantages and disadvantages, thus aiding practitioners during their decisions of what levels of exploration to choose.

## III. METHOD

The following research questions are formulated for our study.

- *RQ1: How can different degrees of exploration be characterized?* In previous studies the question was raised whether or not to do exploratory testing and how to distribute effort between exploratory and scripted testing [9]. Practitioners can use the characterization of different degrees of exploration as decision options that go beyond the distinction of scripted versus exploratory testing when defining their testing strategies. That is, the question is not whether to conduct exploratory testing or not, but rather which degree of exploration is preferred.
- *RQ2: What are the factors that influence the levels of exploration?* Knowledge of such factors, e.g. with regard to defect detection ability may support practitioners when deciding at which level to perform exploratory testing.

### A. DESIGNING THE CLASSIFICATION (RQ1)

In our earlier work we identified 35 potential information items that may be included in test charters [8]. Examples of these include descriptions of the test setup, test techniques to be used, purpose of the testing session, priorities, quality characteristics to focus on, etc. Potentially one could include all information items in a test charter, however, we argue that this would be counterproductive for various reasons, such as:

- Not all items in the test charter may be of equal importance.
- Including all the items would overload the test session, as there is too much to check, resulting in confusion for the tester. This would be counterproductive from the idea of exploratory testing, where the testers are driven by short iterations of learning and reacting during a test session.

Test charters can be used to steer the degree of exploration where the ET level changes as information items are included/excluded in the test charters. When no extensive information (except the test object or system under test) is provided to the tester, the tester is free to fully explore. The more information that is added to the test charter, the more the tester is restricted. Testers are restricted by test steps or biased by information provided in the test charter. As a consequence we used test charters as a means to define the levels of exploration.

The checklists to support test charter design, defined by Ghazi *et al.* [8] in combination with existing test charters formed the basis for designing the classification of test levels. A total of 15 test charter examples obtained through a literature search (peer-reviewed and grey) were obtained. We ranked the checklists from low exploration to high exploration.

Test charters with very high degrees of exploration only stated the test objects and the main aim/mission of the test session [14]. An example test charter from Suranto [14] states what should be explored i.e. the test object (a “*histogram page*” in Suranto’s example), and provides ideas for what input data to use (“*various data sets and*

**TABLE 1.** Overview of focus groups.

Focus group	Company	Context	Number of participants	Experience (in years)
1	Sony Mobile Communications	Telecommunication systems, mobile applications	6	4 - 25 years
2	Axis Communications	Networked security cameras	4	4 - 25 years
3	Ericsson	Telecommunication systems, charging systems	7	15 - 24 years
4	Softhouse Consulting	Consulting (Banking solutions, telecommunication systems, mobile applications, embedded systems, control systems.)	3	15 - 24 years

different bin-interval setting”) with the goal to discover bugs in “the histogram display.” The charter in the paper by Suranto characterized well as what we later classify as “High degree of exploration.”

An example of a test charter with a low degree of exploration is presented by Claesson [15] showing an example of a Copy/Paste function. The charter includes information about:

- *Actors*: Role the tester should take
- *Purpose*: Goal of the testing session
- *Setup*: Specification of the technical environment
- *Priority*: Importance of the function
- *Reference*: Links to complementary documents (here requirements)
- *Data*: Specification of the types of input data that can be used
- *Activities*: Concrete steps to be taken and their order when conducting the testing

As is evident the test charter leaves little room to explore as all the test steps and types of input data are given, as well as data types when conducting copy and paste. That is, only the concrete test data (e.g. which concrete image to copy) is left to the testers to explore. By providing all this information no room is left for exploration, i.e. this equates scripted testing

Other test charters fell in-between the two examples [14], [15] and thus provided a medium degree of exploration.

## B. EVALUATION

We evaluated our classification of ET levels through a study of test processes at four companies in Sweden involved in large-scale product development in the area of telecommunications and embedded systems. The companies we studied are Sony Mobile Communications, Axis Communications, Ericsson, and Softhouse Consulting. Focus groups were used as the main data collection method for our evaluation.

