Effects of usability measures on the user performance of tool-supported security risk analysis

An evaluation based on lessons learned from partial re-engineering of a risk analysis tool

Fangrongling Fu

Department of Informatics
Faculty of mathematics and natural sciences

UNIVERSITY OF OSLO

August 3, 2020

Abstract

There are many different new software applications and new versions of software applications released every year. The purpose of software development is to provide better and more useful services so that people's lives become more simple and convenient. Hence, with the gradual improvement of people's quality of life, there is a higher requirement for software applications. This has also propelled the software developers to be constantly moving forward and to develop a software application to have more practical efficiency, and one way to increase the practical efficiency is to improve the usability of the application and implement the usability measures on it.

However, different usability measures have different effects on user performance. What kind of usability measures can improve user performance, and what kind of usability measures will reduce user performance, we don't truly know in fact. Hence, I use this re-engineering of the web-based CORAS opportunity to study the effects of usability measures on the user performance of tool-supported security risk analysis.

This empirical study is a controlled experiment conducted by 8 participants where comprehensibility and efficiency were investigated, using the existing web-based CORAS tool and the re-engineered web-based CORAS tool for the participants to complete the given tasks. The results of our empirical study show that there is indeed a significant difference between the two tools in terms of comprehensibility and efficiency.

Our results show that improper implementation of usability measures will have a bad effect on user performance. Most of the usability measures will improve the comprehensibility and the efficiency of users, especially the basic measures that can assist users in performing security risk analysis. Overall, the web-based CORAS tool I have re-engineered has been favorably received by most of the participants.

Acknowledgements

I would like to express my gratitude and appreciation to my thesis supervisors Aida Omerovic from Sintef, Gencer Erdogan from Sintef, and Ketil Stølen from the University of Oslo and Sintef for giving me the opportunity to undertake this project, and for providing great support, advice, encouragement, inspiration and guidance that helped drive the work in the right direction and make me complete my thesis work.

I would like to give a special thanks to my co-supervisor, Gencer Erdogan, for his support and guidance in the final months of my thesis work. Thanks to him for providing critical and constructive feedback on the thesis writing, which in turn helped shape my master thesis.

I would also like to sincerely thank Håkon Antonsen for supporting of the web-based CORAS tool programming. Especially he took the time during his busy schedule to provide me a React and Redux workshop, which enabled me to quickly understand Master the development of Web-based CORAS tools and make redesigning easier.

In addition, I would also like to thank my supervisor, Aida Omerovic, and my mother's husband, Kjell Odd Foss. Thank to them for their valuable comments on the language and structure of my thesis.

Thanks also to the people who participated in this empirical study that was carried out as part of my thesis. Because their participation was crucial. Since their active participation and cooperation, the study process could be carried out so swiftly. Moreover, gratitude for providing me with valuable data and feedback.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this chapter, I present the motivation for our work and describe our objective in this thesis. Further, I present the main contribution and give an overview of the thesis.

The question I try to answer is "How can the "usability" of a software application be enhanced?".

1.1 Motivation

With the continuous advancements in science and technology, computers are now playing a more and more important role in our society. The interactions with computers are mainly done through software applications, and there is no doubt that some software applications have better usability than others.

According to the International Standards Organization (ISO), usability refers to the extent of effectiveness, efficiency, and satisfaction of a specific user in using a product to achieve a specific goal in a specific context of the use. [9]. Moreover, usability has become one of the important factors to be considered during the design phase of software application development. Usability is a part of "usefulness" and is composed of [26]:

- Learnability: How easy is it for users to accomplish basic tasks the first time they encounter the design?
- Efficiency: Once users have learned the design, how quickly can they perform tasks?
- Memorability: When users return to the design after a period of not using it, how easily can they re-establish proficiency?

- Errors: How many errors do users make, how severe are these errors, and how easily can they recover from the errors?

- Satisfaction: How pleasant is it to use the design?

Usability is an important indicator of whether a product is easy to use. If a product does not have a high usability, the user will put a lot of time and energy into a process that they can hardly learn, see or even return, it will lead to energy leakage and frustration over time. So in order to help users avoid frustration, the usability of the product needs to be improved. Higher usability provides higher efficiency, and higher efficiency often means higher value and effectiveness/benefits, and better results. Therefore, we are always looking for some better and more practical software applications, especially those that come with high usability and can promote our efficiency and performance. Because such software applications can help us improve the quality of work and life.

1.2 Objective

However, not all software has perfect usability when it is first developed, so it needs to be re-engineered. But not all re-engineering can move software development in a good direction. Because we do not really know whether the software with usability measures can really improve and promote our efficiency or performance on work when we use them. Inappropriate re-engineering will result in reduced user performance and worse user experience, even though the re-engineering is usability oriented. Sometimes we think a usability measure is good, but it is not.

Hence, this thesis will look at the effects of usability measures, especially efficiency and comprehensibility, on the user's performance by implementing different features in the web-based CORAS tool that are aimed at improving usability. By looking at the correlation between the different types of features and their effect on the user's performance, we can provide insight into what type of usability features should be prioritized when developing software applications.

1.3 Contribution

This thesis provides two kinds of contributions. First, it provides a better artifact in terms of a risk analysis. A re-engineered web-based risk analysis software with more usability. Second, it provides an empirical study in terms of an experiment.

In the experiment, I compared the existing web-based CORAS tool and the re-engineered web-based CORAS tool to discover how the usability measures affect the user performance of the security risk analysis supported by the tool.

1.3.1 The re-engineered web-based CORAS tool

The re-engineered web-based CORAS tool is a usability-oriented re-engineered tool of the existing web-based CORAS tool. The re-engineering of the tool mainly includes the CORAS approach driven re-engineering and the design principles driven re-engineering. The purpose of re-engineering is to make the features of the web-based CORAS tool more comprehensibel and with better learnability, namely more usability, so that the tool can be easier to use. By re-engineering, the web-based CORAS tool can also indirectly improve the efficiency of users using this tool for security risk analysis.

1.3.2 Empirical Study - Comparison of the existing web-based CORAS tool and the re-engineered web-based CORAS tool

In order to measure the re-engineering of the web-based CORAS tool, an empirical study was conducted to compare the existing web-based CORAS tool and the re-engineered web-based CORAS tool. The overall goal of the study was to evaluate the existing web-based CORAS tool and the web-based CORAS tool I have re-engineered to illustrate the effects of usability measures on the user performance of tool-supported security risk analysis.

This study was conducted at the beginning of May 2020 using an experiment. The findings indicate that there is a significant difference in comprehensibility and efficiency by using the tool that is usability-oriented re-engineered and the existing tool. This is because participants using the re-engineered web-based CORAS tool spend generally less time in completing the tasks than participants using the existing web-based CORAS tool. Moreover, in terms of the task score, participants received higher scores when they used the re-engineered web-based CORAS tool.

1.4 Thesis Overview

This thesis is organized in the seven chapters as follows.

Chapter 1 – Introduction provides the motivation for the thesis, describes the goals, outlines the main contributions of the thesis, and gives an overview of the thesis's chapters.

Chapter 2 – Problem Characterisation introduces the main concepts which are involved in the thesis and the background knowledge concerning the thesis, specifies the problem addressed in this thesis, and then refines the overall objective proposed in Chapter1 to success criteria that the artifact must fulfill.

Chapter 3 – Research Method describes the research method that is used in this thesis. This research method aims to improve or produce new artificial products. Moreover, the most common evaluation strategies, what they are and what they measure, are introduced. Finally, this chapter also discusses and describes the evaluation strategies applicable to this thesis.

Chapter 4 – Innovation describes the entire process of the web-based CORAS tool re-engineering in detail. This process is divided into four steps. These four steps are artifact re-engineering design, likelihood calculation feature implementation, user interface optimization, and other re-engineering implementation. In each step, I introduce what I need to do, how I use the supporting documentation to do the re-engineering, and why it is beneficial.

Chapter 5 – Evaluation - Empirical Study provides an empirical study conducted by using an experiment. In this chapter, I describe the characteristics of this empirical study, set the goal for this empirical study, as well as formulate research questions and what to measure for that. Moreover, I also formulate hypotheses, determine variables, as well as identify the subjects for this empirical study and empirical study design. The execution of the empirical study and the results of the experimental analysis are described later in this chapter. Finally, at the end of this chapter, it is described how I analyze the results of this experiment by visualizing data, applying descriptive statistical data, and conducting hypothesis testing.

The empirical study in this chapter aims to evaluate the existing web-based CORAS tool and the web-based CORAS tool I have re-engineered to illustrate the effects of usability measures on the user performance of tool-supported security risk analysis.

Chapter 6 – Discussion of my achievements concerning the success criteria.

Chapter 7 – Conclusion concludes the thesis and provides possible directions for future work.

Chapter 2

Problem characterization

In this chapter, I first provide some background informations and core concepts that play a vital role throughout this thesis. These related background informations and core concepts are clarified in Section 2.1. In Section 2.2, I present the problems to be solved in this article. Finally, I refine the overall objective of the research described in Chapter 1 into a set of success criteria and present the success criteria that should be met to successfully achieve our objective in Section 2.3.

2.1 Background and Conceptual Clarification

In the following sections, I will look into the background information of the tools and languages used for building web-based software. Furthermore, I provide background information of the CORAS approach, as well as a summary of the latest version of the CORAS tool. Finally, I also provide an introduction to the design principles that I used to modify the user interface of the exiting CORAS tool.

2.1.1 Tools and languages for building a web-based software

The list below contains tools and languages that can use for developing or re-engineering of Ib applications. Figure 2.1 shows the relationship betIen them.

HyperText Markup Language

HyperText Markup Language, often abbreviated as HTML, is the World WideWeb's core markup language. Its initial release was in 1993 and has been

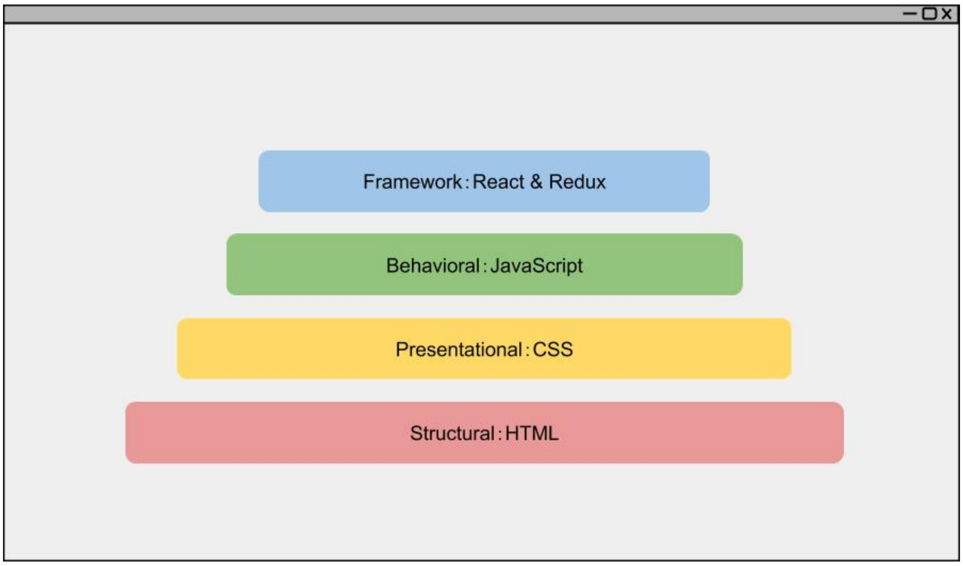


Figure 2.1: Relationship diagram betIen the HTML, JavaScript, CSS, React and Redux

in use for 26 years. It was developed by WHATG[1], and W3C[2] was its former maintainer. It is one of the cornerstone technologies for the World Wide Web, and it often used alongside CSS and JavaScript to design a user interface for web pages, web applications, and mobile applications. One of the main tasks of HTML is to give meaning to the text, also known as semantics, so that the browser knows how to display the text correctly. In the beginning, the main design purpose of HTML was to be a language to semantically describe the scientific documents. But it had been continually adjusted over the years, to make it gradually useful for describing many other types of documents and applications.

JavaScript

JavaScript, often abbreviated as JS, is a high-level, interpreted language. An interpreted language is a programming language that will be executed by the interpreter/engine directly one sentence at a time. It does not need to be compiled into machine code by the compiler first and then executed, like the compiled languages. JavaScript was designed by Brendan Eich and developed by Netscape Communications Corporation, Mozilla Foundation, and Ecma International. It first appeared on 4 December 1995, and it has already been

[1] Web Hypertext Application Technology Working Group https://whatwg.org/
[2] World Wide Web Consortium https://www.w3.org/

used for over 23 years. JavaScript is also one of the cornerstone technologies for the World Wide Web. We can use JavaScript to implement complex functions on the web page. The content displayed on the webpage is no longer simple static information, but real-time content updates, interactive maps, 2D/3D animations, scrolling videos, etc. This language is not only available for HTML and the web, but is more widely used on servers, PCs, laptops, tablets, and smartphones.

Cascading Style Sheets

Cascading Style Sheets, often abbreviated as CSS, is a style sheet language used for describing the presentation of a document written in a markup language like HTML [4]. A computer language used to design style and layout. Generally, CSS can be used to control and adjust the appearance of elements on the web page, such as colors, fonts, and size, etc. CSS can also be used to divide the content into multiple columns, or add animation and other decorative effects. It is the last one of the cornerstone technologies for the World Wide Web. It was developed by Håkon Wium Lie, Bert Bos, and W3C. Its initial release was on 17 December 1996, and it has already been used for 22 years.

React and Redux

React [21] and Redux [22] are the open-source libraries of JavaScript. It was initially released on 29 May 2013 and it is maintained by Facebook and the community. React is used for building the user interface, as a base in the development of web-page or mobile applications. Using React, developers can design simple views for each state in the application, called components, and when the data changes, the code in the library will do the heavy lifting and update the right components. This means that the developers will make encapsulated components that manage their own state, then compose them to make complex User Interfaces. This use of encapsulation has helped the web applications built with React to be more predictable and easier to debug. Since React has excellent performance and the code logic is very simple, more and more people are beginning to pay attention to and use it, and thinking that it may become the mainstream tool for Web development in the future.

When React applications become more complex, the use of additional libraries for state management, routing, and interaction with an API will be needed to reduce the complexity of the overall structure of the software. The additional library of state management that is most commonly used together with React for building user interface is Redux, which is designed to be a predictable container for application state. Using Redux, the complexity of state management and state sharing between different components will be reduced. It can be used not only with other libraries but also independently. It was created by Dan Abramov and Andrew Clark. Its initial release was 2 June 2015. The use of React and Redux can make the design and implementation of data management and interaction between people and servers easier.

2.1.2 CORAS approach

CORAS is a model-based approach to risk analysis with a focus on security. CORAS consists of three artifacts, namely language, tool, and method. The three artifacts together are referred to as the CORAS approach.

The CORAS language is a customized language for risk modeling [35]. The language is diagrammatic, that is to say, the CORAS language is a graphical modeling language. Modeling languages can be used to express information or knowledge or systems in a structure with a specific set of meanings. Graphical modeling languages use charting techniques with named symbols, which represent concepts, while lines connect these symbols and represent the relationship between them, also various other graphical symbols are used to represent constraints [15].

CORAS language provides five kinds of diagrams, each supporting a specific phase in a risk analysis process. The diagrams created by using this language are easy to understand, because it uses simple graphical symbols and the relationships between them are easy to read. Therefore, such diagrams are suitable as a medium for communication between different stakeholders with a diverse background, thereby improving both the efficiency of the risk analysis process and the quality of the risk analysis results.

The CORAS tool is a graphical editor for making CORAS diagrams by using the CORAS language. The tool helps document and present risk analysis results. It is also suitable for creating risk models instantly during brainstorming sessions. Use the CORAS tool to make CORAS diagrams on-the-fly to support, stimulate, and document the discussions and the findings during meetings. Also, the CORAS tool supports model revising and analysis between meetings.

The CORAS method is an asset-driven defensive risk analysis method. It has detailed guidelines explaining how to perform various stages of the CORAS risk analysis in practice. A risk analysis using CORAS consists of eight steps that are shown in Figure 2.2. The first four steps are introductory. They are used to establish a common understanding of the analytical target, and to make the target description as a basis for subsequent risk identification. The remaining four steps are dedicated to the actual detailed analysis. This includes identifying specific risks and their level of risk, as well as identification and assessment of potential treatments for unacceptable risks.

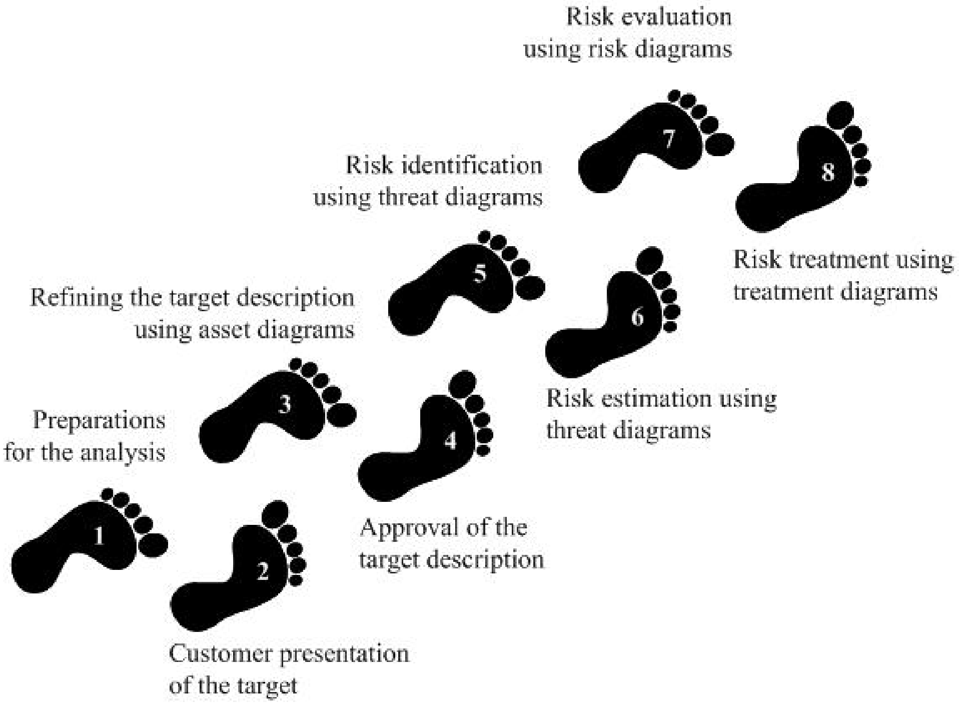

Risk evaluation using risk diagrams
Risk identification using threat diagrams
Refining the target description using asset diagrams
Preparations for the analysis
Risk treatment using treatment diagrams
Risk estimation using threat diagrams
Approval of the target description
Customer presentation of the target
1
2
3
4
5
6
7
8

Figure 2.2: The eight steps of the CORAS method (adopted from K. Stølen. [35])

In the actual detailed analysis process, it also involves analyzing the likelihood of risk occurrence using the CORAS diagram. About the rules of likelihood analyzing and how to use the CORAS diagram to analyze the likelihood, I will explain in more detail in Chapter 4.

Web-based CORAS tool

The existing and latest CORAS tool [27] is a web-based risk analysis tool and was developed by Håkon A. V. Antonsen. Since it is web-based, this tool does not need to be downloaded and installed on a computer and can be used via the web page. The languages used to make the web-based CORAS tool are JavaScript, HTML and CSS. As shown in Figure 2.1, HTML is used to make the structure of a web page, JavaScript is for behaviors, and CSS can decorate web pages and make it look better. Also, this tool uses React and Redux as a framework to make the system easier for people to interact. When Antonsen made the CORAS tool, he followed the Atomic Design concept. Atomic Design was proposed by Brad Forst in 2013. The Atomic Design concept was inspired by chemistry: all matter is composed of atoms, and the bonds between these atomic units work together to form molecules, which in turn combine into more complex organisms, and these organisms ultimately create all matter in our universe. [1]. Atomic design is, therefore, a design concept, a methodology in which a combination of atoms, molecules, organizations, templates, and pages work together to create more efficient user interface systems. It is usually used to build a scientific and standardized design system. Since the CORAS tool is open source, it can be therefore further developed by the open-source community.

