In practice

Efficient and effective exploratory testing of large-scale software systems☆

Torvald Mårtensson [a,*], Daniel Ståhl [b,c], Antonio Martini [d], Jan Bosch [e]

[a] *Saab AB, Linköping, Sweden*
[b] *Ericsson AB, Linköping, Sweden*
[c] *Linköping University, Linköping, Sweden*
[d] *University of Oslo, Oslo, Norway*
[e] *Chalmers University of Technology, Gothenburg, Sweden*

A B S T R A C T

Context: Exploratory testing plays an important role in the continuous integration and delivery pipelines of large-scale software systems, but a holistic and structured approach is needed to realize efficient and effective exploratory testing.
Objective: This paper seeks to address the need for a structured and reliable approach by providing a tangible model, supporting practitioners in the industry to optimize exploratory testing in each individual case.
Method: The reported study includes interviews, group interviews and workshops with representatives from six companies, all multi-national organizations with more than 2,000 employees.
Results: The ExET model (Excellence in Exploratory Testing) is presented. It is shown that the ExET model allows companies to identify and visualize strengths and improvement areas. The model is based on a set of key factors that have been shown to enable efficient and effective exploratory testing of large-scale software systems, grouped into four themes: "The testers' knowledge, experience and personality", "Purpose and scope", "Ways of working" and "Recording and reporting".
Conclusions: The validation of the ExET model showed that the model is novel, actionable and useful in practice, showing companies what they should prioritize in order to enable efficient and effective exploratory testing in their organization.

1. Introduction

1.1. Background and related work

Continuous Delivery was popularized as a term by Humble and Farley (2011), who state that in continuous delivery "every change is, in fact, a release candidate". Even though we in our previous work (Ståhl et al., 2017) have found that there is some confusion around the terminology, we find that the most common interpretation is that continuous delivery is about ensuring that the software can be released and deployed to production at any time, but may not actually be thus released and/or deployed. Humble and Farley (2011) describe the principles behind continuous delivery, and the practices necessary to support it. The central paradigm of the book is the integration pipeline (also referred to as the deployment pipeline). The pipeline, as described by Humble and Farley, consists primarily of automated test activities, but also includes usability testing and exploratory testing.

Exploratory testing was coined as a term by Cem Kaner in the book "Testing Computer Software" (Kaner, 1988), and was then expanded upon as a teachable discipline by Cem Kaner, James Bach and Bret Pettichord in their book "Lessons Learned in Software Testing" (Kaner et al., 2001). Exploratory testing combines test design with test execution, and focuses on learning about the system under test. The test technique is focused on learning, shown in for example Hendrickson's (2013) definition of exploratory testing: "Simultaneously designing and executing tests to learn about the system, using your insights from the last experiment to inform the next". Gregory and Crispin (2015) describe how exploratory testing and automated testing complement each other: "Exploratory testing and automation aren't mutually exclusive but rather work in conjunction. Automation handles the day-to-day repetitive regression testing (checking), which enables the exploratory testers to test all the things the team didn't think about before coding".

James Bach emphasizes that exploratory testing should not be mistaken for unstructured testing (Bach, 2020). Different setups have been presented for planning, execution and reporting exploratory testing: testing can be organized as charters (Gregory and Crispin, 2015; Hendrickson, 2013) or tours (Gregory and Crispin, 2015; Whittaker, 2010) which are conducted as sessions (Gregory and Crispin, 2015; Hendrickson, 2013) or threads (Gregory and Crispin, 2015).

1.2. Research question

In our previous work (Mårtensson et al., 2017) we have shown that exploratory testing plays an important role in the continuous integration and delivery pipeline for a large-scale software system. Whereas automated test activities in the pipeline are able to rapidly provide feedback to developers and to verify requirements, exploratory testing can provide more in-depth insights about the system under test. In a study based on both quantitative and qualitative data, we developed a test method for exploratory testing of large-scale systems, validated in a large-scale industry project (Mårtensson et al., 2017).

Since our previous study on exploratory testing, we have identified a growing interest in exploratory testing from several of the companies we as researchers work with, and an interest to improve how the test technique is used in the companies. A holistic and structured approach is lacking that can point out the most important problem areas in each individual case. Based on this, the topic of this research paper is to answer the following research question:

- *RQ1: How can efficient and effective exploratory testing be realized in industrial organizations?*

In order to work with RQ1 in a step-by-step approach, the research question was during the study broken down into the following detailed research questions:

- *RQ1.1: How are key factors that enable efficient and effective exploratory testing of large-scale software systems described in literature?*
- *RQ1.2: What are, according to practitioners in industry, the key factors that enable efficient and effective exploratory testing of large-scale software systems?*
- *RQ1.3: How can a model be constructed to support practitioners in the industry to optimize exploratory testing in each individual case?*

1.3. Contribution

The contribution of this paper is two-fold: First, it presents results from interviews and workshops with participants from six case study companies, that can give researchers and practitioners an improved understanding how exploratory testing is used in large-scale industry projects. Second, it presents a new model (the ExET model) that can be used by practitioners in industry to identify and visualize strengths and improvement areas, which can be used to optimize exploratory testing in each individual case.

In this paper, we have built upon our previous work (Mårtensson et al., 2019a), introducing the ExET model (Excellence in Exploratory Testing) based on the nine identified key factors. The model is presented with a two-step process describing how the ExET model is used, and the validation of the ExET model (industrial evaluation). This paper also includes a more extensive presentation of the research method and a more extensive analysis of threats to validity. The five steps in the reported study and the results from each step are further described in Section 2.1.

The remainder of this paper is organized as follows: In the next section, we present the research method. This is followed in Section 3 by a study of related literature. In Section 4 we present the analysis of the interview results (the list of key factors), followed by a summary of the follow-up interviews and the cross-company workshop in Section 5. The ExET model is presented in Section 6, followed by a presentation of the validation of the ExET model in Section 7. Threats to validity are discussed in Section 8. The paper is then concluded in Section 9.

2. Research method

2.1. Overview of the research method

The research study reported in this paper consists of five major steps:

- *Step 1: Reviewing literature*: A literature review to investigate if any key factors that can enable efficient and effective exploratory testing have been previously presented in literature (presented in Section 3).
- *Step 2: Identifying the key factors*: A series of interviews to identify key factors that enable efficient and effective exploratory testing of large-scale software systems (presented in Section 4).
- *Step 3: Confirming the key factors*: A second series of follow-up interviews and a cross-company workshop in order to confirm the interpretation of the first series of interviews, as well as looking for negative cases (presented in Section 5).
- *Step 4: Development of the ExET model*: Analysis of the results from the two series of interviews, and development of the Excellence in Exploratory Testing (ExET) model (presented in Section 6).
- *Step 5: Validation of the ExET model*: Validation of the ExET model: Validation of the ExET model in interviews and workshops, and a comparison with the literature reviews (presented in Section 7).

An overview of the research method is presented in Fig. 1: As a first step, a literature review was conducted in order to look for solutions in published work. In parallel, a series of interviews was conducted to identify key factors that enable efficient and effective exploratory testing of large-scale software systems. This was followed by follow-up interviews and a cross-company workshop, to confirm the results from the first series of interviews. The development of the ExET model was based primarily on the results from the two series of interviews, as the literature review did not result in a comprehensive set of key factors. As the final step, the ExET model was validated with interviews and workshops, and a comparison with the literature review.

Fig. 1 also shows how the steps of the study addresses the detailed research questions (defined in Section 1.2) and how the case study companies were included in the different steps of the study. The research method for each part of the study is further described in Sections 2.2–2.6.

The study includes six case study companies, referred to as Company A, Company B, Company C, Company D, Company E and Company F. This study focuses on testing of functions fully or partly implemented in software. The six companies operate in the following industry segments:

- Company A: Automotive products and services
- Company B: Transport solutions for commercial use
- Company C: Video surveillance cameras and systems
- Company D: Services and solutions for military defense and civil security
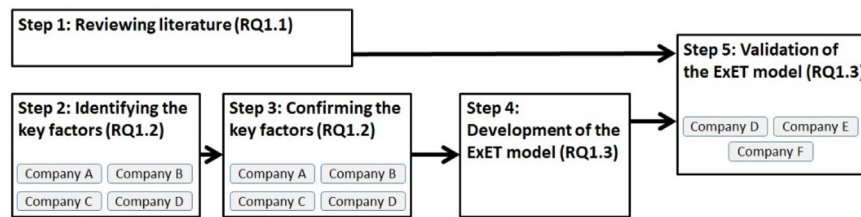- Company E: Communications systems and services

Fig. 1. An overview of the research method.

- Company F: Development, manufacturing and maintenance of pumps

The studied companies are all multi-national organizations with more than 2,000 employees. All companies develop large-scale and complex software systems for products, which also include a significant amount of mechanical and electronic systems.

The companies were considered to be suitable for the study, as they have similar characteristics, but at the same time are operating in different industry segments. The companies were selected as they wanted to increase the amount of exploratory testing in their organization, and improve the ways in which exploratory testing was used. The validation cases in Company D, Company E and Company F were purposely selected for the external validation, as they were separated from the primary studies.

