Perspective Based Risk Analysis - A Controlled Experiment

Sardar Muhammad Sulaman
Department of Comp. Science
Lund University, Sweden.
sardar@cs.lth.se

Krzysztof Wnuk
Department of Comp. Science
Lund University, Sweden.
Krzysztof.Wnuk@cs.lth.se

Martin Höst
Department of Comp. Science
Lund University, Sweden.
martin.host@cs.lth.se

ABSTRACT

Context: The increasing dependence on critical IT systems makes them more and more complex, which results in increased complexity and size. Risk analysis is an important activity for the development and operation of critical IT systems, but the increased complexity and size put additional requirements on the effectiveness of risk analysis methods. There complexity means that there is a need to involve different perspectives into risk analysis. **Objective:** The objective of the research carried out in this study is to investigate the effectiveness of perspective-based risk analysis (PBRA) methods compared to traditional risk analysis (TRA) methods. **Method:** A controlled experiment was designed and carried out. 43 subjects performed risk analysis of a software-controlled train door system using either TRA or PBRA. **Results:** The results suggest that PBRA helps to identify more relevant risks than TRA. On the other hand, our experiment failed to provide supporting evidence that PBRA helps to identify fewer non-relevant risks. This study also found that PBRA is more difficult to use than TRA. **Conclusions:** Some potential benefits of using perspective-based risk analysis are uncovered and experimentally confirmed. In particular, it was discovered that PBRA is more effective than the traditional method and identifies more relevant risks.

Categories and Subject Descriptors

D.2.4 [**Software Engineering**]: Software/Program Verification—*Reliability, Statistical methods*; K.6.4 [**Management of Computing and Information Systems**]: System Management—*Quality Assurance*

General Terms

Reliability, Verification

Keywords

Experiment, Risk analysis, Perspective-based Risk Analysis, Risk Management

1. INTRODUCTION

The increasing complexity of socio-technical IT systems and our dependence on them put additional pressure on the effectiveness of risk analysis methods. More complex IT systems contain more interacting components and sub-systems, which in turn increase the probability of serious failures [1]. Moreover, failures in these complex safety-critical systems are often results of multiple interacting decisions and errors [2].

The complexity, size, and heterogeneity of today's IT systems call for involving different perspectives into risk and hazard analyses. Several authors, e.g., Leveson [2] and Ieracе [3] recognized the benefits from multiple views analysis and encouraged adding internal and external organizational perspectives into the hazard analysis teams. Yoran and Hoffman proposed defining roles and identifying actors before performing risk analysis in order to improve the process [4]. Morevoer, involving different perspectives is also recommended by several risk analysis standards and methods [5, 6, 7, 8].

Perspective-based reading was successfully used for reviews and inspections during software projects, e.g. [9]. However, the potential benefits of involving perspectives into risks analysis have, to our knowledge, not been explored in an experimental way. It can also be observed that only one study listed in a survey about controlled experiments in software engineering was classified as software and system safety [10], which also indicates the need for experimentation in the area.

In this paper, we report the results from an experiment designed to investigate if Perspective-Based Risk Analysis (PBRA) that involves different views and perspectives is more effective and offers higher confidence than Traditional Risks Analysis (TRA). 43 subjects performed risks analysis of a software-controlled train door system using either TRA or PBRA. The effectiveness of the methods is measured by counting the number of relevant and non-relevant risks. A questionnaire was used to assess the difficulty of the methods and the confidence of the subjects concerning the correctness of the identified risks.

This paper is structured as follows: Section 2 provides related work while Section 3 outlines the experimental design. Section 4 describes the execution of the experiment and Section 5 provides the experimental results. Section 6 analyzes the results and Section 7 discusses the validity threats. Section 8 presents the discussion. Finally, the paper is con-

EASE '14, May 13 - 14 2014, London, England, BC, United Kingdom
Copyright 2014 ACM 978-1-4503-2476-2/14/05 ...$15.00
http://dx.doi.org/10.1145/2601248.2601260.

cluded in Section 9.

2. RELATED WORK

There exist a number of risk analysis methods for technical systems in general or for IT-systems in particular, e.g. [5, 6, 7, 8] just to name a few. The Risk Management guidelines for Information Technology Systems [5] highlight that management, CIOs, system owners, business managers and security program managers should be involved into the risk management process. The OCTANE method for risk-based information security assessment also advocates involving business and IT perspectives into the risk analysis processes [7]. The NetRAM method for network security analysis is also adapted for different enterprise structures on different levels and therefore can also involve the business perspective [8].

Some of the most well-known low-level risk analysis methods are Fault Tree Analysis (FTA) [11], Failure Mode and Effect Analysis (FMEA) [1] and Hazard and operability study (HAZOP) [3]. These methods have successfully been used for decades for technical and IT systems. However, these traditional methods do not consider the use of perspectives.

The recent advances in risk analysis methods or techniques include an actuator-based approach that identifies failures in four different severities [12] and the System Theoretic Process Analysis (STPA) method proposed by Leveson that considers safety as a control problem rather than a component failure problem [2]. STPA was applied to various systems with positive outcomes [2, 13, 14].

The idea of using perspectives is not new. Perspectives were utilized for reading software engineering artifacts with the purpose of improved defect identification [15, 9]. Perspective-based reading was also applied for object oriented design inspections [16], code reviews [17] and usability inspections [18]. Different perspectives, e.g. developers, testers and domain experts are often involved in requirements elicitation. This results in increased quality of elicited requirements and often uncovers new requirements based on various views and perspectives.