2.1.3 Design principles

Design principles are written in a prescriptive way, and are some principles derived from a mix of theory-based knowledge, experience, and common sense [40]. They are used to help thinking when designing for the user experience. The advantages of using design principles are that they broaden the thinking of the designer so that the designers can consider different aspects of their designs. Moreover, they can also give designers suggestions on what to provide and what to avoid on the interface. Design principles are like considerations for interaction design and more like a designer's trigger, ensuring that some features have been provided for on an interface. Therefore, the objective of design principles is intended to help designers explain and improve their designs rather than specifying how to design an actual interface [40].

Affordance

Affordance is a term used to refer to an attribute of an object that allows people to know how to use it, and to afford means 'to give a clue' [40]. From a simple perspective, If an object has an obvious perceptual affordance, it would be like giving a hint to make the user easy to know how to interact with it. For example, a button invites people to press it by the way it is designed to afford to push. So when you see a button, it gives you a hint you may press it to make an event start. Affordances support our successful interaction with the virtual objects' world so it is making our life easier.

Consistency

Consistency refers to designing interfaces to have similar operations and use similar elements to accomplish similar tasks [40]. When consistency is contained in the design, people can apply existing knowledge to the new contexts, thus ensuring that users do not have to learn new representations of each task, so that people can quickly learn new things without suffering.

Constraints

Constraints refers to determining ways of restricting the kinds of user interaction that can take place at a given moment [40]. One of the advantages of constraints is that it can prevent users from making wrong operations, thereby reducing the chance of making a mistake. The use of different kinds of graphical representations is a good way to constrain.

Visibility

Visibility is also an important design principle. The more visible features are, the more likely it is that users will be able to know what to do now or next [40]. For example, horns, hazard warning lights, and progress bars, all of them indicating what can be done. Instead, when certain functions are invisible, it will make them more difficult to find and know how to use. Such as doors, lights, and faucets that using induction technology. The activating zones of them are invisible and ambiguous and make it harder for people to control, especially when activating or deactivating them, people have to guess where to put their hands, bodies or feet to make them work.

Feedback

Feedback is related to the concept of visibility. It involves sending back information about the action that has been performed and the operation that has been completed, so that the person can know what has been done and what activities could be continued. For example, when we press the doorbell, we hear the music, and then we know that we have pressed the doorbell. In addition to audio feedback, there are many other kinds of feedback that are available, such as tactile, verbal, visual, and combinations of these. Using feedback in the right way can also provide the necessary visibility for user interaction [40].

2.2 Problem specification

The purpose of the research in this thesis is concerned with evaluating the effects of usability measures on the user's performance by implementing new features and new user interface in the web-based CORAS tool that are aimed at improving usability. The problems addressed in this thesis involves creating new usability features and modifying the user interface for the web-based CORAS risk analysis tool. Therefore, there are several challenges to discuss concerning the development of the CORAS tool.

First, it requires good preparation before starting starting the re-engineering of the CORAS tool. The preparation work is carried out during the problem analysis in technology research and is to help to find the right direction for re-engineering. At the beginning of the problem analysis process, I need to fully understand the current situation of existing Web-based CORAS tools. What features are already included by the existing Web-based CORAS tool, and what the basic CORAS approach features that should be provided, but not yet have. Then find out if there are any existing designs on the web-based CORAS tool that reduce user performance, and modify or improve them. Through the use and investigation of the Web-based CORAS tool, talking with experts of the CORAS approach, and reading literature on the CORAS approach and security risk analysis [35], I have obtained the contents of the current situation of the existing Web-based CORAS tool at the following table 2.1.

Through the initial use and understanding of the web-based CORAS tool, I know that users can use this tool to draw all types of CORAS diagrams. After drawing the diagram, you can save the diagram. Diagrams can be downloaded as SVG format or saved as JSON format. The SVG format is a scalable vector graphic, while the JSON format is used for the latter upload to the tool and manipulating the diagram again. This CORAS tool has a great advantage: if the user accidentally closes the CORAS tool web-page after drawing a picture, there is no need to worry about losing pictures due to unsaved, because it will be stored in the browser's cache. When the user opens the web page of the CORAS tool again with the same browser, the picture that was just drawn will reappear. Finally, there is a function that if you are not satisfied with the drawing, you can directly clear the drawing paper by pressing the "clean" button.

By reading kinds of literature on the CORAS approach and security risk analysis, and by talking with experts of the CORAS approach, I found out what the existing web-based CORAS tools lack. For instance, the tool lacks the function "Using CORAS diagrams to calculate the likelihood" that is introduced in the book "Model-Driven Risk Analysis". Therefore, the tool cannot calculate probability and frequency in CORAS Diagrams. "Using CORAS diagrams to calculate likelihood" is a really important part of doing security risk analysis by using the CORAS approach. In order to achieve the calculation function of likelihood, it is necessary to add the likelihood input box. This is also what the existing web-based CORAS tool does not have. Another features the existing web-based CORAS tool does not have is restrictions on connections between elements. This will be explained in more detail later.

In addition, by using web-based CORAS tools and investigating the user experience of people who have used it, I have identified its current problems and shortcomings. The first is the zoom function of the drawing panel. Because the zoom function is operated by the scroll bar function of the mouse, it will interfere with the up and down sliding of the page and indirectly affect the user performance. So this needs to be modified. Another improvement is the position of the toolbar. It could be placed in a place that can provide a better user experience and user performance.

Available features	• Draw all the different types of CORAS diagrams • Download the completed CORAS diagrams as a SVG file • Download the completed CORAS diagrams as a JSON file • Upload JSON files of previously drawn CORAS diagrams for refurbishment or modification of diagrams • Clean up drawing paper clean up
Missing basic features	• Input box for assigning likelihood value to the element • Calculating likelihood in CORAS diagrams • Restrictions on connections between elements • Likelihoods consistency checking in CORAS diagrams • Prompt related error messages when an error occurs
Areas for improvement	• Position of Tool Bar • The zoom function of drawing paper

Table 2.1: The current situation of the existing Web-based CORAS tool

Through the analysis results of the web-based CORAS tool, the correct re-engineering direction is pointed out for the development of new features and user interface.

2.3 Success Criteria

In order to fulfill the requirements mentioned in the previous section, as well as the overall aim of the thesis, we need to identify and establish the success criteria. Success criteria are established based on the interests of stakeholders that will benefit from the creation and using of this tool, they are developer and user. The success criteria are as follows:

Success Criterion 1.

The tool should correctly calculate the risk likelihood based on the drawn CORAS diagram.
The Likelihood calculation of the CORAS approach is carried out by strictly following the rules. Based on the drawn CORAS diagram, different rules are used to calculate the likelihood under different conditions. Therefore, it is necessary to follow the rules to correctly implement the likelihood calculation feature.

Success Criterion 2.

The re-engineered tool must be more comprehensible for users.
The users who use the tool for risk analysis will be the main stakeholders and benefit from the tool. Therefore, both the existing tool and the new re-engineered tool must be easy to understand for the users. Hence, it is important to express the features and user interface in a simple, understandable, and correct way.

Success Criterion 3.

The re-engineered tool must improve the efficiency of carrying out a security risk analysis.
In real life, whether it is an enterprise or an individual, people focus on efficiency. High-efficiency means saving in terms of time, effort, and resources. So it would be great to be able to use a tool that improves work efficiency.

Success Criterion 4.

The re-engineered tool must improve users' satisfaction with using the tool.
No matter how good a tool is, no matter how innovative it is. As long as users are not satisfied with it, the existence of such a tool has no meaning at all. Therefore users' satisfaction with using the tool is important to improve.

Chapter 3

Research Method

Computer science is considered as the fourth largest domain of science, and is usually associated with physical sciences (which focus on the non-living matter), life sciences (which focus on the living matter), and social sciences (which focus on humans and their society). As an interdisciplinary scientific domain, it is important to use appropriate research methods when executing a research project in a given computer science environment. In this section, I discuss the research method to be applied in the proposed thesis that will help fulfill the success criteria.

According to Merriam-Webster's definition of research, research is:

> *Investigation or experimentation aimed at the discovery and interpretation of facts, revision of accepted theories or laws in the light of new facts, or practical application of such new or revised theories or laws [5].*

In other words, what we seek is information that will either increase knowledge, modify existing knowledge, or find new uses for already existing knowledge. First, Researchers must formulate hypotheses as a starting point when conducting research. A hypothesis is a question, a basis for the research, and provides a tentative explanation for answering this question [46]. The researchers use verifiability- or falsifiability-oriented experiments and observations to test whether the hypothesis is true in reality in hypothesis testing. Hypothesis testing is also referred to as evaluation [46]. It is common to make predictions regarding the outcome of the observations and investigations. Predictions are statements that are only proven true if the hypothesis is true [46]. If an evaluation of a hypothesis confirms the predictions, the hypothesis is strengthened. However, if the predictions are proven false, it can cause a rejection of the hypothesis [46]. This approach is commonly called the "scientific method". Solheim and Stølen also defined this approach as "classical research". Even if a hypothesis is strengthened through evaluation, the ultimate answers can never be given, so hypotheses consequently always remain assumptions. However, there may arise some new questions worth studying, so classic research is an iterative process. The definition of "classical research" as defined by Solheim and Stølen is as follows:

> *Classical research is research focusing on the world around us, seeking new knowledge about nature, space, the human body, the society, etc. The researcher asks: What is the real world like?*

This research method is mainly rooted in what Solheim and Stølen call basic research, and the definition of basic research is "Research for the purpose of obtaining new knowledge" [46]. The main steps of basic research are problem analysis, innovation, and evaluation. However, my proposed thesis is more concerned with asking questions regarding technology. It is about finding better ways of solving practical problems, and especially how to help the software engineering domain. To this end, I will use a research method that Solheim and Stølen call technology research.

The difference between technology research and basic research is that technology research is “research for the purpose of producing new and better artifacts”. It is mainly rooted in what Solheim and Stølen call applied research, defined as "research seeking solutions to practical problems". While the main difference between basic research and applied research is that basic research aims to discover new general information about the real world that may not be directly applicable. Applied research is specifically designed to directly apply and solve practical problems. Technical researchers seek the principles and ideas for making new and better artifacts in an attempt to use them to create better artifacts than existing products. According to Solheim and Stølen, an artifact is an object manufactured by humans, an object intended to be useful for humans. These artifacts may be materials, medicines, computer programs, algorithms, techniques, etc. This is similar to the definition of artifacts in design science, artifacts as something that people create for some practical purpose. As similarly as basic research, technology research is an iterative method that consists of three main steps, problem analysis, research-based design, and evaluation. This is also similar to design science, where the objective is to design artifacts to interact with a problem context to improve something in that context [48]. Design science is also an iterative method that consists of three main steps, which correspond to the steps in technology research. These steps are problem investigation, treatment design, and treatment validation. The difference between classic research and design science is that classical research aims to understand reality, but design science aims to develop artifacts that serve human purposes.

In the following sections, I will further explain technology research, give an overview of evaluation strategies, and give an overview of the selected evaluation strategies for my thesis.

3.1 Technology Research

Technical research is a research method whose motivation is a need for a new artifact, or a need to improve an existing artifact, e.g. a new robot, new algorithms for a computer program, a new construction method for a bridge, a new medicine, a new treatment of patients, etc. [46]. The first thing a researcher must do when conducting technology research is problem analysis. For instance, in my case, I need to analyze the existing web-based CORAS tools to understand what the web-based CORAS tools already have and what the basic CORAS method functions should have but not yet. This analyzing process is for finding the right redesign direction and identifying needs for improvement. Furthermore, doing problem analysis is also to collect a set of requirements for the artifact. Such requirements are given by existing users (in case the artifact needs to be improved) or new or potential users (in case the artifact does not yet exist). Also, the researcher collects requirements and perspectives from other stakeholders, such as those who make money on it. In constrast to classic research, instead of asking the question: What is the real world like?, we ask How can we improve artifacts or produce new artifacts that benefit humanity in solving practical issues in the real world? After the requirements have been identified, the researcher starts to design, to construct an artifact that satisfies the needs. In this innovation phase, the researcher uses his/her creativity and technical insights to invent the artifact. Finally, after the artifact is developed, the researcher can use several evaluation strategies to evaluate the artifact to proposed whether it meets the proposed requirements and thus the (potential) needs on which it is based. If the results of the evaluation are successful, the researcher may argue that the artifact satisfies the need and the new artifact is useful, or that it is better than its predecessor. Conversely, if the results of the evaluation deviate, the researcher may try to adjust the requirements or the artifact accordingly, and reiterate the evaluation results. This stimulates new iterations in the research cycle. Thus, the technology research method is also an iterative method, and follows the same basic steps as the classical research method, as Figure 3.1 shows.

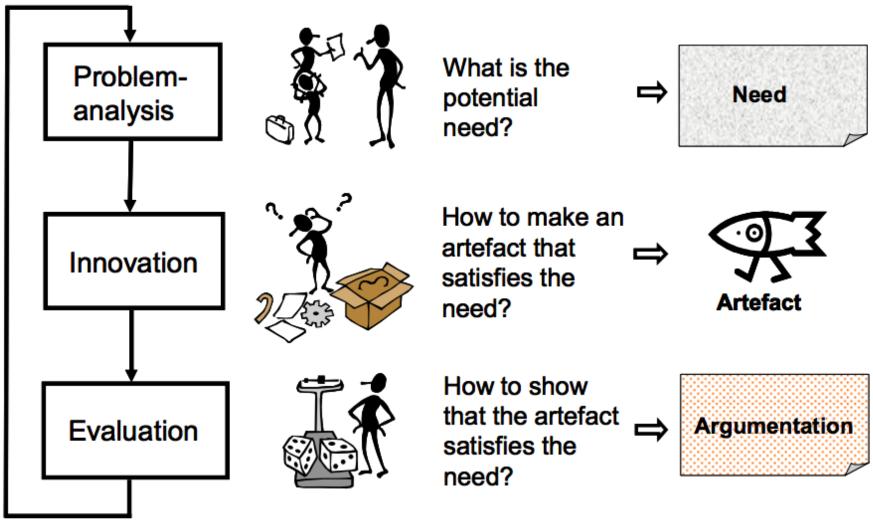

Problem-analysis
Innovation
Evaluation
What is the potential need?
Need
How to make an artefact that satisfies the need?
Artefact
How to show that the artefact satisfies the need?
Argumentation

Figure 3.1: Method for Technology Research - Main steps (adopted from I. Solheim and K. Stølen [46])

3.2 Evaluation Strategies

Evaluations are used to find out whether the predictions for something are true and whether the artifact satisfies the requirements established during the problem analysis. The evaluation strategy is the process of providing information about how the artifact meets the requirements. There are a variety of evaluation strategies to choose from. Which strategy to choose depends on the following factors. According to McGrath [38], there are three desired properties' degrees that the artifact must consider, when strategies are carried out to collect evidence under evaluation:

- **Generality**
  A measure of whether the results are valid across populations.
- **Precision**
  The measurement of the obtained results precise, and control of external variables that are not within the field of the study
- **Realism**
  To what extent does the evaluation reflect reality (the evaluation is conducted in a realistic environment)

The best desirable choice of strategy would be to maximize the scores of all these three factors simultaneously, but McGrath argues that this is not possible, because every research strategy is flawed. Different strategies have different flaws, therefore one must choose evaluation strategies that complement each other to attain acceptable levels for each property. Therefore, when choosing strategies, the researcher also has to consider other factors. For instance, the available resources that support strategy. When choosing an evaluation the strategy, time and cost are two important constraints. Also, it includes individuals who can participate in the assessment. These resources can determine whether the strategy is feasible.

In the following, we give a brief description of each of the eight most common evaluation strategies, as depicted in Figure 3.2.

- **Laboratory experiment** – Providing the researcher with a greater degree of control, and it is possible to isolate the variables to be examined to obtain high precision at the cost of reducing generality and realism;
- **Experimental simulation** – A laboratory test that simulates relevant parts and processes in the real world;
- **Field experiment** – experiments conducted in a natural environment, but certain factors that are intervened and manipulated by researchers;
- **Field study** – with highly realistic, because it is a direct observation of "natural" systems, with little or no interference from the researcher;
- **Computer simulation** – operating on a model of a given system;
- **Non-empirical evidence** – argumentation based on logical reasoning;
- **Survey** – Information collected from a broad and carefully selected group of informants, and is usually collected through questionnaires and interviews;
- **Qualitative interview** – a collection of information from a few selected individuals. The answers are more precise than those of a survey but cannot be generalized to the same degree.

Further, these eight most common strategies are divided into the following four pairs.

i The evaluation is performed in a natural environment.

ii The evaluation is performed in an artificial environment.

iii The evaluation is independent of the environment.

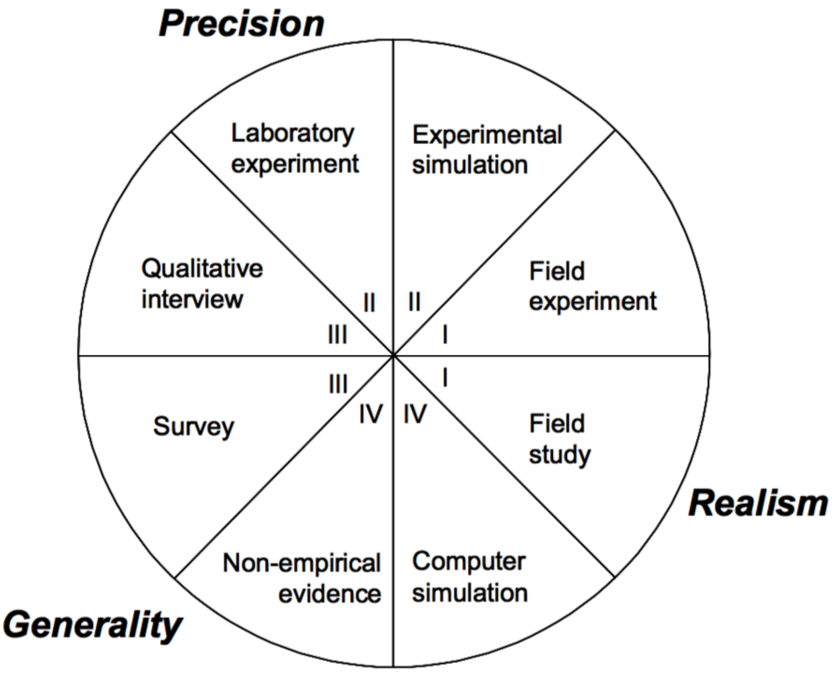


Figure 3.2: Evaluation strategies (adapted from McGrath [38])

iv The evaluation is independent of empirical measurements.

In addition to the above evaluation strategies, there are also some other strategies that are pointed out in Section 3.5.

3.3 Selection of appropriate evolution strategies

As mentioned in the introduction of this chapter, I have used the technology research method as my main research method. Following its iterative nature, as I gain new insights, I make changes and improvements to the artifact and its related success criteria. Documentation for these three phases of this iterative method, namely problem analysis, innovation, and evaluation, are found in Chapters 2, 4, and 5, respectively. As the starting point of the evaluation process, I must first select a suitable evaluation strategy. In order to select a suitable evaluation strategy from the evaluation strategies described in the previous section, I re-examined the requirements or success criteria identified in section 2.3:

1. The tool should correctly calculate the risk likelihood based on the drawn

CORAS diagram.