### 1) COMPANIES

All of these companies use agile software development as their development methodology. However, two of these companies, Sony Mobile Communications and Ericsson, have a strong focus on developing telecommunication systems ranging from mobile applications to telecommunications charging systems. Axis Communications, mainly works in the area of networked security cameras and embedded software. Softhouse Consulting provides consultation to a wide range of companies working on banking solution, telecom-

munication systems, mobile applications, embedded software and control systems.

### 2) SUBJECTS

Exploratory focus groups were conducted at Sony Mobile Communications and Axis Communications, to elicit the influence factors that affect the level of exploration in testing. The participants were selected by the companies considering the research needs of the study presented herein.

To validate the factors, that influence levels of exploration, two focus group at Ericsson and Softhouse Consulting were performed. Overall, 20 practitioners participated in these focus groups. These participants were experienced testers with 4 to 25 years of experience in software testing. Table 1 provides an overview of the participants during each focus group, context of their assignments in the company and their experience in software testing.

### 3) DATA COLLECTION

We used focus groups [16] as the main method for data collection at all four companies and conducted them in two main iterations. In a focus group, a group of experts is selected by the researchers to discuss and collect their views in a specific area of expertise where these practitioners have considerable experience. Focus groups, as a data collection method, support the researchers to understand a research area in a concise way with a strong involvement of experts from industry [16].

In the study presented herein, the initial two focus groups were exploratory and were conducted at two companies, Sony Mobile Communications and Axis Communications. These companies were interested in extending their use of exploratory testing in their current test processes. The third and fourth focus groups were held at Ericsson and at Softhouse Consulting with the main aim to validate the results from the initial two focus groups. The focus group at Ericsson was performed in two 4 hour sessions on different days.

The first two focus groups contained the following steps:

- 1) Introduce the basic concepts of exploratory testing
- 2) Present our classification of exploration levels
- 3) Share examples of test charters type for each level with the participants
- 4) The participants re-write an existing test case at the different exploration levels using the provided test charter types

- 5) Open discussion of how each level and test charter types matches the context for their current test practices
- 6) Elicit factors that are affected by the level of exploration in testing

Prior to the third and fourth focus group, we conducted a survey with the focus group participants to gauge their views of the factors elicited from the first two focus groups (such as learning) and to which extent these impact the level of exploration. At the (subsequent) focus group sessions, we discussed the outcome of the survey and in particular, how various factors are influenced by the level of exploration and how this affects the decision regarding which level of exploration to apply in a test session, and reached a consensus for each factor.

The companies selected the focus group participants based on their experience of testing and their interest in exploratory testing. We audio recorded and transcribed all focus group sessions, and analyzed these to identify resulting impact of factors on each level of exploration in the proposed classification. Table 1 provides an overview of companies and the participants in the focus groups.

#### 4) THREATS TO VALIDITY

The focus group participants did not have direct experience of all the levels of exploration and the corresponding test charter types discussed in the focus group. However, we believe that the participants could relate to these anyway given their experience of testing. We reduced this threat of lack of first hand experience by letting the practitioners gain hands-on experience of the test charter types during the focus group.

A common threat of studies with companies is the generalizability of the findings. We partially reduce this threat, by involving four companies. Furthermore, we mitigated the risk of researcher bias by involving three different researchers in designing and performing the focus groups, and in jointly discussing the outcome of these.

## IV. RESULTS

### A. LEVELS OF EXPLORATION IN EXPLORATORY TESTING (RQ1)

We have identified five levels of exploration ranging from *free style exploratory testing* to *fully scripted testing* with the intermediate levels of *high*, *medium* and *low level of exploration*. Figure 1 provides an overview of the proposed classification. Each of the five levels are defined by a test charter type that guides the testing. The test charter for each level of exploration adds an element that sets the scope for exploration space for the tester. At the freestyle level, the tester is only provided with the test object. For each subsequent level, the exploration space is reduced by adding further information and instructions to the test charter, e.g. high-level goals. The tester is thus further focused for each decreasing exploration level, and the reduced freedom leads to a less exploratory approach compared to the previous level. The test charter type for the lowest exploration level, i.e. fully scripted,

contains both test steps and test data, and thus leaves no space for exploration during test execution.