Yoran and Hoffman proposed the Role-Based Risk Analysis (RBRA) method that defines roles and identifies actors before performing risks analysis activities in order to reduce the set of vulnerabilities and controls to those appropriate to a given role [4]. RBRA was presented on an illustrative example from the computer software engineering domain but not experimentally investigated. Leveson [2] and Ierace [3] advocated to involve various perspectives during risk analysis, also from external organizations. It is always recommended, in almost all risk analysis methods, to have experts with domain knowledge while performing risk analysis but to our knowledge no one has proposed the use of specific perspectives for risk analysis. In this study we have used specific perspectives for the performed risk analysis. To summarize, the potential of perspectives in risk analysis was not yet experimentally assessed.

3. EXPERIMENTAL DESIGN

In this study, the research is carried out through a controlled experiment based on the guidelines presented by Wohlin et al. [19] and reported based on the reporting guidelines presented by Jedlitschka et al. [20][1]. This research is carried out in the following steps as shown in Figure 1.

1. Experiment design
2. Risk analysis of the selected object for this experiment. This resulted in a first set of "correct risks".
3. Pilot study
4. Presentation about risk analysis to the subjects at a lecture
5. Experiment execution
6. Presentation of results to the subjects

3.1 Research questions

The objective of the research carried out in this study is to investigate the effectiveness of the PBRA method in comparison with the TRA method. Here, effectiveness means a large number of relevant risks and a small number of non-relevant risks. This general objective is broken down to the following research questions:

- RQ1: Which risk analysis method is more effective?
- RQ2: Which risk analysis method is more difficult to use?
- RQ3: How confident are the participants about the risks they find using the studied methods?

RQ1 is important to investigate since a general goal of any risk analysis method is to find as complete set of risks as possible [21] and to minimize the number of non-relevant risks. RQ2 is relevant to investigate since the successful introduction of any method is dependent on that it is not seen as too hard to use by the users. Moreover, if the users do not feel confident (RQ3) with the results of the proposed method, they will be reluctant to apply the method in the real safety-critical systems.

3.2 Variables and hypothesis

The following independent and dependent variables are used in this experiment. The independent variable is the used risk analysis (RA) method. Two methods are compared in this experiment:

- Traditional risk analysis
- Perspective-based risk analysis

The dependent variables for this experiment are:

- N_r: Number of relevant risks found
- N_{nr}: Number of non-relevant risks found
- D: Difficulty level while using risk analysis method. The difficulty is measured on a Likert scale with five possible values, from *very easy* (1) to *very difficult* (5).
- C: Confidence level of the participants about found risks. The confidence level is measured on a Likert scale with five possible values, from *Very Confident* (1) to *Strongly not confident* (5).

[1] The experimental package including all the guidelines and results is available at `http://serg.cs.lth.se/index.php?id=87041`

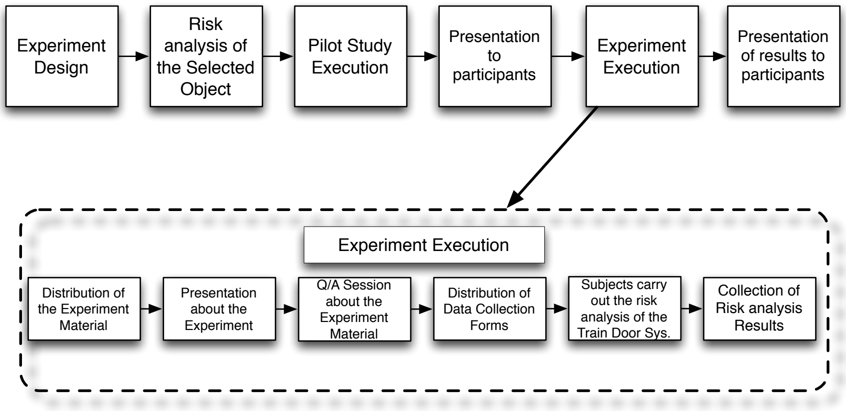
Experiment Design
Risk analysis of the Selected Object
Pilot Study Execution
Presentation to participants
Experiment Execution
Presentation of results to participants
Experiment Execution
Distribution of the Experiment Material
Presentation about the Experiment
Q/A Session about the Experiment Material
Distribution of Data Collection Forms
Subjects carry out the risk analysis of the Train Door Sys.
Collection of Risk analysis Results

Figure 1: Carried out steps for the experiment

The values of the dependent variables, N_r and N_{nr}, are calculated based on the identified relevant and non-relevant risks. The confidence and difficulty levels are determined using a questionnaire. The statistical analysis was performed to accept or reject the hypotheses H_0^1, H_1^1 and H_2^1.

RQ1 is broken down into two null hypotheses, detailed below. The first null hypothesis is that both risk analysis methods, PBRA and TRA, find the same numbers of relevant risks.

- H_0^1: The mean of PBRA and TRA is equal that both found same number of relevant risks (N_r).

The alternative hypothesis is:

- H_1^1: The mean of PBRA and TRA is not equal that both found different number of relevant risks (N_r).

The second null hypothesis for the RQ1 is that both risk analysis methods, PBRA and TRA, find the same numbers of non-relevant risks.

- H_0^2: The mean of PBRA and TRA is equal that both found same number of non-relevant risks (N_{nr}).

The alternative hypothesis is:

- H_1^2: The mean of PBRA and TRA is not equal that both found different number of non-relevant risks (N_{nr}).

The null hypothesis for the RQ2 is that both risk analysis methods, PBRA and TRA, are equally difficult to use.

- H_0^3: Both PBRA and TRA methods have same median that is same difficulty level to use (D).

The alternative hypothesis is:

- H_1^3: TRA method has lower median that it is less difficult to use (D).

The null hypothesis for the RQ3 is that the participants of both methods are equally confident about the identified.

- H_0^4: The median for both methods, PBRA and TRA, is same. i.e. the participants of both treatments are equally confident (C).

The alternative hypothesis is:

- H_1^4: TRA method has small value of median that means the participants that used TRA are less confident (C).