2. The re-engineered tool must be more comprehensible for users.

3. The re-engineered tool must improve the efficiency of carrying out a security risk analysis.

4. The re-engineered tool must improve users' satisfaction with using the tool.

Regarding success criteria one, I need to assess whether the tool can correctly calculate the likelihood from the drawn CORAS diagram. This involves verifying whether the tool can support the "Using CORAS diagrams to calculate likelihood" feature. To this end, I can use an evaluation strategy called prototyping to obtain a better discovery and understanding of the requirements of the artifact. Prototyping will be between experimental simulations and field experiments, as I will try to simulate users and potential users of security risk analysis tools while controlling certain research factors. For a further description of the prototype, see Section 3.4.

Success criteria two, three, and four are mainly related to users' perception and use of the tool. I need to evaluate the users' use and understanding of the tool to evaluate whether the artifact is more comprehensible to users, whether it has improved the efficiency of performing security risk analysis, and whether it has improved users' satisfaction with the use of the tool. In other words, I evaluate whether the artifact meets the three quality components that constitute usability, namely learning, efficiency, and satisfaction. In addition, I need to evaluate whether the usability measures promote the user performance of tool-supported security risk analysis. To evaluate these three success criteria I conducted an empirical study to gain precision and realism. An empirical study is a strategy for evaluation through direct and indirect observations or experience. This approach requires a set of individuals that are available to participate in the empirical study. Therefore, it is often difficult to find the appropriate participants. For empirical studies, students are traditionally selected as participants. However, having industry professionals are often required to gain accurate insight and responses to research questions [44]. In Section 3.5, I have further described empirical studies.

3.4 Prototyping

Prototyping refers to building software application prototypes [28]. A prototype is an initial version and represents the artifact created from requirements established initially in the problem analysis. The prototyping process involves writing programs to discover and understand the optimal design and the construction of these designs. This method helps us identify the strengths and weaknesses of our tools early in development and discover new requirements or success criteria. As described by Balzer et al. [29] "Given a proposed solution to a problem, prototyping is used to answer three types of questions: Is this a method for achieving the solution; does the proposed implementation have acceptable performance, production cost, and reliability; and is it a good solution?" [29]. In order to find a good solution, the prototype we make needs to be evaluated and modified many times. Prototyping is therefore an iterative approach. Through each iteration, new requirements, or a better understanding of existing requirements would be obtained, thereby the prototype that best meets the requirements will be gradually produced.

3.5 Empirical Study

Executing an empirical study often requires investigation strategies to help to evaluate and validate the research results. Depending on the purpose of the evaluation, and depending on the conditions for the empirical investigation, there are four major different types of investigation strategies: experiments, case studies, surveys, and post-mortem analysis [49]. These strategies are needed so that they can be scientifically indicated whether something is better than others. Therefore, the empirical study provides an important scientific basis for software engineering and is essential for researchers.

Surveys and case studies are both qualitative and quantitative evaluation strategies, while an experiment is a quantitative evaluation strategy [49]. The quantitative study is appropriate when testing the effect of some operations or activities. The purpose is to determine causality. While a qualitative study on beliefs and understandings are appropriate to explain the phenomenon, and to find out the reason why the results from a quantitative study are as they are. The two approaches should be regarded as complementary rather than competitive [49].

- **Experiment** Experiments are most often conducted in a laboratory setting. Since they are concerned with a limited scope, experiments are sometimes referred to as research-in-the-small [49]. Experiments are usually highly controlled. During the experiment, the experimenter can control and manipulate one or more variables, while all the other variables remain unchanged. Then apply treatment to them to observe the effect through the experiment's control group, measure the effect of the manipulation, and perform a statistical analysis accordingly. There are two types of experiments: random experiments and quasi-experiments. The former is an experiment in which subjects are randomly assigned to different treatment methods, while in the latter term it is impossible to randomly assign subjects to the different treatments.

- **Case studies** A case study is sometimes referred to as research-in-the-typical, is a study related to researching real projects, activities, or assignments. In the case study researcher execute statistical analysis by collecting data during the observation of an on-going project or activity. The case study is usually designed to track a specific attribute or establish relationships between different attributes [49]. Since a case study is an observational study and an experiment is a controlled study, the amount of control in the case study is less than the control in the experiment. For instance, in my case, a case study may be aimed at performing the security risk analysis with the tool in a real project.

- **Surveys** The survey is often referred to as research-in-the-large and research-in-the-past [49]. It is research-in-the-large because of the collections of information from an extensive and massive target population we have by interview or questionnaire. It is also research-in-the-past because a survey is usually retrospective, such as investigation of a tool or technique that has been in use for some time. Through interviews or questionnaires in the case study, qualitative or quantitative data can be collected. The collected data from the survey are then analyzed to arrive at descriptive and explanatory conclusions.

- **Post-mortem analysis** A post-mortem analysis may be viewed as inheriting properties from both surveys and case studies, it is also research-in-the-past as indicated by the name [49]. Use of this type of analysis can retrospectively study any part of a project. The post-mortem analysis can be performed by viewing the project documentation or by interviewing individuals or groups, who involved in the object that being analyzed in the post-mortem analysis.

Chapter 4

Innovation

In this chapter, I carry out the second step of the technology research method introduced in Section 3.1, the innovation step. In this step, I re-engineer the existing web-based CORAS tool [27]. Figure 4.1 illustrates the steps involved in the re-engineering process. There are two main steps in the process, they are artifact re-engineering design and artifact re-engineering implementation. There are also three sub-steps in the second main step. Therefore, the whole process consists of a total of four steps. For each step, they have their own corresponding input and output. Input is an important element to help carry out the step, while output is the outcome or result of a step and it can also be the input to the following steps.

The first main step is the design of artifact engineering. In this step, I use data from problem specification as the input to further determine what to do and how to do the re-engineering of the existing web-based CORAS tool, as well as what is needed in the re-engineering. And then by completing the first step, I get a re-engineering implementation plan as an output.

The second main step is the implementation of the re-engineered artifact, and it is subdivided into 3 sub-steps, these are likelihood calculation feature implementation, user interface optimization, and other implementations described in Section 4.4. To start the second step, I use the output from the first step and the supporting documentation as the second step's input. The supporting documentation consists of the CORAS approach, design principles, and the feedback of the existing CORAS tool. They are the basis for the artifact I will re-engineer and implement in each sub-steps. By using these supporting documents, I develop an artifact that better fulfills the specified success criteria. The output at the end is the final re-engineered tool, the re-engineered web-based CORAS tool.

In the following sections of Chapter 4, I will introduce each step of re-engineering in more detail.

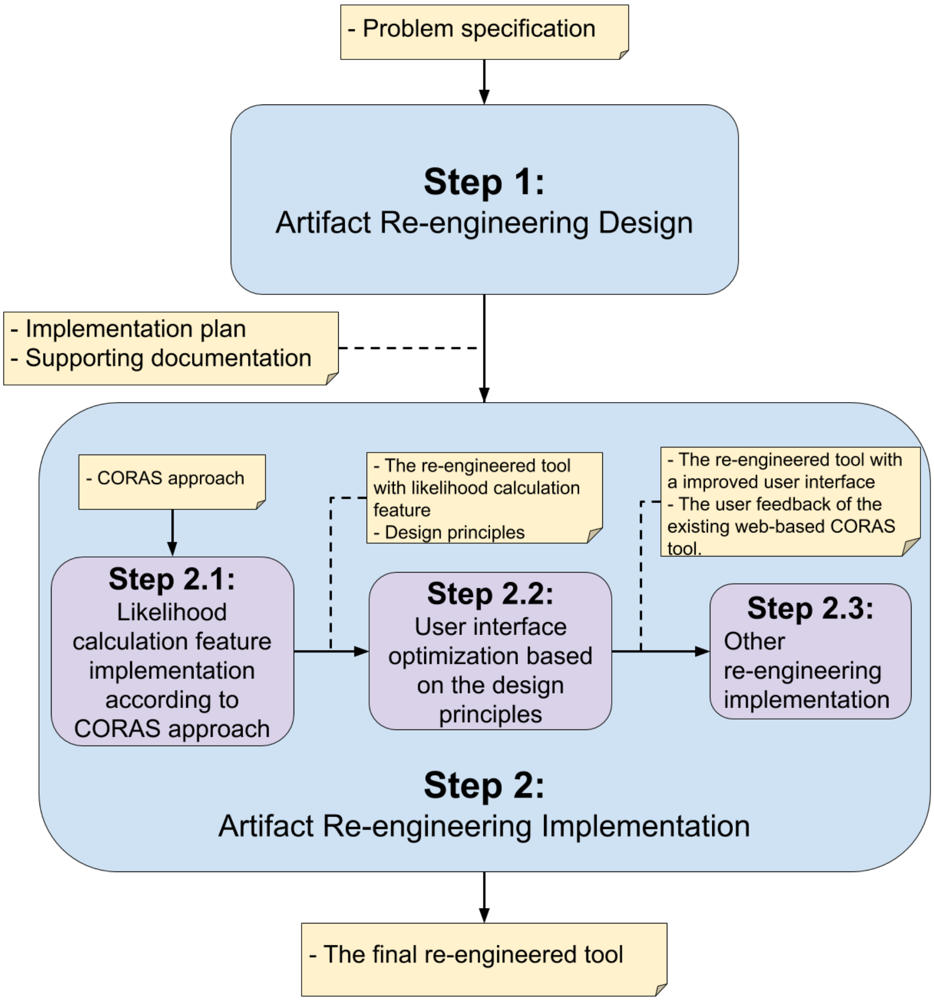

- Problem specification
Step 1:
Artifact Re-engineering Design
- Implementation plan
- Supporting documentation
- CORAS approach
- The re-engineered tool with likelihood calculation feature
- Design principles
- The re-engineered tool with a improved user interface
- The user feedback of the existing web-based CORAS tool.
Step 2.1:
Likelihood calculation feature implementation according to CORAS approach
Step 2.2:
User interface optimization based on the design principles
Step 2.3:
Other re-engineering implementation
Step 2:
Artifact Re-engineering Implementation
- The final re-engineered tool

Figure 4.1: The re-engineering process

4.1 Artifact Re-engineering design

In Section 2.2, I discussed the current situation of the existing web-based CORAS tool, namely currently available features, currently missing features, and areas for improvement. Since the existing web-based CORAS tool uses two JavaScript libraries, React and Redux, for development, in order to maintain consistency I use them and follow the architecture of the existing web-based CORAS tool to do the re-engineering. By analyzing the current situation of the existing web-based CORAS tool, I know where re-engineering should start. First, I need to add the features missing from the web-based CORAS tool according to the CORAS approach described in the book [35]. Secondly, I should ensure that the user interface should be as clean and simple as it possible can, so that the tool can be more comprehensible. Moreover, a clean and simple user interface can also reduce user errors and confusion during its use, thereby indirectly improving work efficiency. Finally, I should make sure that all re-engineered features should improve the usability of the tool, and make the tool easier to use for the users.

According to the CORAS approach introduced in the book [35], the tool needs a new feature, which is using CORAS diagrams to calculate the likelihood of risks. As mentioned in the book [35], the likelihood can be divided into probability and frequency, there are therefore two algorithms that are contained in the likelihood calculation, one is used to calculate the probability of unwanted incidents or threat scenarios and the other is used to calculate the frequency of unwanted incidents or threat scenarios. Therefore, the user can calculate the probability and frequency by drawing the CORAS graph, which is very important for the likelihood calculation and overall security risk analysis. At the same time, the addition of this new feature can greatly improve the efficiency of the security risk analysis. In addition, I also learned that the drawing of the CORAS diagrams has many regulations. For instance, there is no direct relationship between some elements, that is, there is no way to directly connect them together. In order to implement these regulations, I need to implement features that guide the user in creating syntactically correct CORAS diagrams. This feature can reduce the mistakes that users should not make when drawing a CORAS diagram.

Concerning usability, I found there are still some deficiencies in the user interface of the existing web-based CORAS tool. These deficiencies will be further explained in Section 4.4. An unusability user interface may make users feel irritated when using it. As a result of that, the user interface of the existing tool must be optimized. In addition, the user interface of the newly designed function should also be usability. We, therefore, chose to follow the design principles to modify the user interface of the tool and add new features to the tool. Design principles can help me look at the tool from different perspectives. I will further explain in Section 4.3 about which design principles I have used to modify the tool interface, what design principles the new features follow, and why I want to modify them. User interfaces designed according to design principles will have high usability [6].

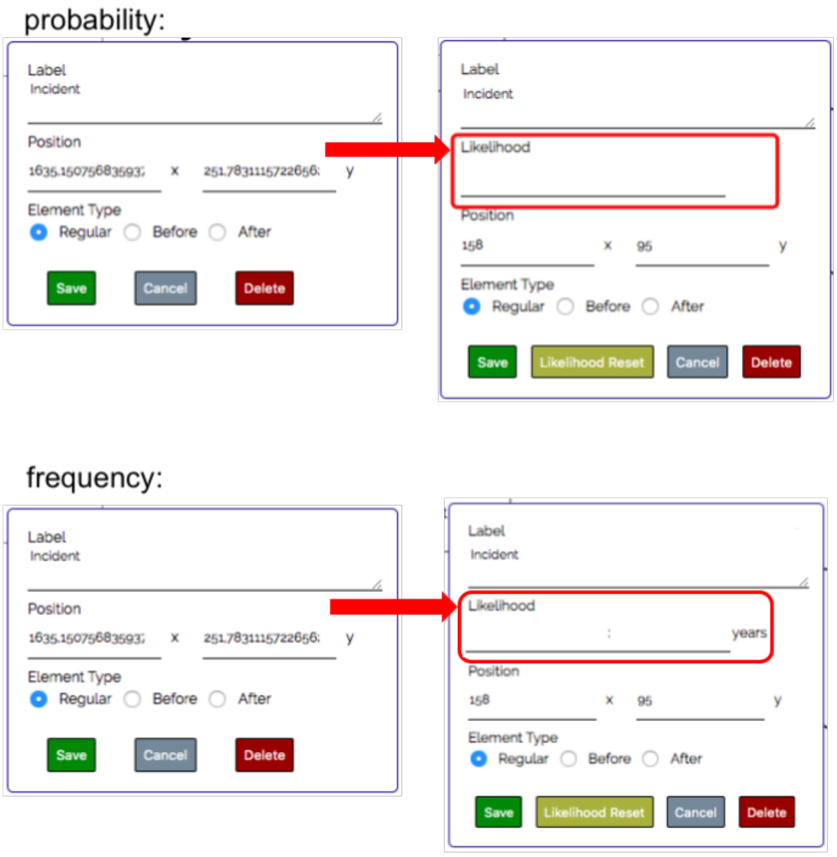
4.2 Likelihood calculation feature implementation according to CORAS approach

To make sure that my web-based CORAS tool is in line with the CORAS approach, I will implement likelihood calculation features in the tool. But before I implement the calculation algorithms, I still have some preparations to do. First, I need a calculation button to start the likelihood calculation. Secondly, I need to have an input box in the interface of the elements editor that allows users to fill in the likelihood value. Moreover, likehoods can be probability or frequency, the form of probability is a value between 0 and 1, and the form of frequency is "$N : My$", that is, an event has occurred N times in M years. Therefore, the interface of the likelihood input box should have two corresponding forms, as Figure 4.2 shows, and display the interface of different input boxes according to the content to be calculated. With this in mind, I also need to add a toggle button that can switch between probability calculation and frequency calculation. When users switch to probability calculation, the interface of the elements editor should have a likelihood input box where values can be entered, and when users click the calculation button, the calculation of the probability algorithm should be performed. Conversely, in the case of frequency calculation, the interface of the editor should be in the form of "$N : My$", and when users click the calculation button, the calculation of the frequency algorithm should be performed.

When all the above work is ready, I can start implementing the algorithm of likelihood calculation. There are two kinds of CORAS language elements to which likelihoods can be assigned, these are unwanted incidents and threat scenarios. There are also two kinds of relations between elements to which likelihoods can be assigned, which are the initiates relation and the leads-to relation. The ultimate goal of drawing CORAS diagrams is to assign likelihoods to the identified unwanted incidents. Furthermore, It is helpful in many cases to estimate the likelihood of unwanted scenarios and their relations to each other by likelihood calculation. Because firstly the threat scenarios will lead to a certain degree of unwanted incidents, and secondly, the unwanted incidents represent the risks [35]. By doing this, the estimation of the likelihood of unwanted incidents is promoted, thereby the documentation of the most important origins of risk, and to identify possible unclarity in the risk picture.

4.2.1 Rules for Calculating probability in CORAS diagram

In this section, I explain how to implement the rules in Table 4.1 for calculating and reasoning about probabilities as specified in CORAS diagrams



probability:

Label
Incident

Position
1635.15075683593; x 251.78311157226561 y

Element Type
● Regular ○ Before ○ After

Label
Incident

Likelihood

Position
158 x 95 y

Element Type
● Regular ○ Before ○ After

frequency:

Label
Incident

Position
1635.15075683593; x 251.78311157226561 y

Element Type
● Regular ○ Before ○ After

Label
Incident

Likelihood
: years

Position
158 x 95 y

Element Type
● Regular ○ Before ○ After

Figure 4.2: The interface of the likelihood input box in the element editor

into algorithms and implement them in the tool.

The first two rule are the initial rule (**Rule 13.1**) and the leads-to rule (**Rule 13.2**), the initial rule captures the semantics of the initiates relation, and the leads-to rule (**Rule 13.2**) captures the conditional likelihood semantics embedded in the leads-to relation. These two rules only involve calculations on a single branch of the threat graph.

Rules for Calculating Probability

Rule 13.1 (Initiates) For threat t and scenario/incident e related by the initiates relation, we have:

$$\frac{t\xrightarrow{p}e}{(t\sqcap e)(p)}$$

Rule 13.2 (Leads-to) For the scenarios/incidents e1 and e2 related by the leads- to relation, we have:

$$\frac{e_1(p) \quad e_1\xrightarrow{l}e_2}{(e_1\sqcap e_2)(p*l)}$$

Rule 13.3 (Mutually exclusive scenarios/incidents) If the scenarios/incidents e1 and e2 are mutually exclusive, we have:

$$\frac{e_1(p_1) \quad e_2(p_2)}{(e_1\sqcup e_2)(p_1+p_2)}$$

Rule 13.4 (Independent scenarios/incidents) If the scenarios/incidents e1 and e2 are statistically independent, we have:

$$\frac{e_1(p_1) \quad e_2(p_2)}{(e_1\sqcup e_2)(p_1+p_2-p_1*p_2)}$$

Rule 13.7 (Composing relations from mutually exclusive scenarios/incidents) If the scenarios/incidents e1 and e2 are mutually exclusive, we have:

$$\frac{e_1(p_1) \quad e_2(p_2) \quad e_1\xrightarrow{l_1}e \quad e_1\xrightarrow{l_2}e}{(e_1\sqcup e_2)\xrightarrow{\frac{p_1*l_1+p_2*l_2}{p_1+p_2}}e}$$

Rule 13.8 (Composing relations from statistically independent scenarios/inci- dents) If the scenarios/incidents e1 and e2 are statistically independent, we have:

$$\frac{e_1(p_1) \quad e_2(p_2) \quad e_1 \xrightarrow{l_1} e \quad e_1 \xrightarrow{l_2} e}{(e_1 \sqcup e_2) \xrightarrow{\frac{p_1 * l_1 + p_2 * l_2 - p_1 * l_1 * p_2 * l_2}{p_1 + p_2 - p_1 * p_2}} e}$$

Table 4.1: Rules [35] that are used for implementing probability calculation of CORAS diagrams in Web-based CORAS tool

The initial rule (**Rule 13.1**) means the probability p' of the occurrences of scenarios/incidents e due to threat t is equal to the probability p with which t initiates e. The implementation of the initiates rule is that as shown in Figure 4.3, the user can add the probability p that t causes e to the link between the two icons, Human threat accidental and Incident. For instance, $p = 0.7$ at the link, which is shown in the first figure of Figure 4.3. Then after pressing the calculation button, at Incident, there shows the probability p' that caused by Human Threat accidental, and $p' = p$, that is, $p' = 0.7$ as shown in the second figure of Figure 4.3.