We provide examples of test charters produced during one of the focus groups in Figure 2. The test charter for the highest level of exploration, i.e. free style (not shown in figure) contains only the goal for the testing, namely to verify a specific function of the system. The test charter for the medium exploration level contains additional information, e.g. suitable starting points for the testing. Finally, the test charter for the low level of exploration contains detailed test activities/steps in addition to goals and other information. The next level, i.e. fully scripted, which is not shown in Figure 2, also contains test data. For example, the test charter for the test activity “Copy content to card from PC” would also specify the content to be copied.

### B. FACTORS INFLUENCING THE CHOICE OF EXPLORATION LEVELS (RQ2)

We evaluated our classification of exploration levels in ET and explored factors/characteristics that influence the selection of these levels and the corresponding charter types through focus groups at four companies, see Section III. We found six main areas that influence the choice for selecting the level of exploration used in testing, namely *defect detection*, *time and effort*, *people-related factors*, *evolution and change*, *traceability* and *quality requirements*. We provide an overview of these factors in Figure 3 by presenting two opposing poles for each factor. For example, better learning (indicated as positive by 😊) versus poor learning (indicated as negative by ☹️). If an exploration level has a neutral impact on a factor this is indicated by 😐.

Overall, the practitioners had a positive view of the higher exploration levels (freestyle and high exploration) for four of the six main areas. The participants noted a positive impact for these levels within the areas of *defect detection*, *time and effort*, *people-related factors*, and *evolution and change*. In contrast, they expressed a negative impact for factors related to *traceability* and *verifying quality requirements*. The participants believed that the more exploration levels in ET have a negative impact on these two areas.

#### 1) DEFECT DETECTION

The participants of all focus groups highlighted that the exploratory approach will identify more significant defects. However, one participant stated that this may only be the case “if you know what the faults may be,” i.e. the tester should have the skills to identify where significant faults are most likely to occur. Thus, the tester’s skills plays a vital role in exploratory testing, as is also confirmed by empirical studies on ET [12]. These skills are also required to judge whether an explored behavior is a critical defect or not. However, the participants also pointed out that people taking a new perspective often find new defects. One practitioner said: “every time we get new people in the team, we find new defects in the system.” This highlights one of the benefits of exploration, namely that of not biasing the search for



**Freestyle:** Only the test object is provided to the tester. The tester can freely explore the system

**High degree of exploration:** The tester is provided with one or more high goals for the test session, also knowing the test object. Besides that, the tester can freely explore the system.

**Medium degree of exploration:** The tester is provided with one or more high goals for the test session. At the same time additional restrictions are required, that may bias and thus limit the tester in his/her testing session. Biasing aspects could be too detailed goals, priorities, risks that the tester is required to focus on, tools used, the functionality that needs to be covered, or the test method to be used.

**Low degree of exploration:** Besides the information for Medium degree of exploration, the tester is also required to follow certain test steps, which further may bias the tester and reduce the exploration space. The tester is encouraged to choose the test data to be used in the test steps.

**Fully scripted:** The tester is provided with the test steps, but also with the test data, which does not provide room for exploration steps.

**FIGURE 1.** Classification of levels of exploration in exploratory testing.

	High degree of exploration	Medium degree of exploration	Low degree of exploration
Test goal/ purpose	To verify adoptable storage	1. Test different adoptable storage (speed and size) 2. Test adoptable storage and internal/external memory 3. Test move adoptable storage to other device (phone/PC etc) 4. Test extract files from adoptable storage	1. To verify that adopt.storage behaves (as expected) according To the requirement 2. To test that no data loss occurs
Set-up (pre-conditions of what needs to be available to test)			1. PC 2. 2 devices supported by adopt.storage 3. Adoptable storage 1 ... n 4. 1 device not supporting adopt.storage 5. Test content
Priority			High. There must not be any data loss
References			See ALMreg.doc
Test activities			1. Insert adoptable storage 2. Setup as internal 3. Copy content to adoptable storage from PC 4. Read content on adoptable storage from device 5. Save
Additional information used		1. Use previously identified problems as input 2. Use Google requirements if needed	