3.3 Subjects

The sample included participants of a project course in software development at Lund University, offered in autumn 2013[2]. The course is an optional advanced-level Masters' course for students from several engineering programs, e.g., Computer Science, Electrical Engineering, Civil Engineering, and Information and Communication Technology. The course gives 7.5 ETCS points that corresponds to five weeks full-time study. This experiment was a non-mandatory part of the course. 43 out of the total 70 students took part in the experiment. The participants were instructed clearly that the results of this experiment were completely anonymous and do not have any effect on the final grade of the course. It was also explained that results of the experiment will be used for research, and if they do not want to participate in the research then they are not required to submit their results.

3.4 Objects

The objects used a software-controlled insulin pump as an example in the guidelines, and a software-controlled train door system during the experiment. Both systems represent embedded socio-technical safety-critical systems.

3.4.1 Train door system

The train door system (see Figure 2) was selected for the experiment because of the following reasons: (1) it is a simple system and it has fewer components than the insulin pump, (2) it is highly possible that almost every participant has used this kind of system, (3) the system is rather simple and should be easy and quick to understand and (4) the participants should be able to find many risks for this system. The automated train door system has four main components, shown in Figure 2, the door sensor, door controller, door actuator and the physical door.

The door sensor sends a signal about the door position and the status of the doorway (if the doorways is clear or not) to the door controller. Then, the door controller receives input from the door sensor with some other inputs from the external sensors about the motion and the position of the train. It also gets an indication about possible

[2] http://cs.lth.se/kurs/etsn05-programvaruutveckling-foer-stora-system/

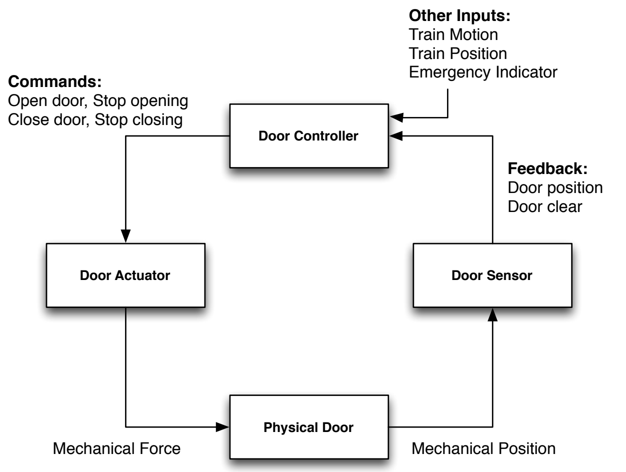

Figure 2: Functional diagram of a Train Door System [14]

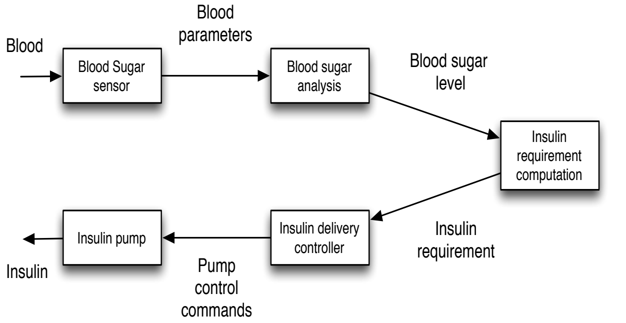

Figure 3: Functional diagram of an Insulin Pump [22]

emergencies from an external sensor. After receiving inputs, the controller performs some computation and then it issues door open and close commands as shown in Figure 2. After this, the door actuator receives commands from the controller and it applies mechanical force on the physical door. Finally, there is a physical door in the system that is closed and opened by the door actuator.

3.4.2 Insulin pump

The software-controlled insulin delivery system (see Figure 3) provides automated insulin delivery by monitoring blood sugar levels. The insulin pump is a portable device that delivers insulin via a needle attached to the body. It was selected to be an example system in the experiment guidelines because it is a representative example of a small and simple safety-critical system. Moreover, it has already been used for risk analysis [22].

3.5 Treatments

3.5.1 Traditional Risk Analysis (TRA)

The TRA method is an iterative activity and it consists of the four following steps [23]:

1. **Planning**: In this step, after forming groups all group members carefully read the system description individually and then decide who will be the moderator and who will be the scribe for the group.

2. **Risk identification**: This step determines a list of possible risks. It is an iterative activity that is normally carried out by brainstorming. In this step every risk analyst in the group attempts to find an individual list of possible risks by answering the following question: What could happen or what can go wrong?

   After performing individual analysis, all analysts in the group compare and merge their individually identified risks with others and make a common risk list. During this process new risks can also be identified.

3. **Determine likelihood**: Step 3 determines the likelihood of occurrence of all identified risks from step 2 by using the qualitative descriptors, i.e., highly unlikely, unlikely, possible, likely, very likely.

4. **Determine consequence**: Step 4 determines the consequences (severity level) of all identified risks from step 2 by using the qualitative descriptors, i.e., insignificant, minor, moderate, major, catastrophic.

3.5.2 Perspective-Based risk analysis

The PBRA method is also an iterative activity that supports the risk analysts to view and analyze the system from different perspectives. For example, one analyst may analyze the system from the point of view of the designer, another from the point of view of the developer, and another from the point of view of the user/client of the system. We believe that by using different perspectives risk analysts can perform a better and more in-depth analysis by thinking about different safety and security requirements. The used guidelines for the both treatments were the same; there was no extra information for the PBRA participants except the used perspectives.

PBRA consists of four steps just like TRA, but in step 1, the planning step, every member of the risk analysis team (group) is assigned one perspective for the identification step (this is similar to the approach suggested by Yoran and Hoffman [4]). The other steps of PBRA are the same as in TRA.