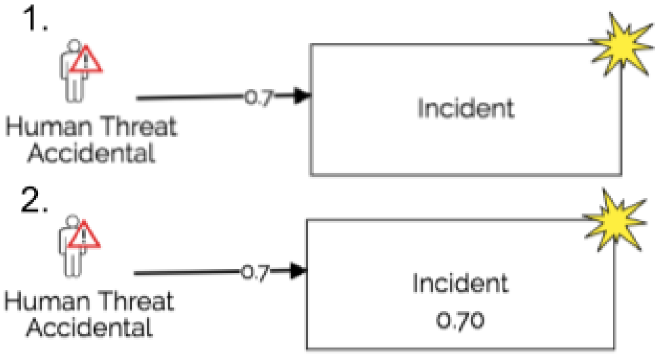


Figure 4.3: Rule 13.1, initiates of probability calculation in the CORAS diagram

While the leads-to rule (**Rule 13.2**) means the probability p' of the occurrences of e_2 that are due to e_1 is equal to the probability p of e_1 multiplied with the conditional likelihood l that e_1 will lead to e_2 given that e1 occurs. As shown in Figure 4.4, the implementation of the leads-to rule is, by multiply the known probability $p = 0.7$ at Incident 1 and the known conditional likelihood $l = 0.3$ at the link between the two incidents to calculate the probability p' at Incident 2, that is $p' = p * l = 0.7 * 0.3 = 0.21$. Thereby, after pressing the calculation button, there shows a probability value is 0.21 at Incident 2.

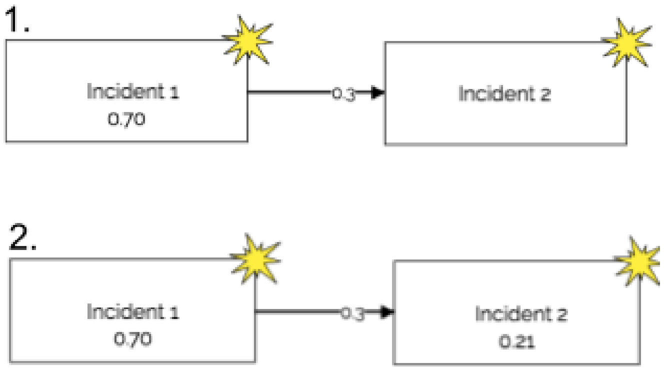
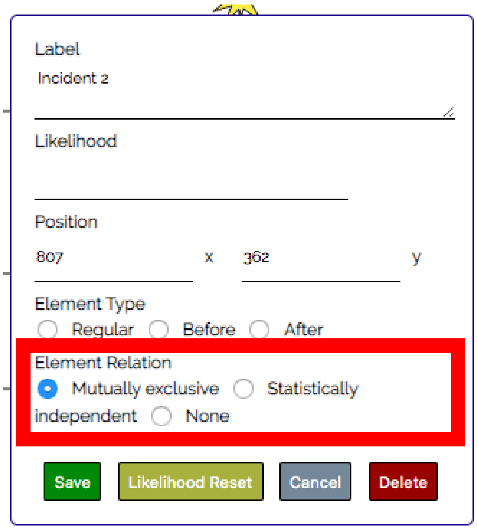


Figure 4.4: Rule 13.2, Leads-to of probability calculation in CORAS diagram

The next rules are about mutually exclusive and statistically independent scenarios/incidents or the relationship between the scenarios/incidents, which shows how to do probability calculations on parallel branches.

What mutually exclusive means, for instance, is that when throwing a die, the outcomes are mutually exclusive. In other words, two events can not occur at the same time, either $e1$ or $e2$ occurs. Only after one event has occurred can the other event occur, since they are mutually exclusive. However, statistically independent means two events can occur simultaneously and do not affect each other if they are statistically independent. Because the probability of one event occurring is independent of the probability of other events occurring. For instance, when throwing two dice to get at least one six. Throwing two dice are two events, and you can throw them at the same time, therefore, these two events are statistically independent.

Since there are two different situations on parallel branches, and every two situations there can have either mutually exclusive, independent or none of them of the scenarios/incidents or the relationship between the scenarios/incidents, this will lead to many different calculation algorithms, and I need to explain and implement one by one. The following is an introduction to what rules are used to implement different parallel branch calculations and implement the algorithm in a tool. And due to the relationship between the scenarios/incidents can be either mutually exclusive independent or none of them. Before I implement these calculations on the tool, I need the tool to be able to distinguish between mutually exclusive and statistically independent relations of the scenarios/incidents, and then the tool selects the corresponding algorithm to calculate. Of course, the tool can not know whether the two events are mutually exclusive or statistically independent through the text on the events' label users entered. So uesrs can distinguish by manual method, that is, when the users create events, they need to indicate what is the relationship between these events. As shown in Figure 4.5, I add a radio button at the element's editor interface. If the relation between two events is statistically independent, the user needs to select the statistically independent option. Conversely, if the two events are mutually exclusive, the user needs to select the mutually exclusive option. But if the relation is neither mutually exclusive or statistically independent, then the user needs to select the none option.



Label
Incident 2

Likelihood

Position
807 x 362 y

Element Type
○ Regular ○ Before ○ After

Element Relation
● Mutually exclusive ○ Statistically independent ○ None

Figure 4.5: The checkbox of element relation in the element editor

Situation 1

The first situation, as shown in Figure 4.6, it is based on the leads-to rule (**Rule 13.2**). In this situation, the calculation of the new incidents does not need to consider how it relates to other incidents. That means whether they are mutually exclusive, independent, or none of them, the calculation method uses only the leads-to rule (**Rule 13.2**). Therefore, the implementation of this situation is as same as the implementation of leads-to rule (**Rule 13.2**). The probability of incident 2 is $p_2 = p_1 * l_2 = 0.2 * 0.3 = 0.06$, and the probability of incident 3 is $p_3 = p_1 * l_3 = 0.2 * 0.6 = 0.12$.

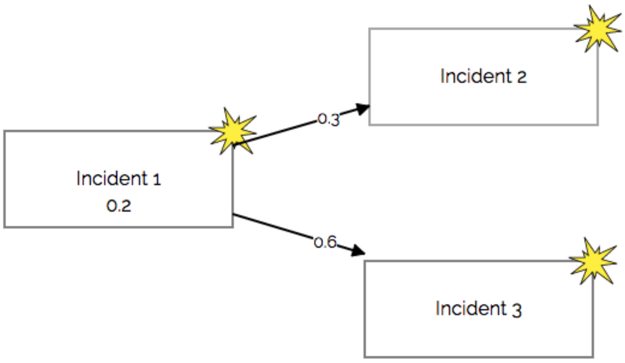


Figure 4.6: Situation 1 of parallel branch probability calculation

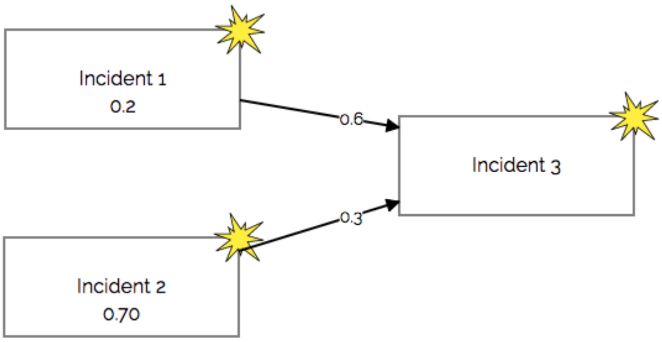


Figure 4.7: Situation 2 of parallel branch probability calculation

Situation 2

As shown in Figure 4.7, the second situation is to calculate the probability of occurrence of the third incident through the probability of the two known incidents and the conditional likelihood between the two incidents and the third incident.

To calculate this situation, I have two approaches. The first approach is for when two events are mutually exclusive. It needs to first calculate the conditional likelihood l that incident 1 or incident 2 will cause incident 3 to occur by **Rule 13.7**. Then use **Rule 13.3** to calculate the probability of the union of incident 1 and incident 2 which is equal to the sum of their respective probabilities. Finally, use **Rule 13.2** to calculate the probability of incident 3. In other words, by combining **Rule 13.2**, **Rule 13.3**, and **Rule 13.7** I get the final calculation formula : $p_3 = p_1 * l_1 + p_2 * l_2$. Turning formulas into algorithms into the tool, the first step is to calculate separately the probability $p1'$ and $p2'$ of the occurrences of incident 3 that are due to incident 1 and incident 2, that is

$p'_1 = p_1 * l_1 = 0.2 * 0.6 = 0.12$ and $p'_2 = p_2 * l_2 = 0.7 * 0.3 = 0.21$. Then store the calculated values in an array list. The second step begins when the probability of all incidents leading to incident 3 has been calculated. The second step is to calculate the probability p' of incident 3. By going through the array list and adding up all the incidents' probability that leading to incident 3, the sum of the values finally obtained is the probability p' of incident 3. To explain it mathematically, that is $p_3 = p_1 * l_1 + p_2 * l_2 = p'_1 + p'_2 = 0.12 + 0.21 = 0.33$.

However, when two events are statistically independent, it needs another approach to calculate the probability of incident 3. This approach also needs to first calculate the conditional likelihood l that incident 1 or incident 2 will cause incident 3 to occur, but it is through **Rule 13.8**. Then use **Rule 13.4** to calculate the probability of the union of incident 1 and incident 2 which is equal to the sum of their individual probabilities minus the product of these. Finally, it will be the as same as mutually exclusive again to use also **Rule 13.2** to calculate the probability of incident 3. Different from mutually exclusive is that the statistically independent calculation formula is by combining **Rule 13.2**, **Rule 13.4**, and **Rule 13.8**, then I get: $p_3 = p_1 * l_1 + p_2 * l_2 - (p_1 * l_1 * p_2 * l_2)$. Turning formulas into algorithms into the tool, the first and the second step is as same as the first two steps in the algorithm establishment of mutually exclusive. Please refer to the introduction in the previous paragraph. After these two steps, the third step is to multiply all the probabilities that are stored in the array list, $p'_1 * p'_2 = 0.12 * 0.21 = 0.0252$. Finally, subtract the multiplied product from the sum of the additions, and the final value obtained is the probability that incident 3 will occur when the relationship between the two events is statistically independent. To explain it mathematically, that is $p_3 = p_1 * l_1 + p_2 * l_2 - (p_1 * l_1 * p_2 * l_2) = p'_1 + p'_2 - p'_1 * p'_2 = 0.12 + 0.21 - 0.12 * 0.21 = 0.33 - 0.0252 \approx 0.30$.

In addition, if the two events are neither mutually exclusive or statistically independent. The probability of incident 3 would be an interval. The tool may calculate the maximum and minimum value. The assumption that there are no other scenarios than incident 1 and incident 2 that may lead to incident 3. It follows that the probability of incident 3 can not be higher than the sum of the probabilities of incident 1 and incident 2. Then by applying leads-to rule (**Rule 13.2**) for each of these, I obtain the probability that incident 1 leading to incident 3, $p_1 * l_1 = 0.2 * 0.6 = 0.12$ and the probability that incident 2 leading to incident 3, $p_2 * l_2 = 0.7 * 0.3 = 0.21$. This means that the maximum probability of incident 3 is $0.12 + 0.21 = 0.33$. Furthermore, the minimum probability of incident 3 can definitely not be lower than the maximum of the probabilities of incident 1 and incident 2. This means that the minimum probability of incident 3 is 0.21. Approximating the maximum and minimum values, the probability of incidents 3 is therefore in the interval $[0.21, 0.33]$.

4.2.2 Rules for calculating frequency in CORAS diagram

In this section, I explain the calculating and reasoning about frequencies as specified in CORAS using the rules in Table 4.2, and implement the rules in the tool.

Rules for Calculating Frequency

Rule 13.9 (Initiate) For threat t and scenario/incident e related by the initiates relation, we have:

$$\frac{t \xrightarrow{f} e}{(t \sqcap e)(f)}$$

Rule 13.10 (Leads-to) For the scenarios/incidents e1 and e2 related by the leads- to relation, we have:

$$\frac{e_1(f) \quad e_1 \xrightarrow{l} e_2}{(e_1 \sqcap e_2)(f * l)}$$

Rule 13.11 (Mutually exclusive scenarios/incidents) If the scenarios/incidents e1 and e2 are mutually exclusive, we have:

$$\frac{e_1(f) \quad e_2(f)}{(e_1 \sqcup e_2)(f)}$$

Rule 13.12 ((Independent scenarios/incidents) If the scenarios/incidents e1 and e2 are separate and statistically independent, we have:

$$\frac{e_1(f_1) \quad e_2(f_2)}{(e_1 \sqcup e_2)(f_1 + f_2)}$$

Table 4.2: Rules [35] that are used for implementing frequency calculation of CORAS diagrams in Web-based CORAS tool

A large part of the implementation of frequency calculation is similar to the implementation of probability calculation. Especially for the implementation of the initial rule and the leads-to rule, you can almost refer to the implementation of probability initial rule and leads-to rule introduced in the previous Section 4.2.1. I just need to replace the probability with the frequency to calculate it. One thing to note is that the format of the frequency and the format of the probability are slightly different. The probability p is a value between 0 and 1, while the frequency f has the form $N : My$, which N and M are both positive integers, and y is a unit that means how many years. Therefore, in the operation, Only N is required to participate, and M remains unchanged. For instance, as shown in Figure 4.8, in the calculation using the leads-to rule, the frequency f' of Incident 2 is calculated by multiplying the known frequency $f = 5 : 1y$ at incident 1 and the conditional likelihood $l = 0.8$ at the link between the two incidents, that is $f' = f * l = 5 : 1y * 0.8 = 4 : 1y$.

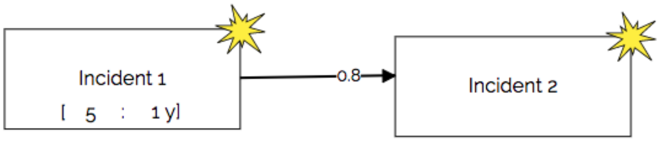


Figure 4.8: Rule 13.10, Leads-to of frequency calculation in CORAS diagram

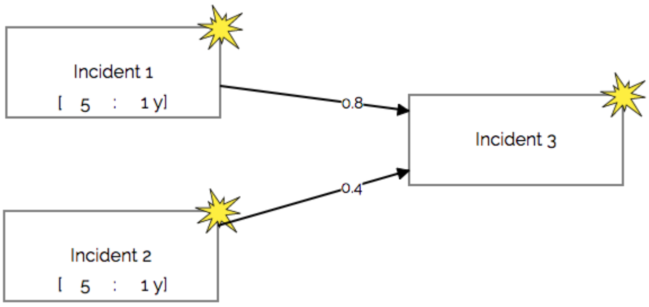


Figure 4.9: Situation 2 of parallel branch frequency calculation

Compared to probability, it is relatively simpler to implement the calculation of the frequency of mutually exclusive events and statistically independent events. However, it is also divided into three cases. The first and last situation are almost the same as the probability calculation, so I will not introduce them here. The only difference is the second situation. As shown in Figure 4.9, when the two events are mutually exclusive, **Rule 13.11** and **Rule 13.10** need to be used to calculate the frequency. Due to the two events are mutually exclusive, the two events cannot occur together, either incident 1 or incident 2 occurs. So no matter which incident occurs, they will happen with the same frequency, that is, $f_1 = f_2$. But the conditional likelihood l_1 and l_2 that the two events lead to the third event may be different. Therefore, in the calculation process, the first step is to calculate the frequencies of the third event caused by incident 1 and incident 2 separately, $f_1' = f_1 * l_1 = 5 : 1y * 0.8 = 4 : 1y$ and $f_2' = f_2 * l_2 = 5 * 0.4 = 2 : 1y$. Then select the maximum value from these two values. This maximum value is the frequency of the occurrence of the third event, $f_3 = 4 : 1y$. In addition, when the two events are statistically independent, the rules, **Rule 13.12** and **Rule 13.10**, to be used for the implementation of the calculation are almost the same as the **Rule 13.2** and **Rule 13.3** that are used to calculate the probability of the mutually exclusive events, except that the probability is replaced by frequency. To explain it mathematically, that is, $f_3 = f_1 * l_1 + f_2 * l_2 = (5 : 1y * 0.8) + (5 : 1y * 0.4) = 6 : 1y$. The detailed algorithm implementation can also refer to the mutually exclusive events in situation 2 of Section 4.2.1.

4.2.3 Generalisation to Intervals and Distributions

So far I have explained how to use CORAS diagrams to reason likelihood, and how I implement likelihood calculation in the web-based CORAS tool. But until now all likelihoods used for calculations have been exact. However, when conducting a risk analysis in practice, I am often forced to use intervals rather than exact values [35], especially when the likelihood is frequency. Because the frequency at which an event occurs is volatile, an exact frequency is less accurate than an interval. In the following, I explain how to modify the likelihood calculation from the exact likelihood calculation to interval likelihood calculation.

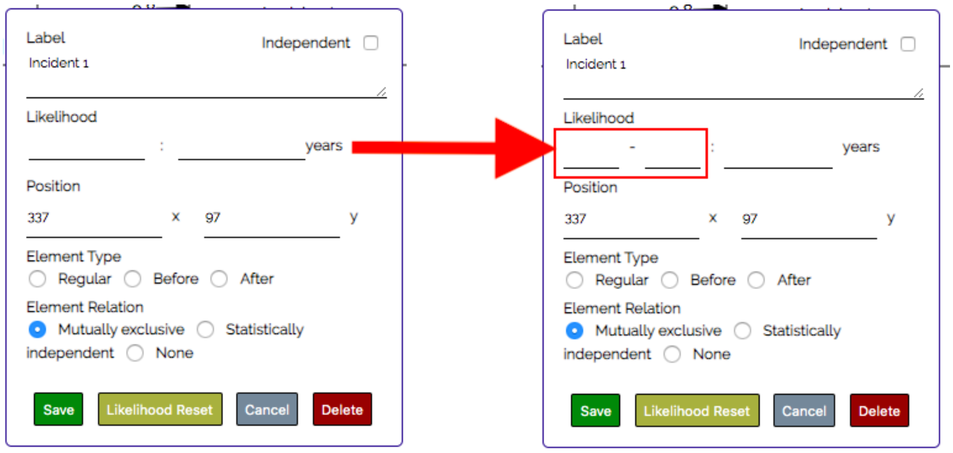
There is a definition, **Definition 13.1**, in Table 4.3, which tells us the generalization of the calculations on exact values captured by the rules to calculations on intervals. By the definition, there is no change in the rules and formulas of the calculation of the likelihood when using intervals that are understood as maximum and minimum values. During the calculation, the original single exact likelihood value is turned into an uncertain likelihood value which is an interval value. The interval value has a maximum value and a minimum value. The likelihood calculation for the interval value is then to use these two values at the same time to operate in the same formula. In other words, the maximum value of one incident is operated with other incidents' maximum values, and the minimum value is operated with other minimum values, and the two values do not interfere with each other.

However, before modifying the likelihood calculation from the exact likelihood calculation to interval likelihood calculation, I need to modify the graphical interface of the elements' editor. As shown in Figure 4.10, the modified interface is an interval that enables the user to type in the minimum

Definition for Calculating interval Frequency

Definition13.1 $[min1, max1]\ op\ [min2, max2] = [min1\ op\ min2,\ max1\ op\ max2]$, where op is one of $+$ (addition), $-$ (subtraction), and $*$ (multiplication).