2.2. Reviewing literature

To investigate whether solutions related to the research question have been presented in published literature, a literature review was conducted, following the guidelines established by Kitchenham (2004). A review protocol was created, containing the question driving the review ("How are key factors that enable efficient and effective exploratory testing of large-scale software systems described in literature?") and the inclusion and exclusion criteria (see Table 1). The review was purposely designed to search for key factors that can enable efficient and effective exploratory testing (i.e. not necessarily presented as a method or a model) in order to include all types of material that could be relevant to the research question defined in Section 1. The stages of the review (according to the guidelines from Kitchenham) were:

- *Identification of research*: Iterative analysis of title, abstract and keywords of publications from trial searches using various combinations of search terms.
- *Selection of primary studies*: Exclusion of duplicates and conference proceedings summaries.
- *Study quality assessment*: The relevance of the selected research papers was assessed in a first review of each paper, and papers considered to be not relevant were excluded.
- *Data extraction & monitoring*: Characteristics and content of the remaining research papers were documented in an iterative process.
- *Data synthesis*: The results from the review were collated and summarized.

The analysis of the results from the literature review was conducted iteratively in a two-step process, involving three of the researchers in order to secure quality and correctness.

The results from the literature review was a better understanding of previously published literature related to the research question, which later in the study was used as input to the validation of the ExET model.

2.3. Identifying the key factors

In order to find a solution for the research question presented in Section 1, a series of interviews were conducted to identify a set of key factors that enable efficient and effective exploratory testing of large-scale software systems. The interviews included 20 individuals from four of the case study companies (Company A, Company B, Company C and Company D). Five interviewees participated from Company A, three from Company B, six from Company C and six from Company D. All of the interviewees had experiences from exploratory testing as testers, and in some cases also as test leaders. The interviews lasted from half an hour up to (in most cases) one hour. They were conducted as semi-structured interviews, using an interview guide with pre-defined specific questions (interview guide 1 in Appendix). The interview questions were designed to provide information leading to the identification of key factors, but also to provide background information about the interviewee and how exploratory testing was used in the interviewee's organization. The interviewer was summarizing and transcribing the interviewee's response during the interview, and each response was read back to the interviewee to ensure accuracy. The interview questions were sent to the interviewee at least one day in advance.

The main question in the interview guide (Q1.6) asked the interviewees to describe the key factors they believed can enable efficient and effective exploratory testing of large-scale software systems. The responses for Q1.6 included a large amount of statements and comments. The interview results were analyzed based on thematic coding analysis as described by Robson and McCartan (2016), outlined in the following bullets:

- *Familiarizing with the data*: Reading and re-reading the transcripts, noting down initial ideas.
- *Generating initial codes*: Extracts from the transcripts are marked and coded in a systematic fashion across the entire data set.
- *Identifying themes*: Collating codes into potential themes, gathering all data relevant to each potential theme. Checking if the themes work in relation to the coded extracts and the entire data set. Revising the initial codes and/or themes if necessary.
- *Constructing thematic networks*: Developing a thematic 'map' of the analysis.
- *Integration and interpretation*: Making comparisons between different aspects of the data displayed in networks (clustering and counting statements and comments, attempting to discover the factors underlying the process under investigation, exploring for contrasts and comparisons). Revising the thematic map if necessary. Assessing the quality of the analysis.

The process was conducted iteratively to increase the quality of the analysis, reaching consensus within the group of researchers through discussions and visualization in diagrams and text. Comments and statements from the interviewees were as a first step sorted into categories, which during the process were

reorganized into new structures. For example, statements regarding a test environment suitable for debugging and statements regarding tools for recording for were originally handled separately, but were then merged into the same category. The remaining nine themes were then described, with two or three representative quotes selected from the transcripts included in the descriptions. Special attention was paid to outliers (interviewee comments that do not fit into the overall pattern) according to the guidelines from Robson and McCartan (2016) in order to strengthen the explanations and isolate the mechanisms involved.

The results from the analysis of the first series of interviews was a list of identified key factors which can enable efficient and effective exploratory testing.

2.4. Confirming the key factors

The next step of the study was follow-up interviews with the same 20 individuals as the first series of interviews, to collect feedback on the identified key factors. The purpose of the follow-up interviews was to confirm that the interpretation of the first series of interviews was correct, as well as looking for negative cases.

The interviews in the second series of interviews lasted for about half an hour, and were conducted in a similar way as the first series of interviews: The interviews were semi-structured interviews, using an interview guide with pre-defined specific questions (interview guide 2 in Appendix). The interview questions were designed to provide information about how the identified key factors were received by the interviewee. The interviewer was summarizing and transcribing the interviewee's response during the interview, and each response was read back to the interviewee to ensure accuracy. The interview questions were sent to the interviewee at least one day in advance.

In order to achieve method and data triangulation (Runeson and Höst, 2009), the follow-up interviews were complemented with a cross-company workshop with 14 participants representing all four case study companies participating in the two series of interviews (Company A, Company B, Company C and Company D). The participants at the workshop had roles in their companies as tester, test leader, test specialist and line manager. At the workshop, two of the researchers presented the results from the literature review and the two series of interviews. The workshop participants discussed the presented key factors, and discussed the identified differences between the companies.

Researcher bias could be a threat to reliability during the interpretation of interview results. This threat was mitigated with member checking (the follow-up interviews) and a focus group (the cross-company workshop) as described in Section 5, following the guidelines from Robson and McCartan (2016) who consider this to be "a very valuable means of guarding against researcher bias" and a good way to "amplify and understand the findings".

The results from the analysis of the follow-up interviews and the cross-company workshop was a better understanding of the identified key factors, and confirmation of the validity of the key factors in the case study companies.

2.5. Development of the ExET model

The Excellence in Exploratory Testing (ExET) model was developed primarily based on the analysis of the results from the two interview series described in Sections 2.3 and 2.4, which included interviewees from Company A, Company B, Company C and Company D. The analysis and design process was conducted iteratively to increase the quality of the results.

Table 1
Inclusion and exclusion criteria for the literature review.

Inclusion criterion	Yield
Publications matching the Scopus search string `TITLE-ABS-KEY ("exploratory testing")` on October 27, 2018	129

Exclusion criterion	Remaining
Excluding duplicates and conference proceedings summaries	122
Excluding publications not related to development of software systems	71
Excluding publications with no available full-text	65

The ExET model is a representation of strengths and improvement areas in an organization, which can be used to improve exploratory testing. The ExET model was designed to provide as much information as possible about the strengths and possible improvement areas in the organization. One factor (factor six) represents balance between freedom and structure, and was therefore split into two statements (6a and 6b) in order to isolate the root causes as much as possible.

The ExET model is used in a simple two-step process, in order to both involve the testers and the stakeholders for the test activities. A simplistic approach was purposely selected, based on our experiences and positive feedback received in the validation phases in our previous studies (Mårtensson et al., 2018, 2019b).

The results from this part of the research study was a model (the ExET model) which can show companies what they should prioritize in order to optimize exploratory testing in their organization.

2.6. Validation of the ExET model

The ExET model was validated using the following methods to achieve method and data triangulation according to the guidelines from Runeson and Höst (2009):

- *Validation interviews and workshops*: Interviews and workshops in five validation cases in three companies (Company D, Company E and Company F). The interviewees and workshop participants used the ExET model to evaluate the status in their organization to identify improvement areas. The purpose of the validation interviews and workshops was to evaluate if the ExET model was considered actionable and useful in practice.
- *Validation cross-company workshop*: A cross-company workshop with participants from all six companies in the study (Company A, Company B, Company C, Company D, Company E and Company F). The workshop participants discussed the results from the validation cases, further strengthening the reliability of the results from the study.
- *Comparison with literature review*: Comparison of the ExET model and related work found in literature, primarily to discuss the novelty of the ExET model.

The validation cases in Company D, Company E and Company F were purposely selected for the external validation as they are separated from the primary studies (resulting in the ExET model). The organization in each validation case wanted not just to improve how they used exploratory testing, but also to increase the amount of exploratory testing in the organization. The design of the validation of the ExET model is described in-depth in Section 7.1.

The validation interviews and workshops (in all validation cases) were held within a time-frame of three months, and the

validation cross-company workshop about one month later. Altogether, all validation activities were conducted within a time frame of four months.

The results from the validation of the ExET model was an industrial evaluation of how the model delivers value to the industry, and identified suggestions for how the model can be improved.

3. Reviewing literature

3.1. Criteria for the literature review

In order to investigate if any key factors that can enable efficient and effective exploratory testing have been previously presented in related work, a literature review (Kitchenham, 2004) was conducted. The question driving the review was "How are key factors that enable efficient and effective exploratory testing of large-scale software systems described in literature?" The inclusion criterion and the exclusion criterion for the review are shown in Table 1. To identify published literature, a Scopus search was conducted. The decision to use only one indexing service was based on that we in previous work (Ståhl and Bosch, 2014, 2016; Ståhl et al., 2016) have found Scopus to cover a large majority of published literature in the field, with other search engines only providing very small result sets not already covered by Scopus. This threat to validity is also discussed in Section 8.1

Our previous work (Mårtensson et al., 2017) includes a similar literature review, covering 52 publications. For the literature review in this study we expanded the search scope, which yielded 129 publications. After removing duplicates and conference proceedings summaries, the abstract of the remaining 122 publications were reviewed manually. Publications not related to development of software systems were excluded, i.e. removing publications related to archeology, chemistry et cetera. As a final stage, publications for which we could not find any available full-text were excluded from the literature review.