The selection of perspectives can be done by the participants in the groups, or can be assigned to the group before they start, in this case by the experimenter. In this experiment, during the pilot study the experimenter assigned the specific perspectives to the participants according to their experience. In the experiment execution with subjects, the perspectives were selected by the participants themselves according to their own choice.

In this experiment, PBRA was performed from the following three perspectives for the train door system.

- System Engineer (SE)
- Tester (T)
- Train Staff Member (TS)

The participants were informed that it is possible and even likely that several of the identified risks from the different perspectives are the same.

3.6 Instrumentation

The detailed guidelines were written in an understandable language and reviewed by the authors to execute the experiment effectively. Minor changes were introduced in the guidelines for the PBRA method about the use of different perspectives, in PBRA the participants have to use different perspectives unlike TRA[3].

The first section of the guidelines is about motivation to perform the experiment and the risk analysis. The main motivation for the subjects to participate in the experiment was to use the gained knowledge and experience from the experiment in their own course projects since risks analysis was a mandatory part of the course.

Then, the guidelines present the risk analysis method in detail with step-by-step instructions to perform it. The guidelines also present one example system (insulin pump) with some of the identified risks to give a solid idea about the risk analysis process to the participants. The guidelines also present qualitative descriptors for the likelihood of occurrence and the consequence levels with their definitions. The example presented in the guidelines shows all steps of risk analysis for the example system (insulin pump) with likelihood and consequences and example risks.

The description of the system (train door system), selected for the experiment, was appended in the appendix of the guidelines. The system description contains the technical details of the system and shows the boundaries of the system and the system context. For risk analysis, defining the boundaries (scoping) of the system being analyzed including all dependencies between components is very important otherwise risk analysts could easily become confused or could find many non-relevant risks. The system context, i.e., where the system is used, how and by whom, is also very crucial for the risk analysis. To perform an effective and efficient risk analysis the risk analysts should have clear understanding of the system context [24].

A post-experiment questionnaire was designed to measure the understanding of the guidelines, system description, and prior experience of risk analysis process. It contains 8 questions in total, where 6 of them are quantitative and 2 are qualitative.[4]

Two different data collection forms were designed to be used by the participants, one for each risk analysis method. The participants were asked to write identified risks on the provided data collection forms. The example presented in the guidelines used the same data collection forms. The motivation behind this was to give good understanding of the risk analysis process to the participants.

For data collection a complete set of risks was needed to decide which risks, identified by the participants, are relevant and non-relevant. The set of risks was incrementally developed in several phases. The first author performed the initial analysis and identified 28 risks. Then, one independent researcher working in the software safety domain evaluated the list. After evaluation and discussion 13 more risks were added. During the pilot study, 23 additional new risks were identified and added to the list. As a result, the final risk list contains 64 risks. The participants of the experiment found 5 new risks during the experiment execution that were not identified by the experimenters before. After adding these 5 new risks in the risk list the total identified risks became 69.

3.7 Pilot study/experiment

After preparing the instrumentation a pilot experiment was carried out on 13:th September 2013. The pilot study was carried out to evaluate the instrumentation of the experiment. Therefore, the results of pilot study are not used in the analysis of the experiment.

The sample contained 9 participants, where 5 participants were from the IT industry with 1.5–5 years of experience in software testing and development. The four other participants were researchers; one was PhD in biology and the other three were PhD students in computer science and electrical engineering.

Since there were 9 participants, they formed three groups each with three members. Each group was in the separate room when they performed the risk analysis for the pilot experiment. Two groups performed risk analysis by using PBRA and one group by using the TRA method.

The pilot experiment was carried out by following same steps for the main experiment mentioned in Section 4. The participants of the pilot study were asked to give feedback verbally after the experiment. After this, the feedback was noted down by the experimenter for the later analysis of instrumentation.

The participants of the pilot study mentioned the following problems or ambiguities in the guidelines, and system description.

- The example system and the experiment system are not clearly distinguished in the guidelines.
- Some information was missing in the given presentation for the experiment, e.g., example about one risk having multiple causes and other way around.
- In the functional diagram of the train door system there is one ambiguous input (*control command*) and one ambiguous output (*status*).
- *Other inputs* mentioned in the functional diagram of the train door system are un-clear.
- The detail of mentioned *emergency indicator* in the functional diagram of the train door system is missing.

Based on the identified problems and suggestions from the participants in the pilot study, changes were made to the instrumentation for the main experiment. The guidelines were improved by explaining the differences between the both example (insulin) and experiment (train door) systems. For the missing information in the presentation, it was decided that the example risks should be explained in the main experiment presentation clearly. The functional diagram was improved by removing the mentioned ambiguous input and output (*control command* and *status*). Here, the problem was unclear system boundaries because the mentioned ambiguous input and output were connected with some external systems. There were three *other inputs* (train motion, position and emergency indicator) to the train door systems that were not clearly explained in the system description. This problem was fixed by adding explanation for each input with headings (clearly visible).

[3] The guidelines can be accessed at `http://serg.cs.lth.se/index.php?id=87041`

[4] The questionnaire can be accessed at `http://serg.cs.lth.se/fileadmin/serg/Questionnaire.pdf`

Table 1: Results from the Pilot Study

Group label	Applied treatment	# of relevant risks found	# of non-relevant risks found
G1	PBRA	26	1
G2	PBRA	14	0
G3	TRA	11	0

Table 2: Summary of groups at seminars

Seminar	# of Subjects		# of Groups
	PBRA	TRA	
I	21	-	7
II	-	$4 - 1 = 3$	1
III	-	18	6
Sum	21	21	14

3.7.1 *Results from the pilot study*

Table 1 shows the results of the pilot study. There were three groups in the pilot study. Two groups (G1 and G2) used PBRA and one group (G3) used TRA. It can be noticed that the number of identified risks of G1 are significantly higher than for the other two groups. This may be because of differences in experience of participants. Group G1 had one member with 5 years of experience working as a system tester and a second member was a PhD in biology. The experience of the participants in G2 and G3 was almost same (1.5-2 years) and there is not a significant difference in the number of found risks between them. However, more risks were found with PBRA than TRA.