Table 4.3: Definition [35] that is used for implementing invertal frequency calculation of CORAS diagrams in Web-based CORAS tool



Label: Incident 1 | Independent □
Likelihood: ____ : ____ years
Position: 337 x 97 y
Element Type: ○ Regular ○ Before ○ After
Element Relation: ● Mutually exclusive ○ Statistically independent ○ None

Label: Incident 1 | Independent □
Likelihood: ____ - ____ : ____ years
Position: 337 x 97 y
Element Type: ○ Regular ○ Before ○ After
Element Relation: ● Mutually exclusive ○ Statistically independent ○ None

Figure 4.10: The interface of the elements' editor from exact likelihood calculation to interval likelihood calculation

and maximum values. After the interface modification, the algorithm for interval calculation is implemented as described above, the maximum value and the minimum value are operated simultaneously with the same calculation formulas. Suppose that under the leads-to rule, the frequency of incident 1 is $f_1 = 4 - 10 : 1y$, and conditional likelihood $l = 0.5$ as shown in Figure 4.11. The calculation for the minimum frequency value of incident 2 is $4 * 0.5 = 2$ and its maximum value is $10 * 0.5 = 5$. Therefore, the frequency of incident 2 occurs that is due to incident 1 is $f_2 = f_1 * l = 4 - 10 : 1y * 0.5 = 2 - 5 : 1y$.

According to definition 13.1 and the calculation implementation of the leads-to rule, I can deduce the rest of the interval calculation implementations in other situations and rules from this. For instance, when two events, incident 1 and incident 2 are statistically independent. Frequency of incident 1 is $f_1 = 4 - 10 : 1y$ and conditional likelihood is $l_1 = 0.5$, frequency of incident 2 is $f_2 = 5 - 15 : 1y$ and conditional likelihood is $l_2 = 0.8$ as Figure 4.12 shows. Then the frequency of the third event, incident 3, can be calculated as $f_3 = f_1 * l_1 + f_2 * l_2 = (4 - 10 : 1y * 0.5) + (5 - 15 : 1y * 0.8) = 2 - 10 : 1y + 4 - 12 : 1y = 6 - 22 : 1y$

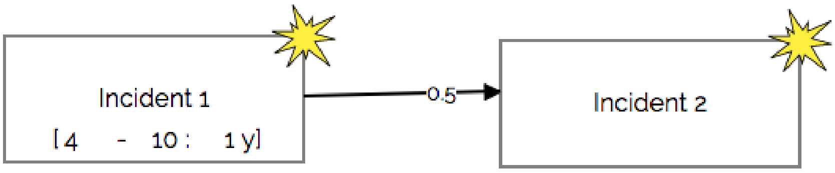


Figure 4.11: Interval frequency calculation

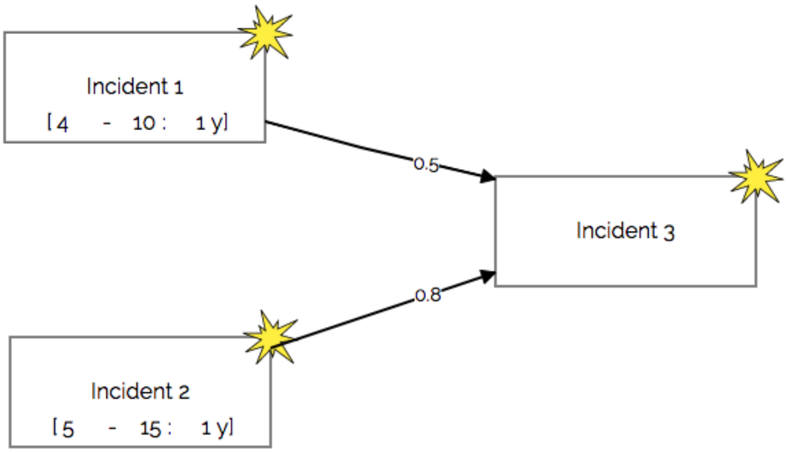


Figure 4.12: Situation 2 of parallel branch interval frequency calculation

4.2.4 Using CORAS Diagrams to Check Consistency

Now the tool can help users using CORAS diagrams that they draw to calculate the likelihood. Next, I implement features that help users check the consistency of the likelihood assignments. In other words, considering the semantics of CORAS diagrams, whether they make sense. More specifically, the tool can check whether the likelihoods assigned to the various parts of a CORAS diagram are consistent with each other. Therefore, the task now is for the tool to check whether the estimates that have been obtained are consistent with each other, instead of calculating missing values from the values assigned to other elements of the diagram.

For a given threat scenario/unwanted incident $e(l)$, where l is the likelihood of assignment. Then the tool uses calculation rules to deduce $e(l')$ based on the likelihoods assigned elsewhere in the diagram. After that, the tool compares the two likelihoods l and l'. It is equivalent to checking consistency. The requirements for consistency depends on whether the likelihoods are exact or given as intervals, and on the completeness of the diagram. In the following, I base on guidelines for consistency checking of likelihoods in Table 4.4 to first discuss how I implement the checking feature of the consistency of the complete diagrams, and then the consistency of the incomplete diagrams.

How to check consistency of likelihoods in CORAS diagrams

Exact values in complete diagrams
Assigned value: $e(l)$
Calculated value: $e(l')$
Consistency check: $l = l'$

Exact values in incomplete diagrams
Assigned value: $e(l)$
Calculated value: $e(l')$
Consistency check: $l \geq l'$

Intervals in complete diagrams
Assigned interval: $e([l_i, l_j])$
Calculated interval: $e([l'_i, l'_j])$
Consistency check: $[l'_i, l'_j] \subseteq [l_i, l_j]$ or, equivalently $l_i \leq l'_i$ and $l_j \geq l'_j$

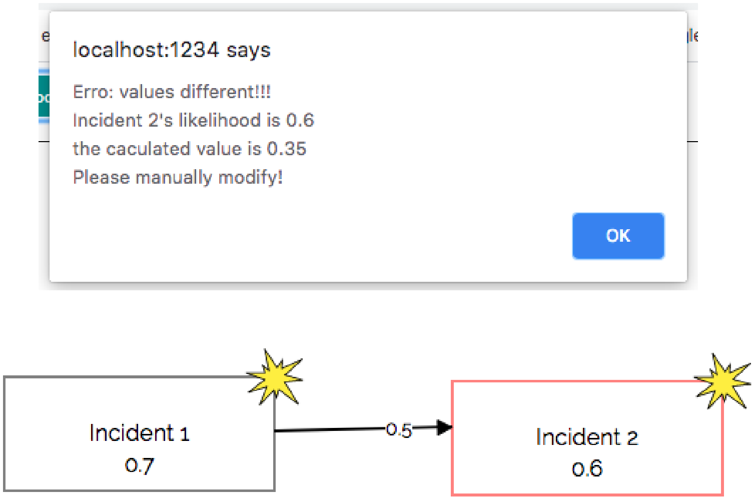
Intervals in incomplete diagrams
Assigned interval: $e([l_i, l_j])$
Calculated interval: $e([li', l'_j])$
Consistency check: $l_j \geq l'_j$

Table 4.4: Guidelines for consistency checking of likelihoods [35]

The consistency of the complete diagrams

According to Table 4.4, for exact likelihoods in the complete diagrams, the consistency requirement is that the assigned likelihood l is equal to the likelihood l' that can be deduced from the other parts of the diagram, $l = l'$. Here the checking feature of the tool is implemented like this: If a scenario/unwanted incident e is not assigned a likelihood value, use the calculation rules to deduce its likelihood based on the likelihoods assigned elsewhere in the diagram. But if likelihood l is given, then use the calculation rules to deduce e(l') likelihood. Then compare if the two likelihoods l and l' are equal. If the two values are not equal, the tool automatically prompts what went wrong and why, and the border of the wrong scenario/unwanted incident will turn red, as shown in Figure 4.13. If l and l' are equal, nothing will happen.

For likelihood intervals in the complete diagrams, the consistency requirement is that the assigned interval $[l_i, l_j]$ of the threat scenario/unwanted incident e is wider than the interval $[l'_i, l'_j]$ that is deduced from the parts that precede scenario/unwanted incident e in the diagram. This means that the tool needs to check whether the assigned minimum value l_i is less than or equal to the deduced minimum value l'_i, and whether the assigned maximum value l_j is greater than or equal to the deduced maximum value l'_j, that is, $[l'_i, l'_j] \subseteq [l_i, l_j]$. Therefore, When the tool checking the likelihood values, if the assigned interval $[l_i, l_j]$ is not wider than the deduced interval $[l'_i, l'_j]$, the tool prompts error message has almost the same form as Figure 4.13 shows.



localhost:1234 says

Erro: values different!!!
Incident 2's likelihood is 0.6
the caculated value is 0.35
Please manually modify!

Figure 4.13: Error message

The consistency of the incomplete diagrams

In cases of incompleteness, there are one or more threat scenarios and/or unwanted incidents that can occur as a result of some other threats or threat scenarios that are not described in the diagram. As a result, the assigned likelihoods can never be smaller than those deduced in the diagram. Therefore, as shown in Table 4.4, for exact likelihoods in incomplete diagrams, the consistency requirement is that the assigned likelihood l to a threat scenario or unwanted incident e is equal to or higher than the likelihood l' that can be deduced from the other parts of the diagram. While for likelihood intervals in incomplete diagrams, the consistency requirement is that the highest value l_j of the assigned interval $[l_i, l_j]$ of e is equal to or higher than the highest value l'_j of the deduced interval $[l'_i, l'_j]$.

However, if I also want to implement the consistency checking of likelihood when the diagram is incomplete in the tool at the same time, it would confusing to check the consistency of likelihood for the completed diagram. Because the requirement of consistency checking in the completed diagrams is more restricted. The requirement of the completed diagrams is if and only if $l = l'$, the diagram is consistent. While the requirement of the incompleted diagrams is if $l = l'$ and $l > l'$, the diagram is consistent. If the tool should meet both requirements, this makes it impossible to judge whether the likelihood is consistent in a completed diagram when $l > l'$. Therefore in my tool, I assume that the diagrams are checked are always completed diagrams. Thus, the tool only checks whether the requirements of the completed diagrams are met.

4.3 User interface optimization based on the design principles

Here, I explain how I can use these design principles that I have introduced in Section 2.1.3 to re-engineer the tool, and what usability the re-engineered tool can provide.

Affordance & Consistency

According to the definition of each design principle introduced in Section 2.1.3, we know that both affordance and consistency can promote the comprehensibility and learnability of an interface. Since the interface is consistent, users only need to learn a single mode of operation for all objects. And since the interface is affordance, users can know and understand how to operate from their existing knowledge. These make the product easier to learn and use, which also indirectly improves the efficiency of using the product to work.

Following these two design principles, the re-engineering I did is shown in Figure 4.14.

Following the affordance design principle, when I designed the new calculation feature, I chose to use the button to trigger it. In addition, I have also designed some other auxiliary features to help the calculation feature better, such as the likelihood value reset feature and the likelihood calculation type conversion feature, and again I choose to use buttons to trigger these features. The corresponding name is added to each button, which greatly improves the comprehensibility, learnability, and memorability of the tool, because when

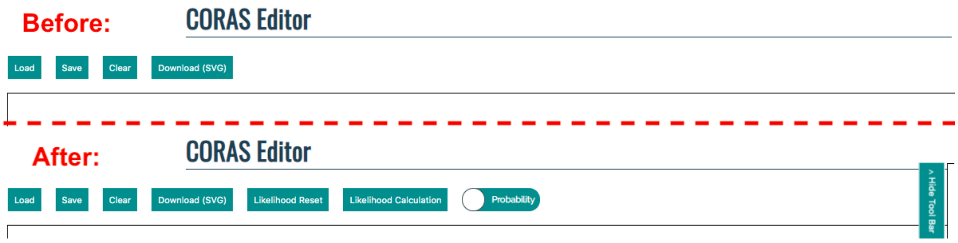


Figure 4.14: Re-engineering by following affordance and consistency design principles - Button & button color

the users see the button, they can quickly understand that they can trigger the corresponding feature by clicking on them.

Following the consistency design principle, I applied the same color as the old button for the newly added buttons. By using the same color, the users who have used the previous version can quickly understand that the newly added green rectangular boxes are also buttons. New users can also quickly understand their similarities through all the rectangular boxes with the same color. Therefore, this can also greatly improve the comprehensibility, learnability, and memorability of the tool.

Constraints

Different kinds of graphics constrain a person's interpretation of an information space [40]. For example, as shown in Figure 4.15, in the existing web-based CORAS tool, different graphics are used to represent different elements, thereby preventing users from mistakenly using one element as another element and making the semantics of graphics wrong.

In the existing web-based CORAS tool, although it uses the graphical representations of different kinds to achieve constraints, the tool does not constraint the connection between elements. According to the CORAS language grammar described in Appendix A of "Model-Driven Risk Analysis-The CORAS Approach" [35] as well as a good representation of reality, I have summarized the following points, which are important restrictions that apply to the relations:

- The symbol "Asset" can only be a source when the target is another symbol "Asset", and it can be only connected with the symbol "incident" as a target.
- The symbol "Threat" can only be a source and connect to the symbol "Threat scenario" and "Incident".

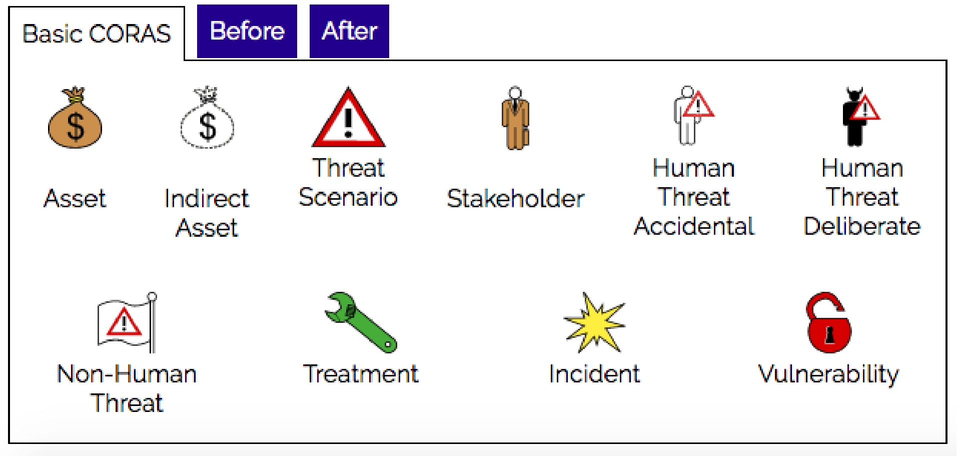


Figure 4.15: Symbols of the CORAS risk modeling language

- When the symbol "Threat scenario" is a source, it can only point to the symbol "Incident" and another symbol "Threat scenario", and it can only be a target when it is connected with the symbol "Threat" and another symbol "Threat scenario".

- When the symbol "Incident" is a source, it can only point to the symbol "Asset" and another symbol "Incident", and it can only be a target when it is connected with the symbol "Threat scenario" and another symbol "Incident".

- The symbol "treatment" can only be a source that connects with other symbols.

- The source and target of a relation cannot be the same element.

By implementing these restrictions, the constraints of the tool have also been improved, so that users can better avoid drawing diagrams that do not conform to the semantics of the CORAS language.

In addition, the toggle button to switch the likelihood value input interface and to switch the calculation method of the likelihood value is also designed following the design principle constraints, as Figure 4.2 showns. We have constrained the users to only fill in one kind of the likelihood values at a time for calculation, namely the probability value or frequency value, to avoid calculation error caused by entering the wrong value. For example, prevent the users from accidentally filling in the frequency value when they want to calculate the probability or fill in the probability value when they want to calculate the frequency.

Visibility & Feedback

As the checking feature introduced in Section 4.2.4 and as shown in Figure 4.13, I followed the design principle visibility designed by an error display feature. This feature makes elements that do not match the calculated value turn into a red border after clicking the calculation button. This allows users to see at a glance what went wrong. Then I followed the design principle feedback, also added a prompt box pop-up feature. The content of the prompt box contains: which elements are wrong, why are they wrong and what is the correct value of them. Another design that follows design principle feedback is that when a user wants to connect two elements that cannot be connected according to CORAS syntax, a prompt box will also pop up to tell the user what went wrong. In addition, as shown in Figure 4.16, a blue border will appear around a button when it is clicked, and a button turns white when the mouse hovers on it. This feature is designed according to the design principle visibility. The blue border tells us that the button was clicked without doubting.

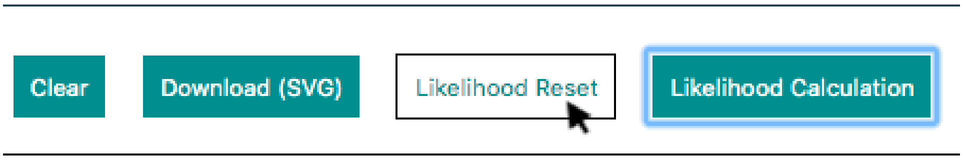


Figure 4.16: Re-engineering by following design principle visibility - Blue border and Color changing

4.4 Other re-engineering implementations

Finally, in order to further improve the usability of the tool, I also made three additional interface modifications to the tool. Through the personal trial of the tool, the observation of the students using the tool in the “IN5130 – Unassailable IT-systems” class [10], as well as discussions with them, I found that the existing zoom feature of drawing panel is very difficult to use and has two defects. Since the zoom feature is achieved by sliding the mouse wheel, this conflicts with the scrolling feature of the web page. So the first defect is that when I want to zoom in and out of the drawing panel, the web page will scroll up and down. The second defect is based on the use of a laptop touchpad to zoom in and out of the drawing panel. Because the laptop’s touchpad is very sensitive, sometimes small movements can be easily detected, and therefore this could cause the user to accidentally drastically zoom in the drawing panel when their intentions were to zoom in a small amount.

In order to avoid the existence of the first defect, I need to separate the feature of controlling the zooming of the drawing panel and the function of the scrolling web page. In order to avoid the existence of the second defect, I should choose a more controllable way to zoom in and out. So here I choose to use buttons instead of the mouse wheel to zoom the panel. As shown in Figure 4.17, I added a plus button and a minus button on the drawing panel. The user can zoom in the drawing panel by pressing the plus button and zoom out by pressing the minus button. Furthermore, since this modification also follows the design principles of affordance to modify, it can better improve the usability of the tool.