**FIGURE 2.** Example of test charters for the high, medium and low degrees of exploration.

defects, for example, through pre-existing test cases and prior knowledge embedded in scripted tests. This may be the case for the lower exploration levels, namely low exploration and fully scripted. Some participants pointed out that high exploration comes with challenges with reproducibility of detected defect. One participant said that the “*problem is when you have higher level of exploration.. the developers want to have very detailed steps to reproduce it.*” However, when “*you focus on the reproducing you lose the exploration*” which is a drawback with the fully scripted level.

## 2) TIME AND EFFORT

Many participants highlighted time efficiency as one of the benefits of the high and medium levels of exploration. One practitioner explained this by saying that “*we can get a better overview quickly*” with higher exploration levels. At these higher exploration levels less effort is required to prepare the tests, compared to the lower and fully scripted exploration levels. One participant explained that the many details of the low levels of exploration require “*an upfront investment to develop test cases*” before you can execute

Negative (☹️)	Neutral (😐)					Positive (😊)
	Freestyle	High exploration	Medium Exploration	Low Exploration	Scripted Manual	
Defect Detection						
Finds less critical defects	😊	😊	😊	😐	😊/☹️	Finds more significant/ critical defects
Does not help finding unknown defects to a great extent	😊	😊	😊/☹️	😐	☹️	Helps to uncover unknown defects
Difficult to reproduce defects	☹️	😐	😐	😐	😊	Easier to reproduce defects
Time and effort						
Time inefficient	😊	😊	😐	☹️	☹️	Time efficient
Effort intensive test preparation	😊	😊	😐	☹️	☹️	Less effort to prepare tests
More effort to maintain test cases	😊	😊	😐	😐	☹️	Less effort to maintain test cases
People factors						
Poor learning	😊	😊	😊	😐	😐	Better learning
Hinders critical thinking to challenge expected outcomes	😊	😊	😊	😐	☹️	Motivates critical thinking to challenge expected outcomes
Demotivates the tester	😊	😊	😊	😐	☹️	Motivates the tester
Evolution and change						
Difficult to design new tests	😊	😊	😐	☹️	☹️	Easy to design new tests
Resilient to change test cases	😊	😊	😊	😊/☹️	😊/☹️	Easier/ provides freedom to change test cases
Difficult to fill knowledge gap when adding new requirements	☹️	😐	😊	😐	😊	Easier to fill knowledge gap when adding new requirements
Traceability						
Hard to trace coverage	☹️	☹️	😊	😊	😊	Easy to trace coverage
Inefficient in checking verification of requirements	☹️	☹️	☹️	😊	😊	Efficient in checking verification of requirements
Quality requirements						
Difficult to verify conformance/legal requirements	☹️	☹️	☹️	😊	😊	Easier to verify conformance/legal requirements
Does not help in checking performance issues	😊/☹️	😐	😐	😐	😊	Helps to check performance issues

**FIGURE 3.** Overview of factors influencing the levels of exploration derived from the focus groups and the survey.

them. Another participant expands on this by saying that there is “less administration if you have a high level of exploration because then you have quite openness and it is much easier to write test cases.” The participants indicated that less effort is required to maintain test cases at the higher levels of exploration due to the fact that changes are more likely to affect details within test cases at the lower exploration levels. For example, the tester would need to update the test steps.