3.8 Data collection procedure

The data collection procedure was kept same for both the pilot experiment and main experiment. The subjects were given a presentation including the motivation for the experiment, explanation of the risk analysis method to be used (TRA or PBRA) with an example. The system that they should work with (train door) was also described.

Then, there was a short answer/question session about the guidelines, system description etc. Then, each group was asked to perform risk analysis. All members of each group performed an individual risk analysis as mentioned in the instrumentation. After this, each group was asked to compare and merge the individual risk lists to come up with a common group risk list.

Data collection forms, designed by the experimenter, were distributed among the participants for writing the identified risks during the risk analysis. After completion of the risk analysis, data collection forms were collected group by group. Then, all the participants were given a post-experiment questionnaire. The results from the experiment were collected by analyzing the information written in the data collection forms and then these results were also checked against the post-experiment questionnaire.

Each group was assigned a label and asked to write that on the data collection forms. Group labels were used to know that both data collection forms and post-experiment questionnaires are from one specific group, which was required for the analysis. The participants of the experiment were completely anonymous.

4. EXPERIMENT EXECUTION

There were total 43 participants of the experiment, see Section 3.3. These participants were divided into 14 groups (7 groups for each treatment) with 3 members in each as shown in Table 2.

Three course seminars were assigned for this experiment. The first seminar was on 9:th September 2013, at 15-17, the second on 10:the September at 8-10, and the third was also on 10:th September at 10-12. It was decided to perform the experiment with the only treatment PBRA in the first seminar, and with the treatment TRA in the second seminar. The third seminar was allocated to balance the number of groups for both treatments.

21 students attended the first seminar, forming 7 groups, and they all participated in the experiment with the PBRA treatment. 4 students attended the second seminar and used the TRA treatment by forming 1 group of three members. The remaining one student was not part of the experiment. In the third seminar, 18 subjects participated. All attendees of the third seminar used the TRA method and formed 6 groups, which balanced the experiment so that there were equally many groups for both treatments.

This experiment was carried out by following steps.

1. The experiment guidelines were distributed among the participants.

2. The participants were given a brief (10 minutes) presentation about the experiment task. This included an explanation of the risk analysis method, the example presented in the guidelines and the system description. Some examples of relevant and non-relevant risks about the example system were also presented in order to show concrete examples of what type of risks that can be identified, and on what level of abstraction the risks can be formulated on.

3. There was a short (5 minutes) session with questions and answers about the guidelines, system description, etc. The participants were given a chance to ask immediate questions that they had after reading the guidelines, but they were also allowed to ask questions during the later sessions.

4. The data collection forms were distributed for writing the risks found in the system during the risk analysis.

5. Each group was asked to perform the risk analyses.

   (a) 10 minutes were given for the planning step of the risk analysis. It was possible to have as short time as this since the participants had already read the system description.

   (b) 35 minutes were given to perform the remaining steps (risk identification, determine the likelihood level and the consequence level) of the risk analysis. During this time, every member of a group performed individual risk analysis.

6. Each group was given 20 minutes to compare and merge the individual risk lists to come up with a common group risk list.

7. After the collection of data forms the post-experiment questionnaire was given to all the participants.

5. RESULTS

Table 3 shows the experiment results carried out in seminars I, II and III. In seminar I, the PBRA treatment was used by 7 groups (21 participants). It can be seen that group M2 found the highest number of relevant risks, 19, and group M5 found the lowest number of relevant risks, 8. Group T7 found the highest number of non-relevant risks, 5, and M1 found 3. Groups M3, M5 and M7 found 1 non-relevant risk each. The remaining two groups (M2 and M4) found only relevant links.

In seminar II and III, TRA was carried out by the 7 groups. It can be seen that group T4 and T5 found most relevant risks, 11, and group T1 found least relevant risks, 7. Group T7 found most non-relevant risks, 5, and T3 found 1. All other groups did not find any non-relevant risk. The remaining five groups found only relevant risks.

As described in section 3.6, the experiment participants identified 5 new risks that were not present in the risk list identified by the experimenters. These new identified risks were also added in the risk list.

Table 3: The results from the main Experiment

Seminar	Group label	Applied treatment	# of relevant risks found	# of non-relevant risks found
I	M1	PBRA	14	3
I	M2	PBRA	19	0
I	M3	PBRA	13	1
I	M4	PBRA	14	0
I	M5	PBRA	8	1
I	M6	PBRA	10	2
I	M7	PBRA	14	1
II	T1	TRA	7	0
III	T2	TRA	9	0
III	T3	TRA	10	1
III	T4	TRA	11	0
III	T5	TRA	11	0
III	T6	TRA	10	0
III	T7	TRA	9	5

6. ANALYSIS

The data collected from the experiment (the number of found relevant risks) was analyzed for normality. Figure 4 shows the normal distribution plot for both datasets (results of PBRA and TRA). The line on the left is from the TRA dataset, the data points are quite clearly forming a straight line. However, the line of PBRA dataset, on the right, does not look clearly straight. Since the datasets are rather small, it was decided to use the Shapiro-Wilk normal distribution test. It is used to test the null hypothesis that data comes from a normally distributed population. The null hypothesis was not rejected with the p-values 0.306 for TRA and 0.505 for PBRA. Both datasets proved to be normally distributed by using the Shapiro-Wilk normality test, which is one of the most powerful normality tests [25].