Characteristics and content of the remaining 65 research papers were then documented in a consistent manner in a data extraction protocol: for each paper a summary of how the paper was related to the research question, and representative keywords and quotes (sorted into categories which emerged during the process). The process was conducted iteratively to increase the quality of the analysis. Finally, the results from the review were collated and summarized.

3.2. Results from the literature review

An overview of the publications found in the literature review is presented in Table 2. The review of the 65 publications retrieved from the search revealed that nine of the publications were not directly related to exploratory testing (only mentioning exploratory testing in passing while discussing other test techniques).

Ten of the publications are comparing exploratory testing and other test techniques, typically comparing exploratory testing and scripted testing (also referred to as test case based testing, specification based testing or confirmatory testing). The comparisons were based on literature reviews, true experiments, and case studies from industry projects. The papers describe or touch upon the strengths and weaknesses of exploratory testing (e.g. Shah et al., 2014a), but do generally not define key factors for efficient or effective exploratory testing.

Twenty publications propose new methods that in different ways involve or include a reference to exploratory testing. Eleven of those publications present new methods or approaches, which combine exploratory testing and another test technique (Basri

Table 2
An overview of the publications found in the literature review.

Topic of the publications	Number of papers
Not relevant	9
Comparing exploratory testing and other test techniques	10
Methods	20
Tools	7
How exploratory testing is used	9
Reporting experiences	10
Summary	**65**

et al., 2019; Calpur et al., 2017; Frajtak et al., 2016; Gebizli and Sozer, 2014; Hellmann and Maurer, 2011; Hudson and Denzinger, 2015; Kim and Lee, 2014; Kuhn, 2013; Mihindukulasooriya et al., 2016; Rashmi and Suma, 2014; Schaefer and Do, 2014). These methods try to combine the flexibility of exploratory testing with the structure provided by scripted test cases. As one example, Frajtak et al. (2016) describe that the testers can use "their skills and intuition to explore the system", but "it is hard to measure the effectiveness of the [exploratory] testing". As a solution, Frajtak et al. propose a technique where recording of the (exploratory) testers actions are used to create test case scenarios. Ghazi et al. (2017) provide a different kind of structure, aiming to support practitioners in the design of test charters through checklists. Sviridova et al. (2013) discuss effectiveness of exploratory testing and propose to use scenarios. The level of freedom in exploratory testing is discussed by Ghazi et al. (2018), presenting a scale for the degree of exploration and defining five levels. Raappana et al. (2016) report the effectiveness of a test method called "team exploratory testing", which is defined as a way to perform session-based exploratory testing in teams. One of the papers is our previous work (Mårtensson et al., 2017), presenting a test method for exploratory testing of large-scale systems (as described in Section 1). Finally, Shah et al. (2014a,b) take a somewhat different approach, describing exploratory testing as a source of technical debt, and propose (as a solution to this problem) that exploratory testing should be combined with other testing approaches.

Seven of the publications present different types of tools, developed to increase the efficiency in exploratory testing. However, three of the papers does not include any validation of the presented tool. The remaining four papers describe tools developed to visualize how the executed testing covers the system under test (Bures et al., 2018; Frajtak et al., 2017), visualize code changes in the system under test (Reis and Mota, 2018), and refine system models based on recorded testing activities (Gebizli and Sözer, 2017a).

Nine publications describe in different ways how exploratory testing is used by the testers. Four of those publications (Gebizli and Sözer, 2017b; Itkonen et al., 2009, 2013; Micallef et al., 2016) focus on the tester's knowledge and experience: Itkonen et al. (2013) discuss how testers recognize failures based on their personal knowledge without detailed test case descriptions ("domain knowledge, system knowledge, and general software engineering knowledge"). Gebizli and Sözer (2017b) present results from a study showing that both educational background and experience level has "significant impact" on the efficiency and effectiveness of exploratory testing. In contrast to that, two papers (Pfahl et al., 2014; Shoaib et al., 2009) focus on the tester's personality: Shoaib et al. (2009) simply conclude that "people having extrovert personality types are good exploratory testers". Pfahl et al. (2014) analyzes the results from an online survey, and finds that exploratory testing "is as an approach that supports creativity during testing and that is effective and efficient".

Tuomikoski and Tervonen (2009) embrace both approaches, stating that "the effectiveness of exploratory testing is strongly based on individual test engineer's skills and ability to analyze system and its behavior" but also that "exploratory testing doesn't fit for everyone, and really requires experienced test engineers".

Finally, ten papers report experiences from exploratory testing in industry, but without presenting any documented quantitative or qualitative data. Pichler and Ramler (2008) present experiences from development of mobile devices, and find that "tool support enhances the capability of human testers". Kumar and Wallace (2013) describe that for the exploratory tester, it is "easy to get lost in a thicket of well-intentioned heuristics", and proposes the use of "problem frames" as a solution for this problem.

In summary, we found no publication summarizing key factors that enable efficient and effective exploratory testing. Some of the identified papers discuss isolated factors, e.g. how the effectiveness of exploratory testing is based on the testers' knowledge and personality, or how tools can increase the efficiency in exploratory testing. However, the results from the literature review shows that previously published work lacks a holistic perspective and instead tend to focus on one aspect, leaving out areas that other authors consider to be the core issues.

4. Identifying the key factors

4.1. Background information

The literature review (presented in Section 3) did not result in a comprehensive set of key factors. In order to find a solution for the research question presented in Section 1, a series of interviews were conducted to identify a set of key factors that enable efficient and effective exploratory testing of large-scale software systems. The key factors were later in the study used in the development of the ExET model, which is described in detail in Section 6.1.

The interviews were based on an interview guide with pre-defined specific questions (interview guide 1 in Appendix). Twenty individuals participated in the interviews, with an average of 13 years of experience of industry software development (spanning from 4 to 46 years). The interviewees all had experiences from exploratory testing as testers, and in some cases also as test leaders. The interviewees had a very positive attitude towards exploratory testing, rating "the value of exploratory testing as a test technique for large-scale software systems" as 4 or 5 on a Likert scale from 1 ("very low") to 5 ("very high").

Interviewees from all four case study companies described that exploratory testing was used for two purposes in their organization: to find bugs during development of new functions and systems, and for testing of the complete system. The interviewees described that exploratory testing was used primarily for new functions, whereas automated testing and manual scripted testing was used primarily for regression tests.

The interviewees were also asked to describe strengths and weaknesses with exploratory testing. Generally, the interviewees described that exploratory testing is a good way to find problems quickly and efficiently. Exploratory testing is also a more creative way to work for the testers, and was therefore considered to make better use of the testers. Exploratory testing was also described as a good way to test system-wide and to test large-scale systems, especially exploratory testing with an end-user perspective. The interviewees also described a few weaknesses with exploratory testing: Some interviewees described that they believed that exploratory testing was more difficult for new testers (as experience is more important). Another viewpoint was that it could be more difficult to describe what you have tested, compared to if you follow a scripted test case.

4.2. Key factors for efficient and effective exploratory testing

The main question of the first series of interviews was: "What are the key factors that you think enable efficient and effective exploratory testing of large-scale software systems?" The responses for this question included a large amount of statements and comments. Extracts from the interview responses were sorted into categories, which during the process were reorganized into new structures (as described in Section 2.3). A thematic network were constructed (Robson and McCartan, 2016), resulting in a thematic map with four main themes, which in turn consist of several sub-themes: a list of key factors which can enable efficient and effective exploratory testing.

The four main themes (groups of key factors) and their sub-themes (the key factors) are shown in Table 3, together with information about of how many interviewees that provided statements supporting each factor and group of factors.

All 20 interviewees talked about the importance of *the testers' knowledge, experience and personality*. In order to test the system, the testers must *know how the system is built, and the correct behavior of the system*. One interviewee asked for "testers with different types of experiences". Another voice asked for "good system knowledge". One interviewee described that the tester must have "test confidence", meaning that as a tester you should "trust your instinct that this is wrong". The testers must also *know how the product is used by the end-user* (or the end user should be represented in the test team). To quote one of the interviewees: "If you have the end-user perspective, then you know if a problem is a problem". One interviewee even stated that "you should always have the end-user perspective". The testers must also *be curious and want to learn about the system*. This means that even if the tester has good system knowledge, he/she still wants to learn more. The interviewees described that this calls for certain types of personalities, e.g. "the right personality, to be interested in new perspectives, curiosity, imagination". Some of the interviewees also described this as an interest in tracking down the problems in the system, e.g. "someone who want to find the bugs – curious and creative people".