### 3) PEOPLE FACTORS

The participants highlighted that the higher levels of exploration are beneficial for encouraging critical thinking, for challenging the system when testing, and that these levels support learning. One participant said that at the highest exploration level (freestyle) learning “might take longer time. But that it would probably be a better approach from the

beginning to understand what the testing is.” This participant also said that “you only do fully scripted when you know the system and it is monotonous and you can also get tired of it.” However, some participants also expressed positive learning effects from fully scripted testing and that “it is definitely easier to start learning about testing when it is fully scripted. If we do freestyle then it would be difficult” because “it requires skills and some form of competence or otherwise you are completely lost.” One participant suggested that to make full use of the higher levels of exploration, i.e. freestyle and high exploration “you need a mentor that tells you explore this, and then you explore and test. When you have questions, you go back and ask/discuss with the mentor.” Several participants pointed out that a tester’s experience plays an important role vis-a-vis the exploration levels. Less experienced testers are often able to identify new defects

since they bring a new perspective to a project. At the same time, a tester with less experience may find it hard to conduct freestyle or high exploration testing since they do not have the domain knowledge required. Hence, there are additional factors that affect the level of learning. Participants also stressed that with scripted testing *“one problem can be that if you just keep following the test steps then there is a chance that you miss the approval criteria.”* Finally, the participants pointed out that learning occurs during the derivation of test case from the detailed requirements, therefore it is important to consult requirements documents, no matter what level of exploration you select to drive your test sessions. They also highlighted motivation as an important distinguishing factor, where testers quickly get bored when testing at low levels of exploration included fully scripted testing. Furthermore, the participants highlighted that the impact and effect of the exploration levels may very well vary throughout the development cycle. They said that the higher levels of exploration might be particularly useful during the early phases of testing to explore and learn about the system. The testers may then design new tests that later become scripted tests, which are used for regression testing in later stages when the project is closer to releasing software.

#### 4) EVOLUTION AND CHANGE

The participants highlighted that it is easier to design new tests for higher levels of exploration (freestyle and high level of exploration) since this requires less effort; *“you have transparency and it is much easier to write test cases.”* In line with this, the participants also expressed that changes can be more easily implemented given that the *“higher exploration levels are less resistant [to change] since you don’t need to change a lot of details.”* They also expressed that the communication around changes to tests is simplified for these higher exploration levels and that when *“some behavior has changed and you just discuss and notify that this has changed instead of going in details every time.”* However, the practitioners also said that the higher exploration levels are more challenging when requirements are added or changed, since information of the new requirements is needed to guide the testing.

#### 5) TRACEABILITY

All focus groups highlighted that the difficulties of tracing requirements coverage is a major drawback for higher levels of exploration both regarding coverage of code and of requirements. One participant said: *“The sense of coverage is much lower as compared to when you ticked off 100 test cases in scripted tests.”* This issue also applies to requirements coverage since test cases at the higher levels of exploration as per definition do not include any mapping to individual requirements.

#### 6) QUALITY REQUIREMENTS

Several participants highlighted that the higher levels of exploration are not suitable for verifying conformance requirements. One participant said that *“We do have a lot*

*of conformance with different standards and legal requirements”* and *“if you don’t have this kind of [low] exploration level then it is easier to miss.”* The participants expressed different views on this regarding performance. When testing the load of a system, scripted automated tests are often preferred. In one case, a participant highlighted that for performance testing *“you have to continuously compare it to different firmware and we need to have similar tests again. Then we can’t really explore a lot.”* However, it is also important to consider perceived quality and the end user perspective, for this the higher levels of exploration are suitable since they allow for making observations during testing.

### V. CONCLUSIONS AND FUTURE WORK

Exploratory testing (ET) leads to different outcomes in comparison to scripted testing, thus fulfilling different purposes. While an exploratory testing approach can enable find critical and otherwise missed defects by utilizing the skill and creativity of the tester, it is also insufficient for verifying conformance to requirements due to providing weak coverage of requirements. In contrast, scripted testing provides this and is a vital component in regression testing. Thus, the question about exploratory testing is not whether or not to apply it, but rather when to apply which level of exploratory testing (full or none) to achieve the desired outcome.

There have been some previous attempts to provide structure to ET and guide the test process by defining clear test missions and time-boxing test sessions. We provide practitioners with a better understanding of ET practices by introducing the concept of a sliding scale of exploration from fully exploratory, or freestyle, to fully scripted testing. In this paper, we propose five levels of exploratory testing and present factors that are influenced by these levels. We define each level of exploration by providing test charter types with distinct elements that help practitioners to design tests at different levels of exploration.