After testing the datasets for the normality, the T-test was performed to check for statistically significant difference between the efficiency of the TRA and PBRA methods measured by the number of identified relevant risks (research question RQ1). The T-test was applied to investigate the null hypothesis that the data from the two methods are normally distributed with equal means and equal but unknown variance, against the alternative that they are not. It revealed a statistical significant difference between TRA and PBRA methods by rejecting the null hypothesis H_0^1 with the p-value 0.027. As a result, we could accept the alternative hypothesis H_1^1 that the subjects found more relevant risks using the PBRA method.

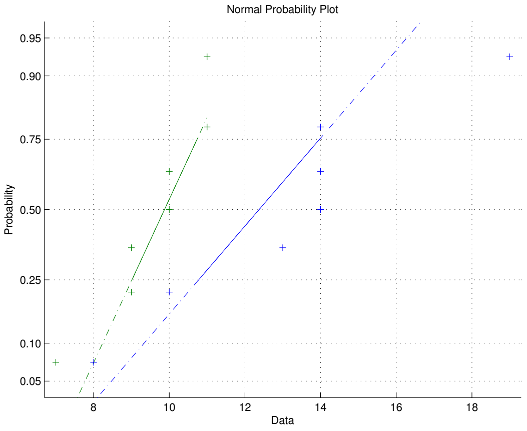

Figure 4: Normal distribution plot for the data

The box plots for the number of found risks are shown in Figure 5. It can be seen that there is a difference in the number of found risks by TRA and PBRA methods. The participants that used TRA method found on average 9.57 relevant risks and the participants that used PBRA found 13.14 relevant risks.

To answer the second hypothesis regarding research question RQ1, the number of identified non-relevant risks with both treatments was first analyzed for normality. The *Shapiro-Wilk* normal distribution test was used to test the null hypothesis that data comes from a normally distributed population. The null hypothesis was rejected for TRA dataset with the p-value 0.0014 and it was not rejected for the PBRA dataset with the p-value 0.587. Thus, it was decided to use non-parametric tests.

The results were tested for the statistical difference using the *Mann-Whitney U*-test. No statistically significant difference was revealed by the test resulting in the p-value of 0.249. Therefore, we cannot state that the PBRA method helped to identify fewer non-relevant risks.

For the research question RQ2 and RQ3, the following two questions were asked using an ordinal scale in the post-experiment questionnaire[5] respectively.

1. How difficult was the risk analysis method to use? (RQ2)

2. How confident are you that you have found all the relevant risks? (RQ3)

[5] Due to space limitations we do not present complete survey results in this paper. We present the frequencies of the answers in Table 4. The questionnaire and the complete set of answers are available at `http://serg.cs.lth.se/index.php?id=87041`.

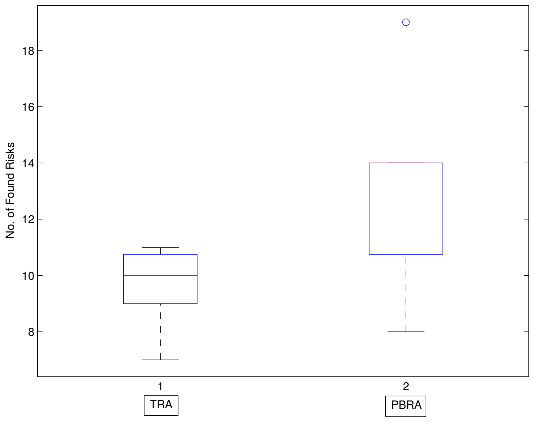

Figure 5: Box Plot of Found Risks

Table 4: Summary of result regarding RQ2 and RQ3 given in frequencies of answers given for each option

Question Method	Frequencies of answers				
Difficulty (RQ2)	Very Easy	Easy	Fair	Difficult	Very difficult
TRA	4	4	11	2	0
PBRA	0	2	12	7	0
Confidence (RQ3)	Very confident	Confident	Fair	Not confident	Strongly not confident
TRA	0	1	5	9	6
PBRA	0	2	4	11	4

The data for RQ2 and RQ3 is collected by using an ordinal scale (Likert). Therefore, the collected data has been tested by using a non-parametric test (*Mann-Whitney* U-test).

The collected data regarding RQ2 was saved in two vectors, $x1$ and $y1$, and *Mann-Whitney* U-test with the left tail was used to test the statistical difference in difficulty level while using both treatments. It tests the null hypothesis that data in vectors $x1$ and $y1$ comes from continuous distributions with equal medians, against the alternative that the median of $x1$ (TRA) is less than the median of $y1$ (PBRA). *Mann-Whitney* U-test rejected the null hypothesis H_0^2 with the p-value 0.004 meaning that the TRA method is less difficult to use than the PBRA method. The descriptive statistics, see Table 4, provides additional explanations for the test result. No subject considered PBRA *very easy* while four subjects considered TRA *very easy.* Moreover, seven subjects considered PBRA *difficult* while only two subjects considered TRA *difficult.*

The data regarding RQ3 was also saved in two vectors, $x2$ and $y2$, and *Mann-Whitney* U-test with the left tail was used to test the statistical difference in confidence level between the two samples. *Mann-Whitney* U-test could not reject the null hypothesis H_0^3 with the p-value 0.691 meaning that there is no statistical difference. Looking at Table 4, there could be several indications of lack of difference. Firstly, no subject was *very confident* of any method results. Secondly, six subjects were *strongly not confident* of the TRA method results and four subjects were *strongly not confident* of the PBRA method results. Thirdly, there are only subtle differences between the number of subjects that were *confident*, *fair*, *not confident* or *strongly not confident* about the results.

7. VALIDITY EVALUATION

The validity threats can be divided into four types [19]: conclusion, construct, internal, and external. We discuss the most relevant validity threats below.

7.1 Conclusion validity

Use of wrong statistical tests: In order to reduce this threat, the collected data was investigated for normality before parametric tests (t-test) were used.

Reliability of treatment implementation: In order to reduce this threat, all subjects received the same standard instructions in all seminars. The illustrating example in the guidelines was also same for both treatments.