As many as 18 interviewees did in different ways talk about the *purpose and scope* of the tests. This includes a *well-defined purpose and scope for the tests (system functions ready to be tested)* which the testers can transform into e.g. scenarios or focus areas. One interviewee described that "you should have a list of functions that should be tested". Several interviewees clarified that this should not be interpreted as that exploratory testing only could be used at a final stage of a project, and that this also affect development planning: "Test when the function is ready, and not too early. You must build the function in steps so it can be tested". Another interviewee had a similar comment: "Do exploratory testing early, but test complete functions". A related area is *regression testing to secure basic stability and integrity in the system* (before exploratory testing). Efficient regression testing finds problems in legacy functions (introduced due to dependencies between functions or systems). If this works well, skilled exploratory testers will not waste their time investigating and reporting problems with legacy functions, but can instead focus on testing the new functions. As one of interviewees put it: "Simple problems should be identified and corrected from automated testing". In the same way, the testers' time is not wasted at trouble-shooting problems that has already been analyzed. To quote one of the interviewees: "what is the status of the product, what are the known errors or problems".

Ways of working were discussed by 14 of the 20 interviewees. The interviewees requested *an established way of working*, including planning meetings, preparations, test strategies and heuristics (a balance between structure and freedom). The statements from the interviewees were quite general, e.g. "some kind

Table 3

The nine key factors that can enable efficient and effective exploratory testing, and the number of interviewees that provided statements supporting each factor and group of factors.

		Interviewees
The testers' knowledge, experience and personality		**20**
1.	– The testers know how the system is built, and the correct behavior of the system	14
2.	– The testers know how the product is used by the end-user (or the end user is represented in the test team)	12
3.	– The testers are curious and want to learn about the system	16
Purpose and scope		**18**
4.	– A well-defined purpose and scope for the tests (system functions ready to be tested) which the testers can transform into e.g. scenarios or focus areas	10
5.	– Regression testing secure basic stability and integrity in the system (before exploratory testing)	11
Ways of working		**14**
6.	– An established way of working, including e.g. planning meetings, preparations, test strategies and heuristics (a balance between structure and freedom)	10
7.	– Testers with different experiences and competences work together as a team, helping each other with new ideas and knowledge about different parts of the system	12
Recording and reporting		**17**
8.	– Test environments that support debugging and recording	10
9.	– A well-defined way to report the test results, including a description of areas covered by the tests and a list of identified problems	12

of structure for the testing". However, the interviewees also emphasized that this structure should never be at the same detailed level as manual test-case-based testing, described e.g. by one of the interviewees as to "find the balance between freedom and traceability". One interviewee stated that "You must have fun – play around with things". The interviewees particularly emphasized advantages from that *testers with different experiences and competences work together as a team*, helping each other with new ideas and knowledge about different parts of the system. Some interviewees asked for that the testers should do the testing together ("test together with colleagues"). Others focused on that testers should help each other with new ideas, preparations et cetera ("a structure for how testers should support each other"). Generally, the interviewees asked for individuals with different knowledge and experience ("a team with a mix of individuals").

Seventeen of the 20 interviewees talked about *recording and reporting*. The interviewees *described test environments that support debugging and recording* as a prerequisite for efficient testing, in order to provide detailed data about the identified problems. The interviewees also asked for recording in order to document the testing. Quoting one of the interviewees: "An efficient way of documenting what you do". The interviewees also asked for *a well-defined way to report the test results*, including a description of areas covered by the tests and a list of identified problems. Reporting should not be limited to only problem reports, as it is also important to describe which areas of the product that has been tested. This is important in order to avoid that testers (or test teams) spend time testing the same things, and to secure that the purpose and scope of the testing is fully covered. As one of the interviewees phrased it: "You must document in a good way what you have done, otherwise you might miss important areas". There were different opinions about how the results from the testing should be reported. One interviewee explained that test results should be visualized, i.e. not only described in text. Another interviewee argued that the best way is to "involve the people interested in the test in the test".

In summary, we find that the thematic coding analysis of the interview results resulted in nine factors, all supported by statements or comments from at least ten of the interviewees. The nine factors were arranged in four groups, each group supported by between fourteen and twenty interviewees.

5. Confirming the key factors

5.1. Follow-up interviews

In order to strengthen the validity of the findings from the first series of interviews, a second series of interviews was conducted (interview guide 2 in Appendix). The list of key factors (presented in Table 3) were presented to the interviewees. The interviewees were then asked to rate the importance of each factor on a Likert scale from 1 ("not important") to 5 ("very important"). This means that the interviewees did not only provide feedback on the interpretation of their own responses, but were also providing feedback on the input from 19 other interviewees from four companies.

The interviewees confirmed the list of factors, rating the importance of each factor as 4 or 5. The interviewees often added comments like "All factors seem to be very important" or "All factors are relevant and good – they describe prerequisites for good testing". The interviewees were also asked to explain if they had not for example talked about e.g. test environments in the first interview, but now rated this factor as very important. Generally, the interviewees explained this with e.g. "the things I did not talk about [at the first interview] was what I took for granted" or similar comments. Even though the interviewees described all of the factors as relevant and important, some of the interviewees did also in different ways emphasize the importance of one or several of the factors:

- Four interviewees emphasized the importance of the testers (the first three factors).
- Four interviewees commented on the importance of balance between structure and freedom (factor six)
- One interviewee described "working together" as most important (factor seven)
- Two interviewees described reporting of test results as difficult, and how a good test report is depending on a good test environment (factor eight and nine)

Three interviewees commented on the fact that the factors seem to be correlated, e.g. to work together as a team is less important if every single tester covers all types experiences and competences. One interviewee rated five of the factors as 2 or 3,

and suggested changes such as "I want to include the term domain knowledge in this factor". However, the same interviewee concluded with "the list is great, but it can be better".

In summary, we find that all of the interviewees confirmed the list of key factors, often with positive or very positive comments. We find that the second series of interviews confirm the results from the first series of interviews, showing that the key factors presented in Table 3 in a good way reflect the interviewees' positions and viewpoints.

5.2. Cross-company workshop

To complement the second series of interviews, a cross-company workshop was organized with participants from all four case study companies. Fourteen individuals participated in the workshop, five of them with roles as manager, test leader or test specialist. Four of the participants at the workshop had also been participating as interviewees in the two series of interviews. At the workshop, two of the researchers presented the results from the literature review and the two series of interviews. The workshop participants discussed the presented key factors (from Table 3) but had only minor comments regarding the factors (e.g. "to talk to the developers is probably better than just to write a problem report").

The researchers also presented a summary of differences between the case study companies: All factors were supported by comments or statements from interviewees from all case study companies, except for the factor "A well-defined purpose and scope for the tests" (coming from three of the four companies). However, workshop participants from the fourth company described that they also work with purpose and scope, and considered this to be important. Another difference between the companies identified in the first series of interviews was that "Work together as a team" was implemented differently in the case study companies: In two of the companies the testers worked together in the test facilities. In the other two companies the testers tested separately, but worked together and supported each other as a team in other ways, e.g. with planning and preparations.

The workshop participants were asked to rate their current situation with regards to each factor. This revealed differences between the companies, and encouraged discussions related to some of the factors, e.g.:

- Should the exploratory test teams cover all types of end-users (such as e.g. a service technician), or focus only on the primary user (e.g. the driver of the car)?
- How are stability and integrity best maintained in the system — with e.g. 100% code coverage on component level, or a mix of component tests and system tests?
- What does reporting actually mean — does it include to follow up that the information has actually been received by the R&D department?

Participants representing all four case study companies showed interest in a continued study, with the purpose to construct a method or a model based on the identified nine key factors. The method/model could then be used to evaluate the current situation in an organization, and provide input to improvement initiatives related to exploratory testing. The workshop participants discussed if such a model could include the stakeholders who can enable solutions for the different key factors. It then became evident that the workshop participants had quite different opinions, also among participants from the same company. As it seems, it is difficult to identify one single role who can enable a factor. Instead, many roles are involved, or
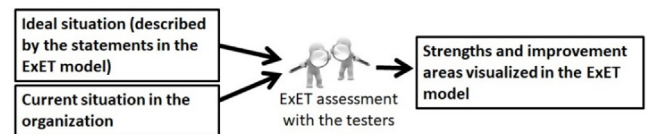

Fig. 2. Strengths and improvement areas for exploratory testing visualized in the ExET model.

it could be difficult to identify any relevant role or part of the organization.

The workshop participants also discussed how the key factors are correlated, e.g. if the testers have the right knowledge, experience and personality (the first three factors), they probably need less support from an established way of working (the sixth factor). Generally, almost all of the identified connections pointed towards the first two key factors (related to the testers' knowledge and experience). We interpret this as that factors related to "the right people" should be considered to be more important than factors related to "the right structure".

Finally, the workshop participants discussed how exploratory testing is related to Agile methodologies. The individuals participating at the workshop expressed very different opinions. One of the workshop participants had the opinion that exploratory testing harmonize well with Agile development ("exploratory testing is enabling Agile"). Another individual had a somewhat opposite opinion ("Agile is killing exploratory testing") based on that Agile methodologies often tend to focus on automated testing.