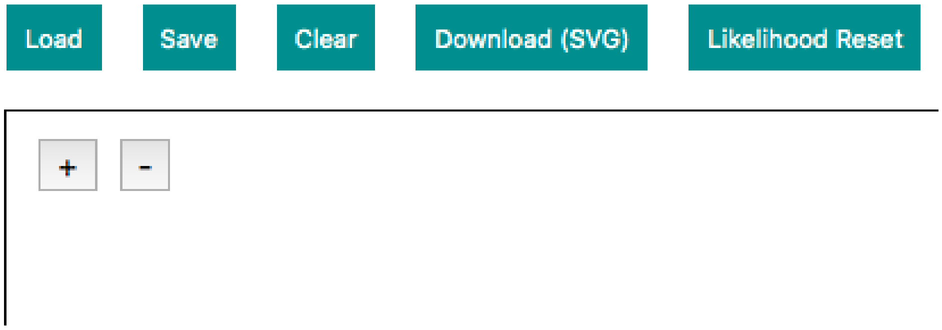
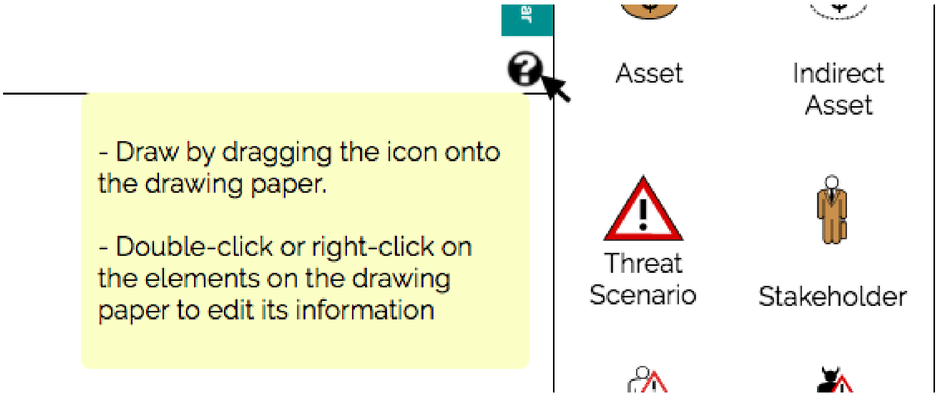


Figure 4.17: Zoom button for the zoom feature

After that, I found that for people who are first time users of the tool, they can not quickly understand how to use the elements in the toolbar to draw a diagram on the drawing panel and how to modify the elements' information. To this end, I added a tip feature, a tip icon beside the toolbar as shown in Figure 4.18, to help novices to master the use method faster. Users only need to put the mouse over this tip icon, the tool usage method will pop up to tell the users how to use it. The last interface modification is inspired by the description of the Eclipse-based CORAS tool in the book [35]. As mentioned in the book in order to get more space for the diagrams, they make the most parts of the tool to be closed or hidden, then only show the drawing area [35]. This idea also happens to meet the two principles, constraints, and visibility, mentioned in the design principles. Inspired by this good idea, I implemented also the hidden feature on the web-based CORAS tool. As shown in Figure 4.19, I have added a button to control the display and hiding of the toolbar. When the users need to add new elements to the drawing panel, they can click "Show Tool Bar" to open the toolbar. If the users just want to adjust the diagram they drew, they can click "Hide Tool Bar" to hide the toolbar. The hidden toolbar makes more areas for the users to adjust the diagram and manage the elements, and makes the users can focus more on their diagram.



- Draw by dragging the icon onto the drawing paper.

- Double-click or right-click on the elements on the drawing paper to edit its information

Figure 4.18: Question mark Icon for tip feature

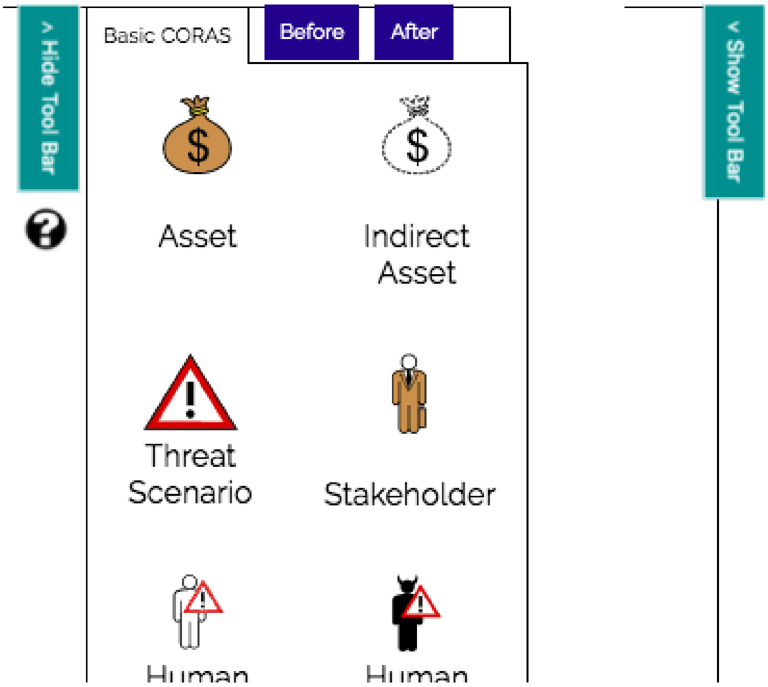


Figure 4.19: Show & Hidden feature

Chapter 5

Evaluation - Empirical Study

So far, I have introduced our re-engineering process of artifacts, which I have used to re-engineer the web-based CORAS tool. In this chapter, I outline and conduct empirical research through an experiment, which is aimed at evaluating the differences in using the original web-based CORAS tool and the re-engineered web-based CORAS tool.

As Sjøberg et al. [45] point out, empirical studies are needed to develop or improve processes, methods, and tools for software development. As in our case, the empirical study is especially concerned with the re-engineering of a tool. In this chapter, the entire evaluation process adopts the six steps of the quality improvement paradigm (QIP framework) [49]. Quality improvement paradigm (QIP) is a generic improvement cycle, which is widely accepted and considered as a recommended way to work with the improvement of software development. Its idea is to improve the quality of software based on the experience gained from previous software development projects. Furthermore, it can be also used as a framework for conducting empirical studies. For example, this framework was also adopted during the ESERNET project [32] to structure empirical studies. The ESERNET project was carried out to increase awareness within the software engineering community to systematically conduct empirical studies [49]. Figure 5.1 illustrates the six steps of the QIP framework. In the ESERNET project, it is referred to as "A Single Empirical Study" [43].

5.1 Characterisation of the Study

Phase 1 of the QIP framework [49] is characterization of the study. The objective is to understand the current situation of the project. Furthermore, at this phase, I need to establish a baseline based on past experiences and characterize their

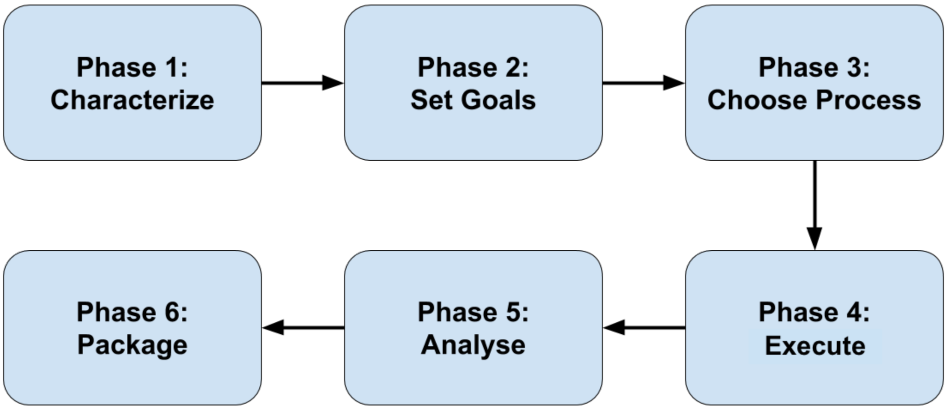


Figure 5.1: The six phases in the "A Single Empirical Study" framework [43]

criticality.

5.1.1 Current situation

My project is to re-engineer the existing web-based CORAS tool [27]. As previously introduced, the CORAS tool is a tool developed using React [21], Redux [22] and JointJS [12] to help people perform security risk analysis. The purpose of my re-engineering is to improve the usability of the tool by implementing added usability measures. The usability measures are those I introduced in Chapter 4, such as the new likelihood calculation feature, usability-oriented interface re-engineering, and so on. Then, through the implemented usability measures, study how these usability measures affect user performance considering tool-supported security risk analysis. Evaluation of this tool will, therefore, focus on the differences between the usability of the existing tool and the usability of the re-engineered tool.

In order to establish a baseline for this study, I conducted a systematic mapping study to get a high-level overview of similar studies. A systematic mapping study is a process conducted before a systematic literature review [66], and which aims to collect evidence of a topic at a higher granularity. This is an ideal strategy when there is little evidence on the subject or the subject is very broad. The focus of the mapping study involves (1) identification; (2) selection of primary studies; (3) study quality assessment [3]. I used digital libraries such as IEEE, ACM using Oria, IEEE Xplore, and Google Scholar to conduct the study. The search terms were: Comparison of AND ("systems" OR "programmings" OR "techniques") tools AND ("empirical study" OR "experiment" OR "controlled experiment" OR "quasi-experiment"). Based on the results of the systematic mapping study, I have identified the following studies.

Comparison of Different techniques

A. Perini et al. [39] conducted an empirical study to compare the accuracy of Analytic Hierarchy Proces (AHP) and Case-Based Ranking (CBRanking) Techniques for requirements prioritization. The experiment took place in a laboratory room equipped with computers. Results showed that AHP should be preferred to CBRanking in prioritization problems for which ordering accuracy is the main issue and the number of requirements is small. When the work used for prioritization is more important than ranking accuracy, CBRank-ing should be the preferred selection.

Comparison of Different tools

Anis Bey et al. [30] studied how well Algo+ could assess students' submissions in a MOOC (Massive Open Online Courses) of programming compared to EPFL (École Polytechnique Fédérale de Lausanne, Switzerland) grader by conducting an empirical study. The scoring results have shown that the EPFL grader scores according to the functionality of the submitted program, while Algo+ grader scores according to whether the program has the correct code even if the program does not run. From this, they concluded that the relation between Algo+ and EPFL grader is that both can complement each other very well.

Maurizio Leotta et al. [34] empirically investigated the difference in terms of robustness that can be achieved by adopting visual and DOM-based (Document Object Model based) locators to understand the strengths and the weaknesses of the two approaches. In addition, as a secondary goal, they investigated the cost/benefit trade-off of visual vs. DOM-based test cases for Web apps. This study involved two human subjects: a Ph.D. student (the first author of that paper) and a junior developer (the second author), and involved the software objects are six open-source Web apps. Results showed that DOM-based locators are generally more robust than visual ones. However, on specific Web apps, visual locators were easier to repair. Overall, the choice between a DOM-based locator and a visual locator is application-specific and largely depends on the expected application structure and visualization development.

Safdar S.A. et al. [42] conducted a controlled experiment with undergraduate and graduate students to compare the productivity of the software engineers while modeling with three of the well-known modeling tools. The three modeling tools are IBM Rational Software Architect(RSA), MagicDraw, and Papyrus. They measured the productivity based on the modeling effort required to correctly complete a task, learnability, time, and the number of clicks required, and the memory load required by the software engineer to complete a task. The results showed that MagicDraw performed significantly better in terms of learnability, memory load, and task integrity. In terms of time and number of clicks, IBM RSA was significantly better at modeling class diagrams and state machines than Papyrus. However, no single tool outperformed others in all the modeling tasks concerning time and number of clicks.

In Wilco J. Bonestroo and Ton de Jong [31] study, they investigated the effects of planning on task load, knowledge, and tool preference by compared two computer software tools designed to generate plans for learning. This study was conducted with first-year university students as participants. The results show that although the task load of learners using a tool where the computer generated the plan was much lower than that of a tool where learners actively created plans, and although the quality of the plans created with a tool where learners actively created plans were lower than a tool where the computer generated the plan, but they could gain more structural knowledge when they were actively involved in the planning process. The knowledge they gained would lead to an increase in their learning outcomes for the whole learning process.

5.1.2 Topic of this Empirical Study

The empirical study will try to uncover if the usability measures aid the participants' performance of security risk analysis. After describing the current state of my project and the topic of empirical study and pointing out similar experiments, I set goals for the research in Section 5.2.

5.2 Set goals

Referring to Figure 5.1, now we come to the second phase. The purpose of this phase is to define the foundation for the empirical study, so in this second phase, the aim is to Set Goals in terms of improvement. The goal is usually formulated based on the problem to be solved. Therefore, the goal and research questions that constitute the empirical study will be introduced in this section. Furthermore, according to the goal constructed as described in Goal Question Metric [47], we need to consider the following points:

- Object of study – what is studied?
- Purpose – What is the intention?
- Quality Focus – Which effect is studied?
- Perspective – From which perspective the object is studied?
- Context – Where is the study conducted?

5.2.1 Formulate the goal

I must establish a goal and research questions along with a hypothesis devised to assist my research. Having said that, the overall goal of this study is to evaluate the existing web-based CORAS tool and the web-based CORAS tool I have re-engineered to illustrate the effects of usability measures on the user performance of tool-supported security risk analysis. The re-engineered web-based CORAS tool is a tool that implemented usability measures on the existing web-based CORAS tools, which I have described in Chapter 4. The evaluation will focus on the efficiency and understandability of the tools used for security risk analysis, as well as participant satisfaction with this re-engineered web-based CORAS tool.

5.2.2 Formulate research questions

The following are the research questions addressed in this experiment:

Research question 1:

> Will the use of the web-based CORAS tool which is re-engineered for usability affect the efficiency of participants in solving the tasks provided?

Research question 2:

> Will the use of the web-based CORAS tool which is re-engineered for usability affect the participants' comprehensibility of how to use the tool?

5.3 Choose process

Now as shown in Figure 5.1 we come to the third phase, which is referred to as Choose Process. In this phase, the execution process of the empirical study will be designed in detail. This phase is divided into five steps to describe, and each step has a task, there are hypothesis statements, variables determination, subjects identification, choice of study design, and experiment material preparation.

5.3.1 Hypothesis statement

First, we have to state the hypothesis of the empirical study, this includes a null hypothesis and an alternative hypothesis. The null hypothesis is usually expressed as H_0. It states the opposite of what the experimenter predicts or expects. While the alternative hypothesis is usually expressed as H_1. It states a potential result or an outcome that the researcher may expect. Both hypotheses need to be defined to follow the goals and research questions established in Section 5.2. So we have:

Null hypothesis (H_0):

> The efficiency and comprehensibility of using the re-engineered web-based CORAS tool for security risk analysis are the same as the efficiency and comprehensibility of using the existing web-based CORAS tool for security risk analysis.

Alternative hypothesis (H_1):

> There is a difference between using the re-engineered web-based CORAS tool and using the existing web-based CORAS tool with respect to efficiency and comprehensibility for doing security risk analysis.

5.3.2 Variables determination

Variables determination is to evaluate the hypotheses established in Section 5.3.1. The task of the second step is, therefore, to identify variables for this study, and which of them includes independent and dependent variables for the experiment.

Independent variables are considered as the input to the experiment. While the dependent variables are considered as the output. In practice, the experiment is to study these dependent variables to investigate how they are influenced by the independent variables. In addition, there are other factors in the experiment, which are called confounding factors. Figure 5.2 illustrates how the independent and dependent variables and confounding factors relate to the experiment. "Confounding factors are variables that may affect the dependent variables without the knowledge of the researcher" [49]. Therefore, to give this study higher validity, it is important to identify these factors that may affect outcomes in an undesirable way and to consider the threats to the study from them. Because these factors are closely related to the threats to the validity of the empirical study.

For this experiments, the independent variables were using different CORAS tools, including the existing web-based CORAS tool and the re-engineered web-based CORAS tool. The dependent variables in the study were comprehensibility, efficiency, and satisfaction.

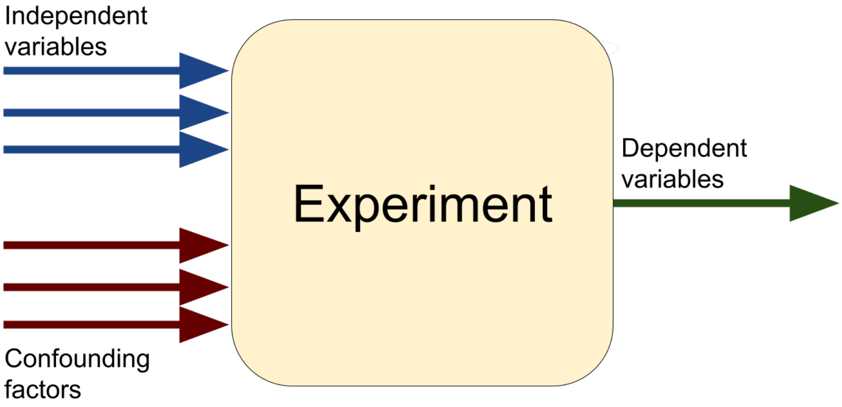


Figure 5.2: Variables in an experiment, Figure adapted from [49]

Comprehensibility is measured by learnability and effectiveness. Learnability refers to how easy a system is to learn to use, and effectiveness refers to how good a product is at doing what it is supposed to do [40]. Therefore, a system has good comprehensibility if the participants can easily learn how to use the tool and use the tool to successfully complete tasks. Furthermore, I will measure comprehensibility by taking into account the participants' scores they obtain by carrying out tasks in the experiment and the time they have spent to carry out tasks.

Efficiency refers to the way a product supports users in carrying out their tasks [40]. For this study, it means the participants are able to perform security risk analysis relatively quickly and correctly by using the CORAS tool. Considering that there is no empirical evidence that the use of CORAS tools for security risk analysis is relatively fast, it is necessary for the efficiency of the two populations to be compared.

Satisfaction refers to how pleasant it is to use the tool. After using the tool, the participants will state their subjective contentment obtained from the experience. I will also collect satisfaction through interviews or questionnaires.

5.3.3 Subjects identification

The participants in this study consisted mainly of undergraduate students and some graduates within the field of computer science. Some of them work while studying, and some are working or studying. There are 8 participants, 4 of them are graduate students and the other 4 are undergraduate students. Recruitment work is done through the researcher's own network, and the selection of participants is based on purposive sampling [44]. Purposive sampling is the selection of participants based on specific characteristics. For instance, I chose computer science students as participants. Please see Section 5.4.2 for more information about the demographics of the participants.

5.3.4 Choice of study design

As discussed in Section 5.2, the study aims to reveal differences between tools across a population. Thus, the study aims to achieve the generality of the comprehensibility and efficiency of different tools across a population. Like most of the identified articles, I also designed an experiment for this study, which is a so-called controlled experiment [49]. The controlled experiment is an experiment in which the researcher can have control over the study and how participants perform the tasks assigned to them. This is done by assigning independent variables to each control group to process the treatment of each control group. There are several available standard designs [49] in the control experiment that can be used. In this study, I chose to use the study design called Standard design 2 by Wohlin et al. Thus, I conduct experiments by assigning each control group the two versions of the web-based CORAS tools in random order. For instance, half of the participants first use the existing web-based CORAS tool and then use the re-engineered web-based CORAS tool, the other half uses the re-engineered web-based CORAS tool first and then the existing web-based CORAS tool. The reason for using treatments in different orders is that the effects of the order should be ruled out [49]. Wilco J. Bonestroo and Ton de Jong [31] compared two tools in a similar way, adopting the same design I used.

Several factors need to be considered when designing the experiment process. Keep in mind that the focus of the experiment is on the differences in using the tools, the process obviously involves interacting with the tools. Therefore, if multiple participants are required to conduct experiments at the same time, multiple computers are required. The existing web-based CORAS tool is already published online, so you can easily visit the website to use it. But the re-engineered web-based CORAS tool has not been published yet, it needs to be manually configured on the computer. If to choose field experiments, the participants want to solve the task on their own computer, the participant must set up the environment for running the webpage on their computer. This may require participants to put in more effort, so that participants may lose interest in participation. Furthermore, an error may occur while setting up the environment, which may frustrate the participants, and may require more dialogue with the researcher to properly set up. Therefore, I chose the laboratory experiment and provided participants with an environment configured computer to save unnecessary trouble as I have described above. If multiple participants are required to conduct experiments at the same time, the environment must be set up on many computers in the laboratory. Obviously, this requires more preparation and resources, because multiple machines will be needed for this purpose. However, due to insufficient resources, and plus the situation of Coronavirus 2019, people cannot gather together in large numbers, I have to choose a small number of times. This choice may lead to challenges in terms of time.

However, in order to gain a deeper insight into the study topic, it is necessary to establish an appropriate experimental process and keep in mind the restrictions discussed previously. For this reason, the laboratory experiment will be conducted first using observations to collect data, and then using online survey tools to conduct additional data collection through questionnaires. Later in this section, I will discuss the selection of an appropriate online survey tool. The chosen experiment process is represented in Figure 5.3.