Random irrelevancies: Elements outside the experimental setting can disturb the experiment's results i.e. noise, and unplanned interrupt in the experiment. In order to reduce this threat, the subjects were not interrupted during the experiment and there was no significant noise in the experiment room. Subjects were instructed to discuss as quietly as possible while merging the individual risk lists.

Random heterogeneity of subjects: We believe that there is a very little chance of this threat because the students were selected from the same level of education (master students of engineering programs) and also had almost similar knowledge and background. That is, the students come from a rather homogeneous group.

7.2 Internal validity

Maturation: In order to reduce this threat the subjects were asked to perform risk analyses in 35 minutes. It was assumed that 35 minutes would be enough to perform individual risk analysis and also subjects will not get bored.

Instrumentation: In order to reduce this threat the instrumentation of the experiment was carefully written and then evaluated by one of the co-authors. After that, an independent researcher evaluated the instrumentation. Finally, a pilot study was carried out to evaluate and improve the instrumentation.

Compensatory rivalry: This threat to internal validity is minimized since the subjects did not know that there is two different treatments.

7.3 Construct validity

Construct validity generalizes the experiment's results to the theory of the experiment. Here, the theory is that PBRA method performs better and finds more relevant risks as compared to TRA. Previous work advocated using perspectives during risk analysis [2, 3, 4] as well as provided supporting evidence that perspectives support reviews and inspections [9]. This theory is based on the assumption that the use of different perspectives can support a better and more in-depth analysis by encouraging the participants to think of different safety and security requirements.

There could be a threat to the construct validity that the participants do not interpret relevant and non-relevant risks as the experimenter intended. There could be difference of risks interpretation between the participants and experimenters. Similarly, the likelihood and consequence levels can also be misinterpreted. In order to reduce the threats to construct validity, the guidelines were written to be as clear and understandable as possible, and help was provided by clarifying any ambiguity to the participants when they asked. An example was also mentioned in the guidelines to make them unambiguous and clear. A very simple and common system was selected for the experiment and we believed that almost all the participants have already used it many time. This means that the selected system was easy to understand without domain knowledge. Finally, a pilot study was also carried out to mitigate any potential ambiguities.

The fear to be evaluated (also known as evaluation apprehension) threat to validity was reduced by clearly stating that the results of the experiment do not have any affect on the studentsÕ final grades. It is not possible to track the individual participants of the experiment for evaluation because the participants were anonymous.

7.4 External validity

Interaction of selection and treatment: There could be a chance of this threat because the subjects for the main experiment were students of a project course and are therefore not representative for the entire population. However, to reduce the affect of this threat, the pilot study was carried out by using experts from industry and academia. There was not a big difference in the number of identified relevant risks found by the industry experts and students.

Another threat to external validity is that the subjects were given 35 minutes for the individual risk analysis and then 20 minutes for the comparison and merger of individual risk lists. They were asked to find as many risks as they can but there was no upper or lower limit for the number of identified risks. The given time was also limited in order to reduce the effect of maturation. The time was chosen as a tradeoff between having the possibility to spend a lot of time and be sure to find "all" risks, and the risks of spending too much time and obtain maturation.

8. DISCUSSION

The experiment confirms that subjects using PBRA found more relevant risks. This result provides supporting evidence about the potential of roles and perspectives in risk analysis, stretching outside a simple scenario of role-based risk analysis given by Yoran and Hoffman [4] and recommendations given by Leveson [2] and Ierace [3]. Moreover, our results suggest that perspectives could increase the efficiency of not only document reviews [9] but also risk analysis and identification. Contrary to expectations, our results do not bring the supporting evidence that PBRA helps to identify fewer non-relevant risks. This indicates that there can be an advantage to assign perspectives to participants in a risk analysis, as a complement to only rely on the more natural differences between different roles in a group.

Our experiment brings statistically significant evidence that PBRA is seen as more difficult than TRA. This does not have to be interpreted as negative for PBRA. It may be that this level of difficulty is necessary, and acceptable, if the results can be more relevant risks. It may also be possible that the higher difficulty level may not be appropriate for the rather inexperienced students participating in the experiment, this calls for a replication of this study with much more experienced practitioners.

Regarding the confidence in the identified risks, no statistical significance may be caused by a lack of experience and domain knowledge in train systems. The results in Table 4 seems to support this interpretation as most subjects were highly not confident about the risks identified using any of the methods. Thus, further studies with more experienced risk managers and engineers are needed to further explore this aspect.

9. CONCLUSIONS AND FUTURE WORK

Involving perspectives into risk analysis brings a potential to increase the efficiency of the risk analysis and confidence in the identified risks. In this paper, we present the results from a study designed to experimentally assess the potential of perspectives in risk management and therefore further experimentally explore the suggestions given in previous work [2, 3, 4, 5, 6, 7, 8]. 43 subjects performed risks analysis of a software-controlled train door system using TRA and PBRA. We measured the efficiency of the methods by counting the number of relevant and non-relevant risks and a questionnaire to measure the difficulty of the methods and the confidence of the subjects in the identified risks.

Revisiting our research questions, we can with a statistical significance claim that PBRA helps to identify more relevant risks than TRA. On the other hand, our experiment failed to provide supporting evidence that PBRA helps to identify fewer non-relevant risks (RQ1). Contrary to expectations, this study did find with a statistical significance that PBRA is more difficult to use than TRA (RQ2). We interpret this result as a consequence of the subjects' limited experience in system engineering and rail domain. It may be that this level of difficulty is necessary, and acceptable, if the results can be more relevant risks. Finally, we cannot say that any of the studied methods generated risks with higher confidence (RQ3). However, most subjects were highly not confident about the risks identified using any of the methods.