We find that the cross-company workshop confirmed the findings from the two series of interviews, and that the identified key factors are valid for all four case study companies. The interest from the workshop participants in the construction of an actionable method or model based on the nine key factors shows that the results from the study are considered to be useful in practice. One workshop participant concluded the workshop with the words "It was good to see the nine factors, and to see that also other companies find these things to be important. This makes it easier for us to change things back home".

6. The ExET model

6.1. A description of the ExET model

In the studies of related work (presented in Section 3) we found several publications presenting material related to the efficiency and effectiveness of exploratory testing. However, all of the reviewed publications tend to focus on one aspect, leaving out areas that other authors consider to be the core issues. In response to this, we developed the ExET model (Excellence in Exploratory Testing) based on the analysis of the interview results presented in Sections 4 and 5.

The ExET model is a representation of strengths and improvement areas in an organization, which can be used to optimize exploratory testing in each individual case. The ideal situation in the model is represented by set of statements (presented in Table 4). At an ExET assessment, the participants compare their current situation to the ideal situation (as shown in Fig. 2). The result is a representation in the ExET model of the current status in the organization. The procedure to use the model is further described in Section 6.2, including an example of visualization of the results from an ExET assessment (Fig. 4).

In the ExET model, the key factors from the thematic coding analysis (presented in Section 4.2) have been transformed into a list of statements. The statements, representing the ideal situation

Table 4
Statements representing the ideal situation in the ExET model.

1.	The testers know how the system is built, and the correct behavior of the system
2.	The testers know how the product is used by the end-user (or the end user is represented in the test team)
3.	The testers are curious and want to learn about the system
4.	We have a well-defined purpose and scope for the tests (system functions ready to be tested) which the testers can transform into e.g. scenarios or focus areas
5.	We have regression testing to secure basic stability and integrity in the system (before exploratory testing)
6a.	We have a structured way of working, e.g. planning meetings, preparations, test strategies
6b.	Our way of working (e.g. planning meetings, preparations, test strategies) provide sufficient freedom for the tester
7	Testers with different experiences and competences work together as a team, helping each other with new ideas and knowledge about different parts of the system
8	We have test environments that support debugging and recording
9.	We have a well-defined way to report the test results, including a description of areas covered by the tests and a list of identified problems

in the ExET model, are presented in Table 4. A simplistic approach was purposely selected in the design of the ExET model, based on our experiences and positive feedback received in the validation phases in our previous studies (Mårtensson et al., 2018, 2019b). The statements in the ExET model were designed to provide as much information as possible about the strengths and possible improvement areas in an organization. To a large extent, the statements in the ExET model are mirroring the key factors from Table 3 in Section 4.2. One of the factors (factor six) speaks about balance between freedom and structure, and is therefore split into two statements (6a and 6b) in order to isolate the root causes as much as possible.

The analysis of the first series of interviews in this study (presented in Section 4) showed that efficient and effective exploratory testing depends on many different things, spanning from recruitment and training of the testers to ways of working and the characteristics of the test environment. From the interviews, we also learned that companies struggle with different problems. For example, the testers' competence about the product could be a huge problem in one company, but this could be working well in other companies (which instead brought forward other problem areas). Here the ExET model plays a role, pointing out on the most important problem areas of each individual case.

6.2. How to use the model

The model is used in a simple two-step process (presented in Fig. 3) resulting in a list of improvement initiatives for the most prioritized problem areas in the organization.

Step 1 – ExET assessment: At the ExET assessment, the exploratory testers rate to which degree (on a Likert scale) each of the statements in the ExET model mirrors the situation in their organization. That is, the participants compare their current situation to the ideal situation (as shown in Fig. 2), and thereby identifies the strengths and potential improvement areas in the organization. The assessment can be conducted as a series of individual interviews, or a group interview (but still with individual responses from each tester). The participants are asked how each statement reflects the current situation in their organization, and are asked to respond on a Likert scale: *Strongly Agree, Agree, Somewhat Agree, Somewhat Disagree, Disagree* and *Strongly Disagree*.

Step 2 – ExET workshop: The purpose of the ExET workshop is to prioritize the most important problem areas and to identify

improvement initiatives. The workshop participants are representatives for the exploratory testers, stakeholders (e.g. a line manager) and enabling functions (e.g. responsible for the test environment). The results from the ExET assessment are visualized in the ExET model, and presented at the ExET workshop, in order to promote discussions. An example of a visualization in the ExET model is shown in Fig. 4. The responses from the interviewees from the ExET assessment are shown as bullets in the cells corresponding to the interviewees' responses (Strongly Agree, Agree, etc.). Identified strengths and improvement areas are emphasized, in the example in Fig. 4 with one green circle (emphasizing the testers' competence and personality as strengths) and two red circles (marking purpose and test environments as improvement areas). Selected quotes from the interviews are shown to the right, highlighting representative comments and statements from the exploratory testers. The values in Fig. 4 are fictitious, and are not related to any of the case study companies in the study.

The expected result from the ExET assessment and the ExET workshop is a list of prioritized improvement areas (which can enable efficient exploratory testing in the organization) and/or a clearly defined list of improvement initiatives. We recommend to schedule a new ExET workshop (e.g. six months after the first one) to follow up the improvement initiatives. If significant improvement have been done in the organization, we recommend to also schedule a new ExET assessment, as it is conceivable that new problems can be identified (not evident in the organization at the time of the first ExET assessment).

7. Validation of the ExET model

7.1. The five validation cases

In the validation of the ExET model, the model was used in five validation cases. In each case the ExET model was used according to the two-step process described in Section 6.2, including an ExET assessment and an ExET workshop. We will refer to the five validation cases as follows:

- Validation Case 1 (in Company D)
- Validation Case 2 (in Company D)
- Validation Case 3 (in Company E)
- Validation Case 4 (in Company E)
- Validation Case 5 (in Company F)

To provide external validation, the validation cases did not involve any of the individuals or organizations involved in the studied to develop the ExET model (described in Sections 4 and 5). Company D was involved in both the primary studies and the validation, but different parts of the organization were selected for the primary studies and the validation. In addition to this, the two validation cases in Company D (Case 1 and Case 2) were located in different cities and worked with different products in the company (an airborne product and a ground-based product). The same circumstances applied to Validation Case 3 and 4 in Company E (a platform system and a mobility system). Validation Case 5 in Company F covered the main product in the company. The organization in each validation case wanted not just to improve how they used exploratory testing, but also to increase the amount of exploratory testing in the organization.

The ExET assessments were done with individual interviews in four of the cases, and as a group interview in the fifth case in order to compare the different formats. The group interview format enabled discussions between the participants, but at the same time the individual interviews gave every participant the same opportunity to be heard. The interviewees had roles in their companies as tester, test leader, test architect or manager. They were all in different ways directly involved in the exploratory
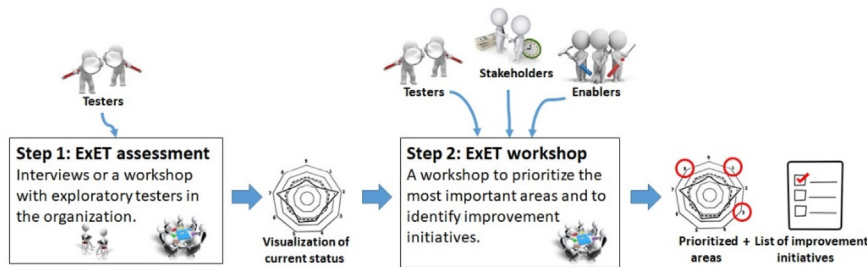
Fig. 3. Using the ExET model to identify a list of improvement initiatives to enable efficient and effective exploratory testing in the organization.

Fig. 4. Strengths and improvement areas visualized in the ExET model, showing how exploratory testing can be improved in the organization (example).

testing activities in their organization. The 36 interviewees had an average of 13.5 years of experience of software development in industry, spanning from 1 to 33 years. Fig. 5 presents the distribution of the interviewees regarding years of experience and the interviewee's role in their company. The number of participants at the ExET assessments in each validation case are summarized in Table 5 in Section 7.3.

The results from the ExET assessment in each of the cases were summarized and visualized in the ExET model (similar to the example in Fig. 4). All validation cases had generally more positive responses (Strongly Agree, Agree and Somewhat Agree) than negative responses. At the same time, responses in all companies pointed at improvement areas in each company. No pattern could be found in all companies with regards to identified problems, i.e. what was considered to be a major problem in one or several of the companies were found to be working well in other companies.

Each participant in the ExET assessments was asked how well the ExET Model succeeded in summarizing the status in their organization, and were asked to respond on the Likert scale from Strongly Agree to Strongly Disagree. All the participants responded Agree or Strongly Agree. The participants were also asked if they thought that there is something important that is not covered by the factors in the model, or if any of the nine factors do not belong in the list. Generally, the responses were that nothing needed to be changed. The other responses were found to be not related to the ExET model, e.g. "the testers should have more time for exploratory testing". There were also many positive, or even very positive comments on the ExET model. Comments such as "very good questions" or "the nine factors are spot on" showed that the factors in the ExET model were well received by the participants. The participants also found the results from the assessment to be valuable (e.g. "This was great, a good summary for management: to show them what needs to be done in order to make things work").