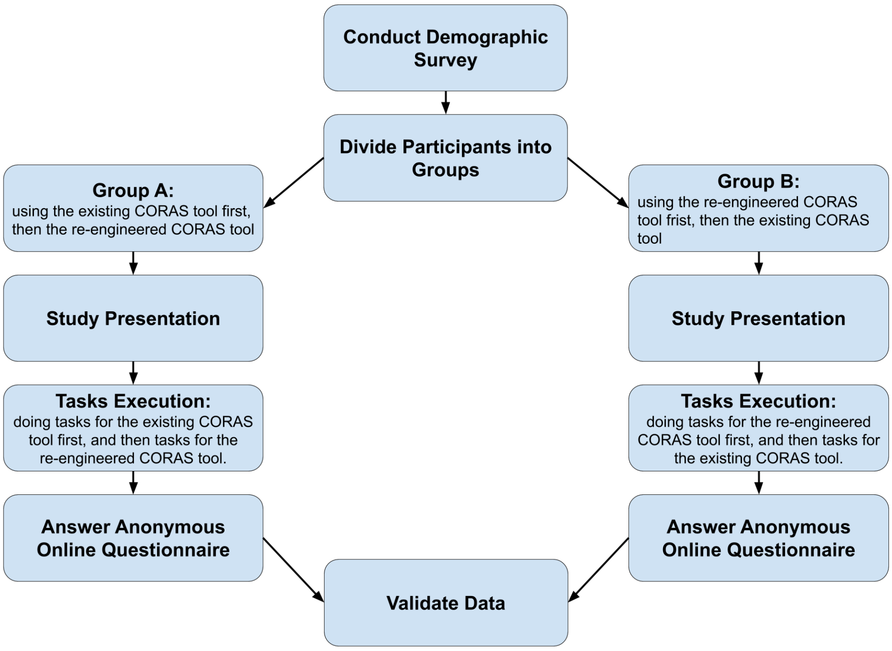

Conduct Demographic Survey
Divide Participants into Groups
Group A: using the existing CORAS tool first, then the re-engineered CORAS tool
Group B: using the re-engineered CORAS tool frist, then the existing CORAS tool
Study Presentation
Study Presentation
Tasks Execution: doing tasks for the existing CORAS tool first, and then tasks for the re-engineered CORAS tool.
Tasks Execution: doing tasks for the re-engineered CORAS tool first, and then tasks for the existing CORAS tool.
Answer Anonymous Online Questionnaire
Answer Anonymous Online Questionnaire
Validate Data

Figure 5.3: The experiment process

First, I collect information about each participant, such as educational background, whether they have used the CORAS tool, and so on. This provides some help for the subsequent data analysis. After the Demographic Survey, the participants will be divided into two groups, Group A and Group B. Group A first uses the existing CORAS tool to do the task, and then uses the re-engineered CORAS tool. Group B uses the re-engineered CORAS tool first, then the existing CORAS tool. After dividing the participants into groups, I provided a presentation, intended to give an introduction to participants about what they need to complete. In addition to the presentation, participants were provided with the main tasks to be solved and questionnaires. Finally, review each participant's response and validate the collected data to ensure that it can be used as part of the statistical analysis.

Ethical Considerations

When collecting research data while conducting an empirical study that involves human participants, privacy issues is one of the most important that must be considered. Therefore, you have to contact the data protection official, which depending on where you conduct the study, for research in the country. The place this study is conducted is in Norway and the appropriate institution in Norway is the Norwegian center for research data (NSD) [17].

On the NSD website, I conducted a test [18] about whether I need to submit a notification form to NSD. In this experiments, I do not process either directly or indirectly identifiable personal data [11]. The online survey tool used to conduct the demographic survey does not store any email address/IP address or link key. Therefore, the research project does not need to submit a notification form to NSD. But it is important to ensure that all data must be anonymous. This means that any information contained in anonymous data material must never directly, indirectly, or through an email/IP address or link key to identifying individuals.

Survey Tool Selection

When choosing a suitable survey tool, there are a lot of things to consider. Does the tool provide all the needed features, such as investigation logic, timer function, file upload function, etc? Moreover, as mentioned earlier, when conducting a demographic survey, the survey tool needs to store personal data about the participants. If one uses an external data processor, one must sign a "Data processor agreement, which regulates and ensures the processing of personal data." [19]. In order to circumvent this issue, I need to find a survey tool that can remain anonymous. Moreover, the survey tool should also have a timer function to help to count the time the participants spent on each task.
Among the identified survey tools were:

- Nettskjema [16]
- SurveyMethods [24].

- SurveyMonkey [25].
- Eval&Go [7].
- LimeSurvey [13].
- Google Forms [8].
- SurveyHero [23].

In the end, my choice is the Eval&Go tool. Because this tool has all the required features, but the only disadvantage is that it cannot view the individual time per question. However, it can record and view the average time spent on each page (i.e., each question). In addition, it is the only survey tool that provides students with all functions for free. Furthermore, it has the feature of submitting anonymously. This can avoid storing e-mail/IP addresses, browser information, or cookies. Also, there is no way to trace the response to a specific participant, because the survey connections obtained by all participants are the same.

5.3.5 Experiment material preparation

The materials to be prepared for the participants in advance are: (1) A demographic survey; (2) A presentation (see Appendix A and Appendix B); (3) The main tasks (see Appendix C and Appendix D) (4) A feedback survey questionnaire (see Appendix F).

Demographic survey

The Likert scale is used in my demographic survey and it has five values, for example: no knowledge, secondary knowledge, some knowledge, good knowledge, experts. Table 5.1 shows the questions for the demographic survey. These questions are mainly related to the skills and working background of the participants. In my demographic survey, I ignore the information which is indirectly identifiable personal data, such as age, gender, etc., and let users submit anonymously. In order to make participation as anonymous as possible, I also intend to limit the answers to these types of questions.

Tasks

For this experiment, the tasks are used to know the use of the tool, that is, to address the tool's comprehensibility and its efficiency in terms of conducting risk analysis. To this end, the task will be related to risk analysis. However,

Q#	Question	Answer(s)	Logic
Q1	Are you a student?	Yes **OR** No	-
Q2	Are you working?	Yes **OR** No	-
Q3	What's your job title?	Open question	If Q2 == Yes
Q4	What's your education level?	Open question	-
Q5	What is your Study Program?	Open question	-
Q6	Do you have any experience in information technology or engineering?	Yes **OR** No	-
Q7	Do you have any experience with security risk analysis?	Yes **OR** No	-
Q8	Knowledge of security risk analysis	No knowledge **OR** Minor knowledge **OR** Some knowledge **OR** Good knowledge **OR** Expert	If Q7 == Yes
Q9	Do you know the CORAS approach?	Yes **OR** No	-
Q10	Do you have used the CORAS tool?	Yes **OR** No	If Q9 == Yes
Q11	Do you have any experience with UI design or UX?	Yes **OR** No	-
Q12	Knowledge of UI design or UX	No knowledge **OR** Minor knowledge **OR** Some knowledge **OR** Good knowledge **OR** Expert	If Q11 == Yes

Table 5.1: The Questions for the demographic survey

because the focus of the experiment is to study the use of the tool rather than doing risk analysis, I do not need participants to know how they should conduct risk analysis, but let them know how to use tools to solve the tasks. For instance, participants do not need to analyze the risk of a company to draw a CORAS diagram. In this experiment, what participants need to do is to draw a CORAS diagram exactly the same according to the CORAS diagram I provide. Therefore, my task only needs to be designed according to the steps of using tools to do a risk analysis. Then I evaluate the comprehensibility and efficiency according to the time spent by the participants in each task and the scores the participants obtained.

Since the only independent variable in the study is the use of different tools, I need to strictly control other variables to remain unchanged and reduce confounding factors. To this end, I make the tasks that participants have to complete consistent when using different tools. In other words, whether the participants are using the existing web-based CORAS tool or the re-engineered web-based CORAS tool, the tasks they should perform are almost the same, as shown in Table 5.2. The only difference may be that the CORAS diagrams they are required to draw will be slightly different when they switch to another tool to continue the experiment, but the difficulty factor and complexity are staying the same. After switching tools, there is no need for the participants to use the new tool to draw the same CORAS diagram again. This is done in order to avoid that the user has mastered all aspects of the chart and thereby the measurement of the dependent variables would not be accurate.

The creation of the task set was improved after many tests and reviews. The final task set is divided into two parts. One part is the task to be completed when using the existing web-based CORAS tool, and the other part is the task to be completed when using the re-engineered web-based CORAS tool. The complexity of the tasks contained in the two parts is the same. These tasks are related to the use of tools to draw risk analysis diagrams and do the risk likelihood calculation. See Table 5.2 for a list of all tasks. Both parts of the task set contain these 10 tasks respectively. The full score of each part is 65 points. Therefore, the two parts of the task set add up to a total of 20 tasks, and the total score is 130 points. For a complete set of tasks and corresponding risk analysis diagrams, please see Appendix C and Appendix D.

Task#	**Description**	**Score**
1	According to the given figure, draw exactly the same diagram on the CORAS tool. Then download the diagram as an SVG file and named task1.svg.	20
2	Adjust the tool to calculate frequency. According to the given table, assign the likelihood value to the elements on your drawing according to the table we give. Then download the diagram as an SVG file and named task2.svg.	20

3	Calculate likelihood by using the CORAS diagram, and fill in the calculated answer of likelihoods by each element.	5
4	Download the diagram as a JSON file and named CORASdiagram.json.	1
5	Clear up the drawing panel	1
6	Upload the newCORASdiagram.json file to the tool. Adjust the diagram to make it look neat and nice. Download the adjusted diagram as an SVG file and named task6.svg.	2
7	Using the diagram to check the consistency of likelihoods. Which elements are inconsistent?	5
8	If it is inconsistent, please amend them until they are consistent. Fill in the correct answer of likelihoods for each element	5
9	Delete all likelihood values. Then download the diagram as an SVG file and named task9.svg.	1
10	How many unique buttons are there, apart from the buttons in the navigation bar?	Cat. 1-5
Total	-	65

Table 5.2: The main tasks

Feedback Survey

The feedback survey is also answered online, and it contains the questions shown in Table 5.3. I use these questionnaires to obtain feedback about the use of the tools from the participants, and it is the most direct way to understand the satisfaction of the participants with the tools and to know whether the tool is comprehensible for the participants. There are a total of 12 questions in the feedback survey questionnaire. The first eight questions that use the Likert scale have five values from strongly disagree to strongly agree, and the last four questions are the open-ended questions.

Q#	Question	Answer
1	I can quickly understand how to use the existing web-based CORAS tool, when I see its interface.	Likert scale.
2	I can quickly understand how to use the re-engineered web-based CORAS tool, when I see its interface.	Likert scale
3	Compared with the existing web-based CORAS tool, I prefer the interface of the re-engineered web-based CORAS tool.	Likert scale
4	I think the tip feature is useful, when I first used this tool.	Likert scale
5	I think the yellow "Likelihood Reset" button is useful, when I should modify one of the elements' likelihood value.	Likert scale
6	I think the dark cyan "Likelihood Reset" button is useful, when I should modify all elements' likelihood value.	Likert scale
7	The Likelihood Calculation feature is helpful for me.	Likert scale
8	Compared with the existing web-based CORAS tool, my satisfaction with the re-engineered web-based CORAS tool is	Likert scale
9	Comparing the zoom feature of the two tools, which one do you prefer, and why?	Open question
10	Do you think it makes sense of the toolbar position modification and the toolbar can hide in the re-engineered web-based CORAS tool? why?	Open question
11	What do you dislike in the re-engineered tool, and how could they be better?	Open question
12	What additional features in the re-engineered tool do you want to have?	Open question

Table 5.3: Questionnaire

Task Scores

The task scores of the task set are given based on the number of correct completed items in each task. That is, for tasks that require multiple items, a single point is given for each completed correctly item. For instance, when drawing a CORAS diagram, one point is given for each element drawn correctly, and then one point is given for each link that correctly connects two elements. Furthermore, about the likelihood value calculation of elements in the figure, a point is given for each correctly calculated likelihood value. In addition, Task 8 scores based on the correct rate of answers, i.e. the number of buttons found, see Table 5.4.

Cat.#	Description	Correct rate	Score
1	≥ 90%	The number of buttons in the tool the participants found is more than 90%.	5
2	[70%,90%)	The number of buttons in the tool the participants found is between 70% and 90%.	4
3	[50%,70%)	The number of buttons in the tool the participants found is between 50% and 70%.	3
4	[30%,50%)	The number of buttons in the tool the participants found is between 30% and 50%.	2
5	[0%,30%)	The number of buttons in the tool the participants found is between 0% and 20%.	1

Table 5.4: Score categories

5.4 Execute

As of now, I have completed the first three phases of this empirical study. In the first three stages, I have characterized the empirical study, set goals, and proposed hypotheses. Then, in order to further test my hypothesis, I designed a controlled experiment and determined various factors such as the variables involved in the experiment. When after all aspects of the first three phases are ready, the fourth phase of the empirical study can be carried out. The fourth phase is the execution phase, which is further divided into three steps: preparation, execution, and data validation. In this section, I will go through these three steps one by one.

5.4.1 Preparation of Study

In the preparation step, the subjects and the needed material are prepared for the empirical study. The study preparation for the experiment mainly includes setting up the demographic survey, the task, and the feedback survey in the survey tool Eval&Go. After setting up the survey, the forms are tested and verified according to the task set, the logical order of each question/task and the way each task is executed or each question is answered. The tool does not have a save feature and a back feature, and does not store any IP address, browser information, or cookies, so participants cannot return to previously unfinished answers and change previous answers. In addition, the participants must be informed about the intention so that I obtain their commitment in the preparation step. Finally, preparation also involved setting up the participant list, which including the demographic survey and letter of consent.

5.4.2 Execution of Study

Demographic survey

The invitations for the demographic survey were sent to all the participants by e-mail on May 13th, 2020. Until May 16, all participants had given and submitted their answers. According to the collected data, we know that among all the participants, there are 6 students, 5 of whom are master students, 1 is an undergraduate student, and there are 2 working persons. Moreover, three of the six students work while studying. Since their study programs are all related to informatics, all of them have experience in information technology or engineering. Their knowledge profiles are shown in Table 5.5. Among all the participants, there are 3 participants knew the CORAS approach, but only 2 have used its CORAS tool.

After this, I assigned them into the two groups A and B randomly, then provided them with relevant documents about the Main Questionnaire part and the Feedback Survey part.

	None	**Minor**	**Some**	**Good**	**Expert**
Security risk analysis	3	3	2	0	0
UI design or usability	3	2	2	1	0

Table 5.5: Knowledge profiles

Main tasks questionnaire & Feedback survey

The main questionnaire and feedback survey were sent to the participants using an anonymous survey link on the day after the demographic surveys were submitted, May 18th, 2020, via email. All answers were submitted before May 20th. The complete task scores are shown in the table in Appendix E. Also, the fan plot of the time spent on using different CORAS tools to solve each task, as well as the feedback from participants are shown in Appendix F.

5.4.3 Data validation of study

Since the questionnaires are executed by using a survey tool, the task of data validation is simply to export the answers via the tool and storing them in an Excel sheet. In addition, the data validation involves scoring based on data according to my scoring system discussed in Section 5.3.5.

5.5 Analysis of results

From the first phase to the fourth phase of the Empirical Study, I have completed the design of the experiment and the execution of the experiment. By executing the experiment, I collected ata, which provided input for this stage. So what needs to be done in this fifth stage is to analyze the collected data.

In the data analyzing phase, there are three tasks that must be done. The first task is to visualize the data, which is to try to understand the data better through visualization. The second task is to use descriptive statistics. Descriptive statistical data helps us better understand the nature of the data, as well as better identify abnormal or invalid data points. The last task is to determine from the analysis whether the hypothesis I have stated is accepted or rejected. This constitutes the basis for decision-making and conclusions about how to use empirical research results, including the motivation for further studies and identification of future possible improvements.

When analyzing the experimental data collected from the experiment, in order to perform good analysis, it is extremely important to look at the data from different perspectives through various methods [44]. In this experiment, the two control groups were required to use the two tools in a different order to complete the two similar task sets. To a certain extent, the order of use will affect task scores and the times it takes to complete the tasks. Moreover, using different tools will also affect task scores and task completion times. Therefore, for the order analysis, I can analyze the data from three different perspectives of the two control groups. I first analyze the total score on the two task sets by combining two task sets' total scores from Group A and B. Second, analyze the total scores on two task sets from Group A. Third, analyze the total scores on two task sets from Group B. In addition, I need to consider the sample size of my control group. Since each group of my control groups has a small sample size of only 4, the strength of my hypothesis test will be affected. As the number of samples increases, the accuracy of statistical testing will be higher. Similarly, larger sample sizes may refute data points that are interpreted as outliers in smaller sample sizes. In statistical data, an outlier is a data point that is significantly different from other data points. Its appearance is probably due to measurement errors or maybe merely extreme manifestations of the inherent random variability of the data, so it should be taken into account in statistical analysis [33]. With this in mind, we move to data visualization.

5.5.1 Data visualization

When visualizing data, it is important for us to use a technique that can appropriately represent the distribution of the data. Graphics (charts, graphs, etc.) are a good choice. Since graphics are suitable for perceiving data patterns, structures, trends, and relationships, they are an invaluable supplement to statistical analysis [36]. For instance, line charts, bar charts, histograms, pie charts, scatter plots, and box plots, these are the most common graphics in statistics. These visual representations can be used to visualize data. For my analysis, I used box plots generated by Microsoft Excel [14]. The box plot can indicate 6 different values and outliers [37]. The box is used to indicate the positions of the upper ($Q3$) and lower ($Q1$) quartiles. Lower quartiles splits off the lowest 25% of data set, while the upper quarlite splits off the highest 25% of data set. The inside of this box represents the range of the interquartile ($Q4$), which is the area between the upper ($Q3$) and lower ($Q1$) quartiles. Interquartile range ($Q4$)can be considered as the length of the interval with the "middle 50%" of the data set. The line in the box indicates the position of the median ($Q2$), which is the value in the middle of the data set. The whiskers (the two lines extending from above and below the box) extend to the extreme values of the distribution, which are the minimum and maximum values in the data set. Additionally, the points outside the box and whiskers represent the outliers.

First look at the box plot of the total score in the main questionnaire on the two task sets. The box on the left shows the distribution of the task set by using the existing CORAS tool, and the box on the right shows the distribution of task

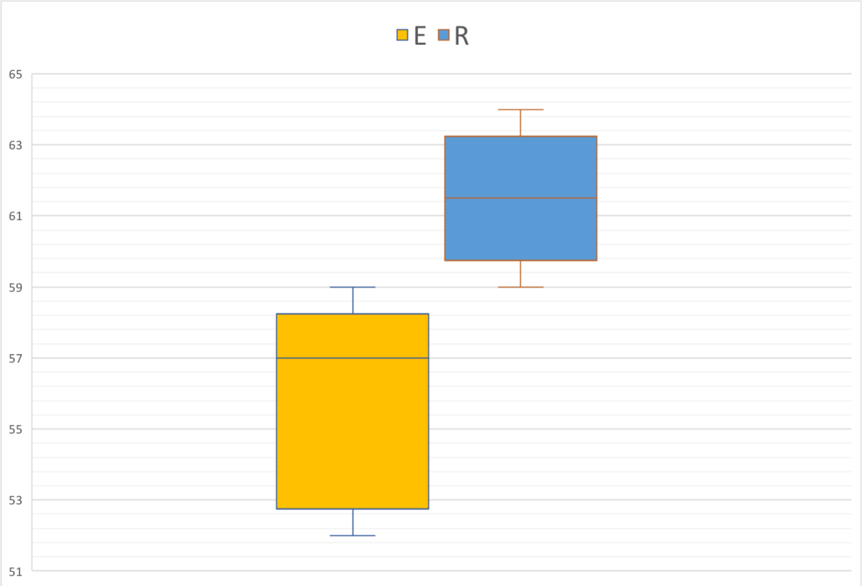


Figure 5.4: Box plot for total scores of the task set of using the existing web-based CORAS tool (E) and the task set of using the re-engineered web-based CORAS tool (R)

set by using the re-engineered CORAS tool. The plots of both task sets do not contain any outliers. This may be due to fewer subjects or maybe everything is alright. From the box plots of the two task sets, we can see that the distribution of both task sets is neither skewed towards the highest score or lowest score. The distributions are approximately normally distributed. However, In the plot of the task set by using the existing CORAS tool, the difference between $Q1$ and $Q2$ is larger than the difference between $Q2$ and $Q3$. That is, the score of this task set is generally concentrated in the smaller part. The median, maximum and minimum values of the task set by using the existing CORAS tool are 57, 59 and 52, and the median, maximum and minimum value of the task set by using the re-engineered CORAS tool are 61.5, 64 and 59. If to compare the two task sets' plots, the overall score of the task set by using the re-engineered CORAS tool is greatly better than that of the task set by using the existing CORAS tool.