In future work, we plan to replicate our study with practitioners experienced in rail domain. We also plan to apply PBRA on more complex systems by involving practitioners that have extensive experience in the system engineering approach and measure their performance. Finally, we plan to explore if different perspectives than used in this experiment (tester, train staff member and system engineer) impact the number of relevant and non-relevant risks identified using the PBRA method.

Acknowledgement

This work was funded by the Swedish Civil Contingencies Agency under a grant for PRIVAD, Program for Risk and Vulnerability Analysis Development.

10. REFERENCES

[1] J.A. McDermid, M. Nicholson, D. J. Pumfrey, and P. Fenelon. Experience with the application of hazop to computer-based systems. In *Computer Assurance, 1995. COMPASS '95. Systems Integrity, Software Safety and Process Security. Proceedings of the Tenth Annual Conference on*, pages 37–48, 1995.

[2] Nancy G. Leveson. *Engineering a Safer World: Systems Thinking Applied to Safety.* The MIT Press, 2012.

[3] Stefano Ierace. The basics of fmea, by robin e. mcdermott, raymond j. mikulak and michael r. beauregard. *Production Planning and Control*, 21(1):99–99, 2010.

[4] A. Yoran and L. J. Hoffman. Role-based risk analysis. In *20th National Information Systems Security Conference*, pages 37–51, 1997.

[5] Gary Stoneburner, Alice Y. Goguen, and Alexis Feringa. Sp 800-30. risk management guide for information technology systems. Technical report, Gaithersburg, MD, United States, 2002.

[6] International Organization for Standardization. Information technology – security techniques – information security risk management. Technical report, 1, ch. de la Voie-Creuse, CH-1211 Geneva 20, Switzerland, 2011.

[7] James Stevens Christopher Alberts, Audree Dorofee and Carol Woody. Introduction to the octaveő approach. Technical report, Pittsburgh, PA 15213-3890, 2003.

[8] Mohamed Hamdi Noureddine Boudriga and Jihene Krichene. Netram: A framework for information security risk management. Technical report, Techno-parc El Ghazala, Route de Raoued, Ariana, 2083, Tunisia, 2007.

[9] Björn Regnell, Per Runeson, and Thomas Thelin. Are the perspectives really different? further experimentation on scenario-based reading of requirements. *Empirical Softw. Engg.*, 5(4):331–356, December 2000.

[10] D.I.K. Sjoeberg, J.E. Hannay, O. Hansen, V.B. Kampenes, A. Karahasanovic, N.-K. Liborg, and A.C. Rekdal. A survey of controlled experiments in software engineering. *Software Engineering, IEEE Transactions on*, 31(9):733–753, 2005.

[11] C. A. Ericson. Fault Tree Analysis - A History. In *Proceedings of The 17th International System Safety Conference*, 1999.

[12] Per Johannessen, Fredrik TÃűrner, and Jan Torin. Actuator based hazard analysis for safety critical systems. In Maritta Heisel, Peter Liggesmeyer, and Stefan Wittmann, editors, *Computer Safety, Reliability, and Security*, volume 3219 of *Lecture Notes in Computer Science*, pages 130–141. Springer Berlin Heidelberg, 2004.

[13] Takuto Ishimatsu, Nancy G. Leveson, John Thomas, Masa Katahira, Yuko Miyamoto, and Haruka Nakao. Modeling and hazard analysis using STPA. In *NASA 2010 IV&V Annual Workshop*. NASA, September 2010.

[14] J. Thomas and Nancy G. Leveson. Performing hazard analysis on complex, software- and human-intensive systems. In *The 29th International System Safety Conference*, 2011.

[15] Victor R. Basili, Scott Green, Oliver Laitenberger, Forrest Shull, Sivert Sørumgård, and Marvin V. Zelkowitz. The empirical investigation of perspective-based reading. Technical report, College Park, MD, USA, 1995.

[16] G. Sabaliauskaite, F. Matsukawa, S. Kusumoto, and K. Inoue. An experimental comparison of checklist-based reading and perspective-based reading for uml design document inspection. pages 148 – 57, Los Alamitos, CA, USA, 2002//.

[17] O. Laitenberger, K. El Emam, and T.G. Harbich. An internally replicated quasi-experimental comparison of checklist and perspective based reading of code documents. *IEEE Transactions on Software Engineering*, 27(5):387 – 421, 2001.

[18] Zhijun Zhang, V. Basili, and B. Shneideman. Perspective-based usability inspection: an empirical validation of efficacy. *Empirical Software Engineering*, 4(1):43 – 69, 1999.

[19] Claes Wohlin, Per Runeson, Martin Höst, Magnus C Ohlsson, Björn Regnell, and Anders Wesslén. *Experimentation in software engineering*. Springer (first edition by Kluwer in 2000), 2012.

[20] Andreas Jedlitschka, Marcus Ciolkowski, and Dietmar Pfahl. Reporting experiments in software engineering. In *Guide to Advanced Empirical Software Engineering*, pages 201–228. Springer London, 2008.

[21] Sardar Muhammad Sulaman, Kim Weyns, and Martin Höst. A review of research on risk analysis methods for it systems. In *Proceedings of the 17th International Conference on Evaluation and Assessment in Software Engineering*, pages 86–96. ACM, 2013.

[22] Ian Sommerville. *Software Engineering*. Addison-Wesley, Harlow, England, 7 edition, 2010.

[23] Swedish Civil Contingencies Agency (MSB). *Guide to Risk and Vulnerability analyses*. DanagårdLiTHO Sweden, 2012.

[24] Christin Lindholm, JesperPedersen Notander, and Martin Höst. A case study on software risk analysis in medical device development. volume 94 of *Lecture Notes in Business Information Processing*, pages 143–158. Springer Berlin Heidelberg, 2012.

[25] B. W. Yap and C. H. Sim. Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation*, 81(12):2141–2155, 2011.