The ExET workshops were conducted in the same way in all the five validation cases. The visualization of the results from the ExET assessment were presented for the workshop participants, who all had roles as tester, test leader, test architect, test manager or line manager (for a group of testers). The number of participants at the ExET workshop in each validation case are summarized in Table 5. The workshop participants were generally confirming the strengths and problem areas presented in the visualization, and discussed root cases as well as solutions for the exposed problems. The workshop participants found that many problems could not be easily solved, especially issues related to the testers' competence and/or personality. The participants also discussed how the presented results applied to that their organization wanted to increase the amount of exploratory testing, e.g. how new testers should be onboarded.

The participants at each ExET workshop were asked to evaluate the workshop and the ExET model. The participants were also asked if they thought that there is something important that is not covered by the factors in the model, or if any of the nine factors not belong in the list. In all cases, the workshop participants found that the workshop had resulted in good discussions and ideas for improvement initiatives. The ExET model was well received, and the only improvement comments were related to how the model could be expanded to cover testing in general (not only exploratory testing). The two-step process to use the model was well received by the participants at the ExET workshops, expressed in comments such as "the method was very good" or "very much information for very little effort". Participants at three of the ExET workshops described how they should use the results from the ExET workshop, adding comments such as "I look forward to discussing this with management".

7.2. Validation cross-company workshop

As a complement to the validation of the ExET model in the five validation cases presented in Section 7.1, we facilitated a validation cross-company workshop including participants from all six companies in the study. The validation cross-company workshop included 29 participants: six from Company A, three from Company B, three from Company C, five from Company D, six from Company E, three from Company F, and three of the researchers. The participants had roles in their companies as
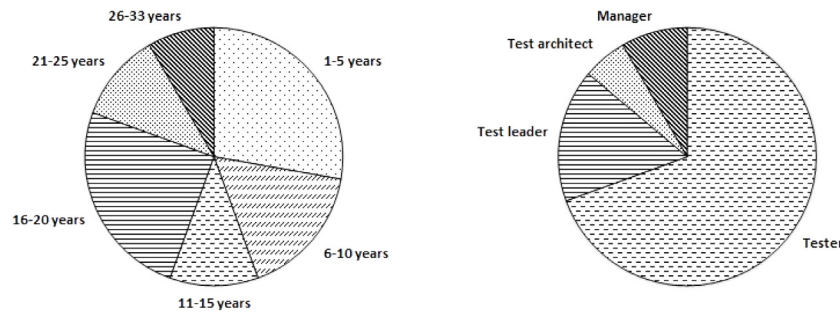
Fig. 5. Distribution of the 36 interviewees in the ExET assessments with regards to years of experience (left) and the interviewee's role in their company (right).

Table 5
Number of participants in each ExET assessment and ExET workshop.

Validation case	Interviewees at ExET assessments with individual interviews	Interviewees at ExET assessment with group interviews	Workshop participants at ExET workshop
Case 1 (Company D)	5	–	5
Case 2 (Company D)	–	10	6
Case 3 (Company E)	12	–	13
Case 4 (Company E)	4	–	6
Case 5 (Company F)	5	–	16
Summary	**36**		**46**

senior tester, test manager, test specialist or line manager (for a group of testers).

At the validation cross-company workshop, the researchers took turns in presenting the results from the study, from the studies to identify the key factors (presented in Sections 4 and 5) to the development of the ExET model (presented in Section 6) and the validation of the model (presented in Section 7.1). The presentations worked as a catalyst, promoting good discussions among the workshop participants.

Especially two topics generated longer discussions: The workshop participants discussed the fact that the ExET assessments in the five validation cases had emphasized different problem areas, without any clear common pattern (as described in Section 7.1). Some of the participants expressed that patterns will probably emerge based on the characteristics of the product, e.g. proximity of hardware. Large-scale was also brought forward as a root cause for several of the identified problems, e.g. one participant commented that "scale is a problem for many factors, especially factor 1-4". The workshop participants did not succeed in agreeing on a list of root causes or patterns, but seemed to agree that ways of working probably are more important than characteristics of the product when it comes to identifying such patterns. The workshop participants also had a longer discussion where they compared exploratory testing in a development team and exploratory testing of the complete product. Different opinions were expressed on the issue of biased testers. Some of the participants argued that exploratory testers in the development team always get biased, making them find less bugs in the product. Other participants at the workshop argued that the most effective and efficient testers are the ones who work close to the development teams.

The ExET model was well received also by the participants from Company A, Company B and Company C (who had not been involved in the validation of the model). Participants from two of the companies asked for ExET assessments to be scheduled in their organization, which we argue further strengthens that the model is considered to be novel, actionable and useful in practice.

7.3. Summary and analysis of the validation

In the validation of the ExET model, the model was used in five validation cases to identify what each company should prioritize in order to enable efficient and effective exploratory testing. In each case the ExET model was used according to the two-step process described in Section 6.2, including an ExET assessment and an ExET workshop. The number of participants in the ExET assessment and the ExET workshop in each validation case is summarized in Table 5.

In total, the ExET model was used in ExET assessments involving in total 36 exploratory testers, and ExET workshops involving in total 46 participants (as described in Section 7.1). The model was well received in all five validation cases. The participants at the ExET assessments found that the ExET model in a good way summarized the strengths and improvement areas in their organization. The ExET workshops resulted in good discussions and identified improvement initiatives. The participants at the ExET workshop described the results from the model as valuable and useful, expressed in comments such as "very much information for very little effort" or "I look forward to discuss this with management".

The results from the validation cross-company workshop (presented in Section 7.2) show that the ExET model was well received by all six companies in the study, which supports the generalizability of the model. The only suggested improvements from the participants in the validation were related to how the model could be expanded to cover testing in general (not only exploratory testing). We argue that this supports the completeness of the model.

In the literature review (presented in Section 3) we found no publication summarizing key factors that enable efficient and effective exploratory testing. Instead, published work tend to focus on one aspect, and are leaving out areas that other authors consider to be the core issues. Based on this, we find that although the different characteristics of the ExET model is mirrored in published literature, we have found no other publication that presents a model that can help a company to enable efficient and effective exploratory testing. We argue that this also supports the novelty of the ExET model.

In summary, we find that the validation of the ExET model has showed that the model is novel, actionable and useful in practice.

8. Threats to validity

8.1. Threats to construct validity

The fact that only one search engine was used for the literature review (reported in Section 3) could be seen as a threat to construct validity. However, the literature review was primarily used to motivate a continued research study, and not used as the primary source for the construction of the ExET model. In Section 7.3, we also use the results from the literature review to discuss the novelty of the results from the study. As the decision to use one search engine is clearly described in Section 3.1, we consider this threat to be mitigated.

Other threats to construct validity are related to the two series of interviews described in Sections 4 and 5: It is plausible that a different set of questions and a different context for the interviews can lead to a different focus in the interviewees' responses. In order to handle this threats against construct validity, the interview guides were designed with open-ended questions. It is conceivable that the interviewees' perception of the key factors for effective and efficient exploratory testing is affected by the current situation in the case study companies (e.g. which type of questions or topics that are currently in focus in each company). Therefore, it is plausible that the exact description of the key factors would have been different if the study had included other case study companies. However, all factors are based on comments and statements from at least 10 of the 20 interviewees. Due to this, we argue that this threat to construct validity can be viewed as acceptable.

In this paper, we also present background material for both the interviewees and the case study companies in order to provide as much information as possible about the context and enable reproducibility of the study.

8.2. Threats to internal validity

Of the 12 threats to internal validity listed by Cook et al. (1979), we consider Selection, Ambiguity about causal direction and Compensatory rivalry relevant to this work:

- *Selection*: All interviewees and workshop participants were purposively sampled (selected as good informants with appropriate roles in the companies) in line with the guidelines for qualitative data appropriateness given by Robson and McCartan (2016). Based on the rationale of these samplings and supported by Robson and McCartan who consider this type of sampling superior for this type of study in order to secure appropriateness, we consider this threat to be mitigated.
- *Ambiguity about causal direction*: While we in this study in some cases discuss relationships, we are very careful about making statements regarding causation. Statements that include cause and effect are collected from the interview results, and not introduced in the interpretation of the data.
- *Compensatory rivalry*: When performing interviews and comparing scores or performance, the threat of compensatory rivalry must always be considered. The questions in our interviews described in Sections 4 and 5 were deliberately designed to be value neutral for the participants, and not judging performance or skills of the interviewee or the interviewee's organization. Generally, the questions were also designed to be opened-ended to avoid any type of bias and ensure answers that were open and accurate. However, our experiences from previous work is that we found the interviewees more prone to self-criticism than to self-praise.

8.3. Threats to external validity

The list of key factors presented in Section 4 was confirmed by series of follow-up interviews and a cross-company workshop with participants from the same case study companies as the first series of interviews (as described in Section 5). Due to this, it is conceivable that the findings from this study are only valid for these companies, or for companies that operate in the same industry segments and have similar characteristics (presented in Section 2). The follow-up interviews and the cross-company workshop (described in Section 5) showed that the list of key factors is valid for all four case study companies. Because of the diverse nature of these four companies, the case study companies represent a good cross-section of the industry (as described in Section 2). Based on analytical generalization (Runeson and Höst, 2009; Wieringa and Daneva, 2015) it is reasonable to expect that the identified key factors that can enable efficient and effective exploratory testing are also relevant to a large segment of the software industry at large.