We have explored factors related to the level of exploration through a series of focus groups. Our research shows that the exploration levels have an affect on factors such as the ability to detect defects, test efficiency, learning and motivation, and that different outcomes are to be expected depending on the chosen exploration level. Awareness of these factors allows testers to select the exploration level according to what they want to achieve with their testing. For example, testers operating at higher levels of exploration, e.g. freestyle, can expect to achieve improved defect detection, savings in time and effort, and facilitated management of evolution and change. They can also expect a positive impact with regards to learning and motivation. Though there are drawbacks too, since the higher exploration levels are weak in supporting traceability and verification of quality requirements concerning conformance and performance. Another characteristic of high levels of exploration is the weak reproducibility of defects, as the test steps are not clearly documented for developers to follow to reproduce the defect. However, we note that recent research studies provide solu-

tions for tracking testing sessions to later derive and repeat the test steps [17].

We encourage practitioners to consider striving for a combination of exploratory and scripted testing. In this way, testers can obtain the positive effects of the higher levels of exploration, while not neglecting other types of testing. As one participant stated at the end of one focus group: “we [now] think that we want both scripted and exploratory; a mix of both approaches, so that we can approach our testing in different ways.” We also found that the test charters we used when defining the exploration levels, provide practical value to the participants. The practitioners quickly grasped the differences between the exploration levels by viewing and applying these test charter types. We suggest that practitioners reflect on the levels of exploration by using a similar approach. They can explore and reflect on how the various levels of exploration could support them by rewriting existing charters or scripted tests according to the corresponding test charter types.

## ACKNOWLEDGMENT

We would like to thank the participating companies and in particular the individuals for their active involvement in and support of this research. This work was partly funded by the EASE Industrial Excellence Center for Embedded Applications Software Engineering, (<http://ease.cs.lth.se>).

## REFERENCES

- [1] W. Afzal, A. N. Ghazi, J. Itkonen, R. Torkar, A. Andrews, and K. Bhatti, “An experiment on the effectiveness and efficiency of exploratory testing,” *Empirical Softw. Eng.*, vol. 20, no. 3, pp. 844–878, 2015.
- [2] K. Bhatti and A. N. Ghazi, “Effectiveness of exploratory testing: An empirical scrutiny of the challenges and factors affecting the defect detection efficiency,” M.S. thesis, Dept. School Eng., Blekinge Institute of Technology, Karlskrona, Sweden, 2010. [Online]. Available: <https://www.diva-portal.org/smash/get/diva2:832837/FULLTEXT01.pdf>
- [3] C. Kaner, J. Bach, and B. Pettichord, *Lessons Learned in Software Testing*. Hoboken, NJ, USA: Wiley, 2008.
- [4] J. Itkonen, M. V. Mäntylä, and C. Lassenius, “Test better by exploring: Harnessing human skills and knowledge,” *IEEE Softw.*, vol. 33, no. 4, pp. 90–96, Jul./Aug. 2016.
- [5] S. M. A. Shah, C. C. Gencel, U. S. Alvi, and K. Petersen, “Towards a hybrid testing process unifying exploratory testing and scripted testing,” *J. Soft., Evol. Process*, vol. 26, no. 2, pp. 220–250, 2014.
- [6] D. Pfahl, H. Yin, M. V. Mäntylä, and J. Münch, “How is exploratory testing used? A state-of-the-practice survey,” in *Proc. 8th ACM/IEEE Int. Symp. Empirical Softw. Eng. Meas. (ESEM)*, Sep. 2014, Art. no. 5.
- [7] J. Bach, “Session-based test management,” *Softw. Test. Quality Eng. Mag.*, vol. 2, no. 6, pp. 1–10, 2000. [Online]. Available: <http://www.satisfice.com/articles/sbtm.pdf>
- [8] A. N. Ghazi, R. P. Garigapati, and K. Petersen, “Checklists to support test charter design in exploratory testing,” in *Proc. 18th Int. Conf. Agile Softw. Develop. (XP)*, 2017, pp. 251–258.
- [9] E. Engström, K. Petersen, N. bin Ali, and E. Bjarnason, “SERP-test: A taxonomy for supporting industry-academia communication,” *Softw. Quality J.*, vol. 25, no. 4, pp. 1269–1305, 2016.
- [10] J. A. Whittaker, *Exploratory Software Testing: Tips, Tricks, Tours, and Techniques to Guide Test Design*. London, U.K.: Pearson, 2009.
- [11] J. Itkonen, M. Mäntylä, and C. Lassenius, “The role of the tester’s knowledge in exploratory software testing,” *IEEE Trans. Softw. Eng.*, vol. 39, no. 5, pp. 707–724, May 2013.
- [12] M. Micallef, C. Porter, and A. Borg, “Do exploratory testers need formal training? An investigation using HCI techniques,” in *Proc. 9th IEEE Int. Conf. Softw. Test., Verification Validation Workshops (ICST Workshops)*, Chicago, IL, USA, Apr. 2016, pp. 305–314.
- [13] J. Bach, “Exploratory testing explained,” Tech. Rep., 2003. [Online]. Available: <http://www.satisfice.com/articles/et-article.pdf>
- [14] B. Suranto, “Exploratory software testing in agile project,” in *Proc. IEEE Int. Conf. Comput., Commun. Control Technol. (I4CT)*, Apr. 2015, pp. 280–283.
- [15] A. Claesson. (2007). *How to Perform Exploratory Testing by Using Test Charters*. [Online]. Available: [http://www.sast.se/q-moten/2007/stockholm/q3/2007\\_q3\\_claesson.pdf](http://www.sast.se/q-moten/2007/stockholm/q3/2007_q3_claesson.pdf)
- [16] M. Daneva, “Focus group: Cost-effective and methodologically sound ways to get practitioners involved in your empirical RE research,” in *Proc. Joint Workshops, Res. Method Track, Poster Track Co-Located 21st Int. Conf. Requirements Eng., Found. Softw. Quality (REFSQ)*, Essen, Germany, Mar. 2015, pp. 211–216.
- [17] E. Alegroth, R. Feldt, and P. Kolström, “Maintenance of automated test suites in industry: An empirical study on visual GUI testing,” *Inf. Softw. Technol.*, vol. 73, pp. 66–80, Feb. 2016.



**AHMAD NAUMAN GHAZI** received the M.Sc., Licentiate of Technology, and Ph.D. degrees in software engineering from the Blekinge Institute of Technology, Sweden, in 2010, 2014, and 2017, respectively. He is currently a Lecturer with the Department of Software Engineering, Blekinge Institute of Technology. He has extensive experience with software industry, where he was a Software Test Engineer for several years. His research interests include empirical software engineering, software verification and validation, exploratory testing, agile software development, software quality assurance, and software process improvement.



**KAI PETERSEN** received the Ph.D. degree from the Blekinge Institute of Technology, Sweden, in 2010. He is currently a Professor with the Department of Software Engineering, Blekinge Institute of Technology. He has authored over 70 research works in international journals and conferences. His research focuses on software processes, software metrics, lean and agile software development, quality assurance, and software security in close collaboration with industry partners.



**ELIZABETH BJARNASON** received the Ph.D. degree from Lund University, Sweden. She was with software and telecommunications industry for several years. She is currently a Senior Lecturer of software engineering with the Department of Computer Science, Lund University. Her research interests include empirical research and theory building on requirements communication and collaboration, in particular towards software testing.



**PER RUNESON** (M’98) is currently a Professor of software engineering with Lund University, Sweden, the Head of the Department of Computer Science, and the Leader of the Software Engineering Research Group and the Industrial Excellence Center on Embedded Applications Software Engineering. He is the principal author of *Case Study Research in Software Engineering* and co-authored the *Experimentation in Software Engineering*. His research interests include empirical research on software development and management methods, in particular for verification and validation. He is a member of several program committees. He serves on the Editorial Board for the *Empirical Software Engineering* and the *Software Testing, Verification and Reliability*.

...