The validation of the ExET model (presented in Section 7) included interviews with 36 individuals and six workshops with in total 72 participants from the six companies participating in the study. The only suggested improvements from the participants in the validation were related to how the model could be expanded to also cover other types of testing. This supports the completeness of the model. The validation of the ExET model is primarily based on the external validation in the five validation cases, which were separated from the primary cases in the study. The validation showed that the model and the list of key factors is valid for all six companies in the study, which further strengthens the generalizability of the findings from this study. Obviously, we see further validation of the ExET model as an interesting area for future work, especially including companies in different industry segments.

9. Conclusion and further work

9.1. Conclusion

In this paper, we have presented interview results from two series of interviews with 20 interviewees from four case study companies, all developing large-scale software systems. The interviewees had generally good experiences from using exploratory testing. According to the interviewees, exploratory testing is a more creative way to work for the testers, and was therefore considered to make better use of the testers. Exploratory testing was also described as a good way to test system-wide and to test large-scale systems, especially exploratory testing with an end-user perspective.

In the analysis of the results from the two series of interviews (presented in Sections 4 and 5) we identified nine *key factors that enable efficient and effective exploratory testing*, grouped in four themes:

- *The testers' knowledge, experience and personality*:
  - The testers know how the system is built, and the correct behavior of the system
  - The testers know how the product is used by the end-user (or the end user is represented in the test team)
  - The testers are curious and want to learn about the system

● *Purpose and scope:*

    – A well-defined purpose and scope for the tests (system functions ready to be tested) which the testers can transform into e.g. scenarios or focus areas

    – Regression testing secure basic stability and integrity in the system (before exploratory testing)

● *Ways of working:*

    – An established way of working, including e.g. planning meetings, preparations, test strategies and heuristics (a balance between structure and freedom)

    – Testers with different experiences and competences work together as a team, helping each other with new ideas and knowledge about different parts of the system

● *Recording and reporting:*

    – Test environments that support debugging and recording

    – A well-defined way to report the test results, including a description of areas covered by the tests and a list of identified problems

Based on the results from the two series of interviews (in total 40 interviews) and a cross-company workshop with 14 participants, we developed *the ExET model (Excellence in Exploratory Testing)*. The ExET model (described in Section 6) allows companies to identify and visualize strengths and improvement areas, which can be used to optimize exploratory testing in each individual case. The model is used in a simple two-step process (the ExET assessment and the ExET workshop) resulting in a list of improvement initiatives for the most prioritized problem areas in the organization.

In the *validation of the ExET model*, the model was used in five validation cases in three case study companies (described in Section 7). In each case the ExET model was used according to the two-step process, including an ExET assessment and an ExET workshop. In total, the ExET model was used in ExET assessments involving in total 36 exploratory testers, and ExET workshops involving in total 46 participants. The results from the five validation cases were presented at a validation cross-company workshop with 26 participants from all six case study companies in the study, which further strengthened the generalizability of the model. The model was well received in all five validation cases, and was considered to be actionable and useful in practice.

The literature review (presented in Section 3) identified 129 publications related to exploratory testing, with 65 publications related to exploratory testing of software systems. No publications were found that presents a model that can help a company to enable efficient and effective exploratory testing, which supports the novelty of the ExET model.

In summary, we find that *the validation of the ExET model has showed that the model is novel, actionable and useful in practice*, showing companies what they should prioritize in order to enable efficient and effective exploratory testing in their organization. The ExET model is based on interview results from four case study companies and has been validated in five validation cases in three case study companies. As the six case study companies included in this study operate in different industry segments, it is reasonable to expect that the ExET model can be applied to a large segment of the software industry.

We believe that the ExET model will be valuable for both researchers and practitioners as it provides a systematic approach, based on the key factors that enable efficient and effective exploratory testing.

9.2. Further work

In addition to the results presented in the analysis and the conclusions, we believe that this study also opens up several interesting areas of future work. The validation of the ExET model in this study showed very promising results (as described in Section 7). Further validation should encourage more companies to use the model, including companies in different industry segments than the case study companies in this study. Further validation could also include other validation methods, e.g. using quantitative data.

The ExET model was well received in all five validation cases (as described in Section 7.1), and the only improvement comments were related to how the model could be expanded to cover testing in general (not only exploratory testing). A future study could expand the studies reported in Sections 4–5 to include a wider scope of test activities. If the study succeeds in isolating a list of key factors, a model could be constructed similar to the ExET model — but applicable also to other types of test activities than exploratory testing.

At the validation cross-company workshop, the workshop participants discussed the fact that the ExET assessment had emphasized different problem areas, without no clear common pattern (as described in Section 7.2). Another area of further work could be to further investigate root causes for the identified problem areas, defining patterns for how different industry segments or other types of characteristics are related to different types of problem areas.

The participants at the validation cross-company workshop also had a longer discussion where they compared exploratory testing in a development team and exploratory testing of the complete product (described in Section 7.2). A future study could continue to investigate pros and cons with exploratory testing on different levels, e.g. discussing the role of the independent tester as a member of a development team.

Another topic for further work is to analyze the relation (described in Section 5.2) between exploratory testing and Agile methodologies: is it so that "exploratory testing is enabling Agile" or is it "Agile is killing exploratory testing"?

CRediT authorship contribution statement

Torvald Mårtensson: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing - original draft, Visualization, Project administration. **Daniel Ståhl:** Conceptualization, Methodology, Validation, Writing - review & editing. **Antonio Martini:** Methodology, Investigation, Writing - review & editing, Visualization. **Jan Bosch:** Writing - review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Appendix. The interview guides

Appendix presents the questions in the interview guides for the two series of interviews which are reported in Sections 4 and 5 of this paper.

- Interview guide 1: Interviews to identify the key factors (described in Section 2.3)
- Interview guide 2: Interviews to confirm the key factors (described in Section 2.4)

Interview guide 1: the interview guide for the *interviews to identify the key factors that enable efficient and effective exploratory testing of large-scale software systems* included the following questions:

- Q1.1: How many years of industry experience with regards to software development do you have?
- Q1.2: How would you describe your individual role and responsibilities?
- Q1.3: How, and for which purpose do you currently use exploratory testing in your company?
- Q1.4: How do you on a scale from 1 (very low) to 5 (very high) value exploratory testing as a test technique for large-scale software systems?
- Q1.5: How would you describe the primary strengths and weaknesses of exploratory testing?
- Q1.6: What are the key factors that you think enable efficient and effective exploratory testing of large-scale software systems?
- Q1.7: Which of the things you have described in your response to Q6 are currently implemented in your company?

Interview guide 2: the interview guide for *the interviews to confirm that the interpretation of the first series of interviews was correct, as well as looking for negative cases* included the following questions:

- Q2.1: Is there anything that you would like to change in the presented key factors?
- Q2.2: How would you rate the importance of each of the key factors with regards to "efficient and effective exploratory testing"?

    – 5 (very high)
    – 4
    – 3
    – 2
    – 1 (very low)

References

Bach, J., 2020. Exploratory testing. https://www.satisfice.com/exploratory-testing (accessed July 18, 2020).

Basri, S., Dominic, D.D., Murugan, T., Almomani, M.A., 2019. A proposed framework using exploratory testing to improve software quality in sme's. Adv. Intell. Syst. Comput. 843, 1113–1122.

Bures, M., Frajtak, K., Ahmed, B.S., 2018. Tapir: Automation support of exploratory testing using model reconstruction of the system under test. IEEE Trans. Reliab. 67 (2), 557–580.

Calpur, M.C., Arca, S., Calpur, T.C., Yilmaz, C., 2017. Model dressing for automated exploratory testing. In: IEEE International Conference on Software Quality, Reliability and Security Companion, QRS-C 2017. pp. 577–578.

Cook, T.D., Campbell, D.T., Day, A., 1979. Quasi-Experimentation: Design & Analysis Issues for Field Settings. Houghton Mifflin Boston, p. 351.

Frajtak, K., Bures, M., Jelinek, I., 2016. Model-based testing and exploratory testing: Is synergy possible?. In: 6th International Conference on IT Convergence and Security, ICITCS 2016

Frajtak, K., Bures, M., Jelinek, I., 2017. Exploratory testing supported by automated reengineering of model of the system under test. Cluster Comput. 20 (1), 855–865.

Gebizli, C.S., Sozer, H., 2014. Improving models for model-based testing based on exploratory testing. In: 38th Annual International Computers, Software and Applications Conference Workshops, COMPSACW 2014. pp. 656–661.

Gebizli, C.Ş., Sözer, H., 2017a. Automated refinement of models for model-based testing using exploratory testing. Softw. Qual. J. 25 (3), 979–1005.

Gebizli, C.S., Sözer, H., 2017b. Impact of education and experience level on the effectiveness of exploratory testing: An industrial case study. In: 10th IEEE International Conference on Software Testing, Verification and Validation Workshops, ICSTW 2017. pp. 23–28.

Ghazi, A.N., Garigapati, R.P., Petersen, K., 2017. Checklists to support test charter design in exploratory testing. Lect. Notes Bus. Inform. Process. 283, 251–258.

Ghazi, A.N., Petersen, K., Bjarnason, E., Runeson, P., 2018. Levels of exploration in exploratory testing: From freestyle to fully scripted. IEEE Access 6, 26416–26423.

Gregory, J., Crispin, L., 2015. More Agile Testing. Addison Wesley.

Hellmann, T.D., Maurer, F., 2011. Rule-based exploratory testing of graphical user interfaces. In: 2011 Agile Conference. Agile 2011, pp. 107–116.

Hendrickson, E., 2013. Explore it!. In: The Pragmatic Bookshelf.

Hudson, J., Denzinger, J., 2015. Risk management for self-adapting self-organizing emergent multi-agent systems performing dynamic task fulfillment. Auton. Agents Multi-Agent Syst. 29 (5), 973–1022.

Humble, J., Farley, D., 2011. Continuous Delivery. Addison Wesley.

Itkonen, J., Mäntylä, .M.V., Lassenius, C., 2009. How do testers do it? An exploratory study on manual testing practices. In: 3rd International Symposium on Empirical Software Engineering and Measurement, ESEM 2009. pp. 494–497.

Itkonen, J., Mäntylä, .M.V., Lassenius, C., 2013. The role of the tester's knowledge in exploratory software testing. IEEE Trans. Softw. Eng. 39 (5), 707–724.

Kaner, C., 1988. Testing computer software. In: TAB Books.

Kaner, C., Bach, J., Pettichord, B., 2001. Lessons Learned in Software Testing. John Wiley & Sons.

Kim, D.-K., Lee, L.-S., 2014. Reverse engineering from exploratory testing to specification-based testing. Int. J. Softw. Eng. Appl. 8 (11), 197–208.

Kitchenham, B., 2004. Procedures for performing systematic reviews. Keele UK Keele Univ. 33 (2004), 1–26.

Kuhn, A., 2013. On extracting unit tests from interactive live programming sessions. In: International Conference on Software Engineering. pp. 1241–1244.

Kumar, S., Wallace, C., 2013. Guidance for exploratory testing through problem frames. In: Software Engineering Education Conference. pp. 284–288.

Mårtensson, T., Martini, A., Ståhl, D., Bosch, J., 2019a. Excellence in exploratory testing: Success factors in large-scale industry projects. In: 20th International Conference on Product-Focused Software Process Improvement, PROFES 2019. pp. 299–314.

Mårtensson, T., Ståhl, D., Bosch, J., 2017. Exploratory testing of large-scale systems – testing in the continuous integration and delivery pipeline. In: 18th International Conference on Product-Focused Software Process Improvement, PROFES 2017. pp. 368–384.

Mårtensson, T., Ståhl, D., Bosch, J., 2018. Enable more frequent integration of software in industry projects. J. Syst. Softw. 142, 223–236.

Mårtensson, T., Ståhl, D., Bosch, J., 2019b. Test activities in the continuous integration and delivery pipeline. J. Softw. Evol. Process 31 (4), e2153.

Micallef, M., Porter, C., Borg, A., 2016. Do exploratory testers need formal training? An investigation using HCI techniques. In: International Conference on Software Testing, Verification and Validation Workshops, ICSTW 2016. pp. 305–314.

Mihindukulasooriya, N., Rizzo, G., Troncy, R., Corcho, O., García-Castro, R., 2016. A two-fold quality assurance approach for dynamic knowledge bases: The 3cixty use case. In: CEUR Workshop Proceedings.

Pfahl, D., Yin, H., Mäntylä, .M.V., Münch, J., 2014. How is exploratory testing used? A state-of-the-practice survey. In: International Symposium on Empirical Software Engineering and Measurement.

Pichler, J., Ramler, R., 2008. How to test the intangible properties of graphical user interfaces?. In: 1st International Conference on Software Testing, Verification and Validation, ICST 2008 pp. 494–497.

Raappana, P., Saukkoriipi, S., Tervonen, I., Mäntylä, .M.V., 2016. The effect of team exploratory testing - experience report from F-secure. In: International Conference on Software Testing, Verification and Validation Workshops, ICSTW 2016. pp. 295–304.

Rashmi, N., Suma, V., 2014. Defect detection efficiency of the combined approach. In: Advances in Intelligent Systems and Computing 249 VOLUME II. pp. 485–490.

Reis, J., Mota, A., 2018. Aiding exploratory testing with pruned GUI models. Inform. Process. Lett. 133, 49–55.

Robson, C., McCartan, K., 2016. Real World Research, fourth ed. Wiley.

Runeson, P., Höst, M., 2009. Guidelines for conducting and reporting case study research in software engineering. Empir. Softw. Eng. 14 (2), 131–164.

Schaefer, C.J., Do, H., 2014. Model-based exploratory testing: A controlled experiment. In: 7th International Conference on Software Testing, Verification and Validation Workshops, ICSTW 2014. pp. 284–293.

Shah, S.M.A., Gencel, C., Alvi, U.S., Petersen, K., 2014a. Towards a hybrid testing process unifying exploratory testing and scripted testing. J. Softw. Evol. Process 26 (2), 220–250.

Shah, S.M.A., Torchiano, M., Vetrò, A., Morisio, M., 2014b. Exploratory testing as a source of technical debt. IT Prof. 16 (3), 44–51.

Shoaib, L, Nadeem, A., Akbar, A., 2009. An empirical evaluation of the influence of human personality on exploratory software testing. In: 13th International Multitopic Conference, INMIC 2009.

Ståhl, D., Bosch, J., 2014. Modeling continuous integration practice differences in industry software development. J. Syst. Softw. 87, 48–59.

Ståhl, D., Bosch, J., 2016. Industry application of continuous integrationmodeling: a multiple-case study. In: Proceedings of the 38th International Conference on Software Engineering Companion. ACM, pp. 270–279.

Ståhl, D., Hallén, K., Bosch, J., 2016. Achieving traceability in large scale continuous integration and delivery: Deployment, usage and validation of the eiffel framework. Empir. Softw. Eng. 1–29.

Ståhl, D., Mårtensson, T., Bosch, J., Continuous practices and DevOps: Beyond the buzz, what does it all mean? In: Proceedings - 43rd Euromicro Conference on Software Engineering and Advanced Applications, SEAA 2017, 8051321, 2917, pp. 440–448.

Sviridova, T., Stakhova, D., Marikutsa, U., 2013. Exploratory testing: Management solution. In: 12th International Conference: The Experience of Designing and Application of CAD Systems in Microelectronics, CADSM 2013. p. 361.

Tuomikoski, J., Tervonen, I., 2009. Absorbing Software Testing Into the Scrum Method. In: Lecture Notes in Business Information Processing 32 LNBIP, pp. 199–215.

Whittaker, J., 2010. Exploratory Software Testing. Addison Wesley.

Wieringa, R., Daneva, M., 2015. Six strategies for generalizing software engineering theories. Sci. Comput. Program. 101, 136–152.

Torvald Mårtensson, is Distinguished Engineer in System Integration and Test, and Technical Fellow at Saab AB. He has a background of fifteen years in the aeronautics industry and another eight years in the telecom industry. His work has primarily revolved around systems integration and system testing of large-scale software systems, and he has published several research papers on the subject at conferences and in journals. He received an M.Sc. degree from Linköping University, Sweden, and a Ph.D. degree from the University of Groningen, The Netherlands.

Daniel Ståhl is a software subject matter expert and development leader at Ericsson AB, where he is responsible for the corporate strategy on continuous integration, delivery and deployment. He has been working professionally in software development since 2007, and has spent most of that time implementing and researching continuous practices in organizations ranging from single teams to tens of thousands of engineers. Daniel is also Associate Professor of Software Engineering at Linköping University since 2018. He received his M.Sc. degree from Linköping University, Sweden, in 2007 and his Ph.D. degree from the University of Groningen, The Netherlands, in 2017.

Antonio Martini is Associate Professor at the University of Oslo and is a part-time researcher at Chalmers University of Technology. He has been employed as a Postdoctoral Researchers at Chalmers, after having obtained a Ph.D. in Software Engineering at Chalmers University of Technology, Sweden, in 2015. He has worked as Principal Strategic Researcher at CA Technologies for a co-financed project for technology transfer related to Technical Debt and Architecture by the H2020 Marie Skłodowska-Curie grant of the European Union. He has collaborated as researcher and consultant with several large software companies.

Jan Bosch, is Professor of software engineering and Director of the software research center at Chalmers University Technology in Gothenburg, Sweden. Earlier, he has worked as Vice President Engineering Process at Intuit Inc and as head of the Software and Application Technologies Laboratory at Nokia Research Center, Finland. Before joining Nokia, he headed the software engineering research group at the University of Groningen, The Netherlands He received a M.Sc. degree from the University of Twente, The Netherlands, and a Ph.D. degree from Lund University, Sweden.