Figure 5.5 shows the box plots of the total score of the two task sets within Group A. From Figure 5.5, we can see that task set by using the existing CORAS tool is skewed toward the maximum value, while task set by using the re-engineered CORAS tool is skewed toward the minimum value. I also noticed that the difference between the minimum value and $Q1$ of the task set by using the existing CORAS tool is small, the difference between $Q3$ and maximum value of the task set by using the re-engineered CORAS tool is small, and when I compare the plots of these two task sets, the values of the task set by using the re-engineered CORAS tool are always higher than the task set by using the existing CORAS tool in all aspects. In Figure 5.5, the median, maximum and minimum values of the task set by using the existing CORAS tool are 52.5, 57 and 52, and the median, maximum and minimum value of the task set by using the re-engineered CORAS tool are 62, 64 and 59.

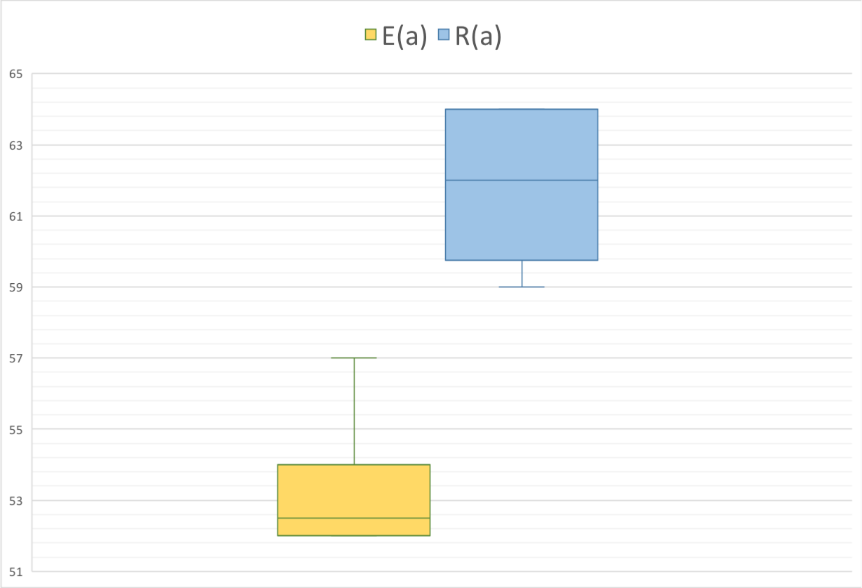


Figure 5.5: Box plot for total scores of the task set of using the existing web-based CORAS tool (E(a)) and the task set of using the re-engineered web-based CORAS tool (R(a)) within Group A

From Figure 5.6, we can see the box plots of the total score of the two task sets within Group B. Both task sets are skewed toward the minimum value, but the task set by using the re-engineered CORAS tool's deviation to the minimum is only slightly higher than the deviation to the maximum, it looks like the task set by using the re-engineered CORAS tool seems to have a normal distribution. I also noticed that the variation in both task sets between $Q1$ and $Q2$ and $Q2$ and $Q3$ look almost the same, but the difference between $Q3$ and the maximum value of the task set by using the existing CORAS tool is small. When I compare the plots of these two task sets in this instance, it seems also that the values of the task set by using the existing CORAS tool are generally higher than the task set by using the re-engineered CORAS tool. In addition, the median, maximum and minimum values of the task set by using the existing CORAS tool are 58.5, 59 and 57, and the median, maximum and minimum value of the task set by using the re-engineered CORAS tool are 61.5, 63 and 59 in Figure 5.6.

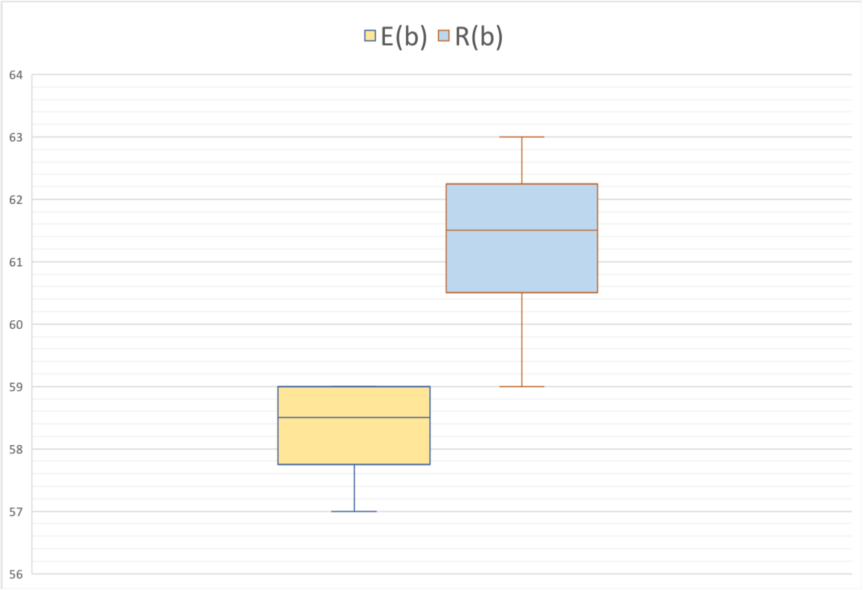


Figure 5.6: Box plot for total scores of the task set of using the existing web-based CORAS tool (E(b)) and the task set of using the re-engineered web-based CORAS tool (R(b)) within Group B

To summarise, from these graphics, it seems that the task set by using the re-engineered CORAS tool's scores are better and has an improvement over the task set by using the existing CORAS tool. I continued to apply more descriptive statistics before deciding which hypothesis testing method to use.

5.5.2 Descriptive statistics

Descriptive statistics is a general term of methods for data describing or summarizing. Descriptive statistics is to process and display the collected data in the form of charts, and then obtain data that reflect the objective fact through comprehensive summary and analysis. In the previous Section 5.5.1, I have described the data visualization. This chapter will continue to describe the statistical calculations I will use in the data analysis, such as mean, median, variance, standard deviation, percentage and so on.

Arithmetic Mean

The arithmetic mean is the sum of all collected data values divided by their number, and is also known as the average value of the data set. It is a measure of the center of data value distribution [36]. In statistics, the arithmetic mean is refined into population mean and sample mean. The population mean is the average of a data set, which is a statistical population and is consisted of every possible observation. While the sample mean is the mean of the data set that is a subset of the population. For this experiment, the data I collected is simply a subset of the population, so I call the mean of this experiment sample mean and calculate it in the following way:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + ... + x_n}{n} \tag{5.1}$$

However, the mean is highly affected by "extreme values". The extreme values are very large or very small values. If the data set contains extreme values, it is possible to "destroy" the whole picture, thereby causing misleading. Therefore, for the data sets with many extreme data values, an alternative is to use the median instead of the mean.

Median

The median is the value in the "middle" of the data set, which divides the values of the data set into two parts with an equal number of values [36]. The method to find the median is as follows:

- First sorting the data values by ascending order.
- If the size of the data value is an odd number, the median is the middle value.
- If the size of the data value is an even number, thus there is no single data value that divides data values into two equal-sized parts, the median is the average of the two central values.

People usually use the median to supplement the average, because the median is less sensitive to extreme values than the average [36].

Variance

Variance is a measure of spread, and is the average of the squared distance between the data value and the mean [36]. The units of variance are measured in square units of the original data values. For instance, if the unit of the original data values are meters, the unit of the variance of the original data are square meters. The following is the calculation formula of variance:

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{5.2}$$

where: σ^2 = variance, n = number of samples, x_i = sample i, $barx$ = sample mean.

Standard deviation

Standard deviation is the most common measure of spread [36]. It can be considered as the "mean distance" between the data values and the average. Standard deviation is the square root of the variance. The larger the standard deviation, the larger the spread of sample values distribution. The following is the calculation formula of standard deviation:

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{5.3}$$

Standard error

Standard error (abbreviated SE) is the standard deviation of the data sample distribution. It is calculated by dividing the standard deviation σ by the square root of the number of data values n. The following is the calculation formula of standard error:

$$SE = \frac{\sigma}{\sqrt{n}} \tag{5.4}$$

Skewness and kurtosis

Skewness and kurtosis are useful statistics for checking whether the data can be described by a normal distribution [36]. They provide an easy opportunity when your data visualization method cannot accurately represent the distribution.

Skewness is a measure of the asymmetry of a normal distribution [36]. If the data is right-skewed, it is called positive skewness, and if the data is left-skewed, it is called negative skewness. If the skewness is 0, it means that the data is symmetrically distributed. To know how large deviations from 0 can be accepted for the skewness for different sample sizes n, Table 5.6 is a rough guide you can use.

Kurtosis indicates how big “tails” the distribution has [36], which is how large the distribution is away from the mean. The kurtosis of the normal distribution is 0. If the kurtosis is larger than in the normal distribution, it is called positive kurtosis. A distribution with positive kurtosis is usually steeper than normal distributions at the top. If the kurtosis is smaller than in the normal distribution, it is called negative kurtosis. Contrary to a distribution with positive kurtosis, a distribution with negative kurtosis is flatter than the normal distribution at the top [36]. The deviations from 0 that can be accepted are larger for the

n	Maximum deviation for skewness
25	1.00
100	0.50
400	0.25
1600	0.12

Table 5.6: Maximum deviation for skewness that can be accepted when it is normally distributed, table adapted from [36]

n	Min. Kurtosis	Max. Kurtosis
25	-1.2	2.3
100	-0.7	1.1
400	-0.4	0.5

Table 5.7: Min. and max. kurtosis that can be accepted when it is normally distributed, table adapted from [36]

kurtosis than for the skewness. Table 5.7 shows the minimum and maximum kurtosis that can be accepted when the distribution of the data values is a normal distribution. For a given sample size that is in Tabel 5.7, if the kurtosis of the sample size exceeds the range in Tabel 5.7, the data cannot be described by a normal distribution. However, for my sample size, the range of this size is not covered in Table 5.7. Therefore, the kurtosis may be biased, but it can still indicate whether the distribution is approximately normally distributed.

Skewness and kurtosis need to be calculated comprehensively. For this purpose, Microsoft Excel [14] will be used for calculations. I apply the above calculations to my data set. The output calculation results can be found in Appendix G.

Descriptive statistics for the total score in the main questionnaire on the two task sets

Table 5.8 is the descriptive statistics of the total score of the task set by using the existing CORAS tool and the task set by using the re-engineered CORAS tool. From it, we can see that the difference in maximum values and minimum values between the two task sets are slightly large, at least 5 points. The average value is relatively close to the median in the task set by using the existing CORAS tool, and the average is equal to the median in the task set by using the re-engineered CORAS tool. However, the difference in the median and the average between the two task sets is a little larger. The median difference between the two task

	Total Score	
	Existing	Re-engineered
Minimum	52	59
Maximum	59	64
Median	57	61.5
Mean	55.88	61.5
Variance	9.27	4.29
Standard deviation	3.04	2.07
Standard error	1.08	0.73
Skewness	0.45	0
Kurtosis	-1.97	-1.79

Table 5.8: Descriptive statistics applied to the total score for the existing web-based CORAS tool and the re-engineered web-based CORAS tool

sets is 4.5 points, and the average difference is 5.62 points. A high average score indicates that participants are more efficient in doing tasks when using the re-engineered web-based CORAS tool. The variance and standard deviation of the task set by using the existing CORAS tool are higher than those of the task set by using the re-engineered CORAS tool. From the value of the standard deviation, we can know that the values of the two task sets are slightly spread. Moreover, the standard errors indicate that my sample means has a deviation of approximately 1 point in both task sets. The skewness of the task set by using the existing CORAS tool is positive, indicating that its data distribution is right-skew. The skewness of the task set by using the re-engineered web-based CORAS tool is 0 and it indicates normal distribution. The distribution of both groups is flatter than the normal distribution at the top since the kurtosis values are negative. Then considering the skewness and kurtosis values, these distributions are more similar to the normal distribution. These are consistent with the results we observed from the box plots in Section 5.5.1.

In summary, the descriptive statistics tell us that the two groups are not similar, the values of the task set by using the re-engineered CORAS tool are slightly higher than that of the task set by using the existing CORAS tool in all aspect. For instance, the measurement accuracy is higher than that of the task set by using the existing CORAS tool because the σ and SE values are lower.

Descriptive statistics for the total score on the two task sets within Group A

Now is the descriptive statistics of the total score of the two task sets within Group A. From Table 5.9, we can see that minimum, maximum, median, and

Total Score within Group A		
	Existing	Re-engineered
Minimum	52	59
Maximum	57	64
Median	52.5	62
Mean	53.5	61.75
Variance	5.67	6.92
Standard deviation	2.38	2.63
Standard error	1.19	1.31
Skewness	1.78	-0.12
Kurtosis	3.13	-5.29

Table 5.9: Descriptive statistics applied to the total score for the existing web-based CORAS tool and the re-engineered web-based CORAS tool within Group A

mean of the task set by using the existing CORAS tool are less than that of the task set by using the re-engineered CORAS tool. However, the variance of the task set by using the existing CORAS tool is larger than the task set by using the re-engineered CORAS tool. This indicates that the distribution of the task set by using the existing CORAS tool is more spread than another. Furthermore, through the positive and negative values of the skewness in Tabel 5.9, we can also confirm the results we observed from the box plots in Section 5.5.1, which are the data distribution of the task set by using the existing CORAS tool is right-skew and the data distribution of the task set by using the re-engineered CORAS tool is left-skew. The skewness of the task set by using the re-engineered CORAS tool is closer to 0, indicating an approximately normal distribution. The kurtosis of the task set by using the existing CORAS tool is steeper than the normal distribution while the kurtosis of the task set by using the re-engineered CORAS tool than the normal distribution, since the kurtosis of the task set by using the existing CORAS tool is positive and the kurtosis of the task set by using the re-engineered CORAS tool is negative.

Descriptive statistics for the total score of the re-engineered web-based CORAS tool task set between the groups

Finally, we take a look at the descriptive statistics for the total score between the two task sets. From Table 5.10, we can see that the minimum, maximum, median, and mean of the task set by using the existing CORAS tool are still inferior to that of the task set by using the re-engineered CORAS tool. However, since the σ and SE values of the task set by using the re-engineered CORAS

Total Score within Group B		
	Existing	Re-engineered
Minimum	57	59
Maximum	59	63
Median	58.5	61.5
Mean	58.25	61.25
Variance	0.92	2.92
Standard deviation	0.96	1.71
Standard error	0.48	0.85
Skewness	-0.85	-0.75
Kurtosis	-1.29	0.34

Table 5.10: Descriptive statistics applied to the total score for the existing web-based CORAS tool and the re-engineered web-based CORAS tool within Group B

tool are lower, the measurement accuracy of it is higher. In addition, we can also see that the skewness of both groups is negative, and it indicates that the distributions are left-skew. Compare to the normal distribution, the data distribution of Group A is flatter and Group B is simply steeper.

After looking at the descriptive statistics for both groups from three different perspectives, I can initially argue that the re-engineered web-based CORAS tool performs better than the existing web-based CORAS tool. Then, I will continue to use statistical methods to further answer my null hypothesis.

5.5.3 Hypothesis testing

In this experiment, I applied two experimental conditions, and each participant participated in each condition. Moreover, my sample size is small, there are only 8 participants. Therefore, the appropriate hypothesis testing method for this experiments is a paired sample t-test, also known as a dependent sample t-test [20]. I use the paired sample t-test function provided by Microsoft Excel [14]. I apply the appropriate paired sample t-test for each perspective, and the function is given by:

$$t = \frac{\bar{d}}{\sigma_{\bar{d}}} = \frac{\sum_{i=1}^{n} d_i / n}{\sigma_d / \sqrt{n}} \tag{5.5}$$

where d is differences between two paired samples and $\bar{d}$ is the sample mean of the differences. After performing this calculation, I get the so-called t-statistic. From the t-statistic, I can use the p-value and degrees of freedom in the t-distribution table to find the critical value. For my t-test, I will use a 95%

confidence interval ($p_{value} = 0.05$) and degrees of freedom (df) of 7 is obtained by the following formula given by:

$$df = n - 1 \tag{5.6}$$

where n is the size of the sample.

Next, we continue to evaluate the null hypothesis established in Section 5.3.1:

> **Null hypothesis (H_0):** The efficiency and comprehensibility of using the re-engineered web-based CORAS tool for security risk analysis are the same as the efficiency of using the existing web-based CORAS tool for security risk analysis.

The results of the paired t-test for H_0 are as follows:

- **Total Score on the two task sets:** The paired samples t-test assuming equal variances yielded values $t(7) = -5.109$, $p = 0.001$. These values indicate that there is a statistically significant effect between two task sets by using different web-based CORAS tool. Thus from this perspective, the null hypothesis H_0 is rejected.

- **Total Score on the two task sets from Group A:** The paired samples t-test assuming equal variances yielded values $t(7) = -8.716$, $p = 0.003$. These values indicate that there is a statistically significant effect between two task sets by using different web-based CORAS tool. Thus from this perspective, the null hypothesis H_0 is rejected.

- **Total Score on the two task sets from Group B:** The paired samples t-test assuming equal variances yielded values $t(7) = -7.348$, $p = 0.005$. These values indicate that there is a statistically significant effect between two task sets by using different web-based CORAS tool. Thus from this perspective, the null hypothesis H_0 is rejected.

After conducting an appropriate t-test for each perspective, I conclude that all tests report rejection of my null hypothesis. This means that, based on these results, for a given set of tasks, using the re-engineered web-based CORAS tool to complete tasks have different efficiency and comprehensibility from using the existing web-based CORAS tool.

5.5.4 Findings related to comprehensibility and efficiency

Since the Eval & Go survey tool does not provide a feature of individual time recording and the reported time spent on each task in the report is the average

Task #	$\bar{x}_e$	$\bar{x}_r$	Δt
1	334 s	284 s	50 s
2	392 s	371 s	21 s
3	728 s	382 s	346 s
4	72 s	80 s	-8 s
5	9 s	14 s	-5 s
6	113 s	108 s	5 s
7	243 s	166 s	77 s
8	209 s	476 s	-267 s
9	128 s	42 s	86 s
10	32 s	71 s	-39 s
Total	2260 s	1994 s	266 s

Table 5.11: Average time of each task in each task sets. $\Delta t = t_e - t_r$

time, outliers in time measurement cannot be identified. Moreover, since I cannot calculate the standard deviation of the time spent on each task, I cannot use a t-test to compare the means. Table 5.11 shows the average time spent on each task by using different tools, the total tasks completed, as well as the difference in time spent on each task between using different tools. I examine the average time of each task set and noticed that for all subjects, the overall time spent by using the re-engineered web-based CORAS tool to complete the tasks took is less than using the existing web-based CORAS tool. Also, in the fourth column of the table, most of the reported differences indicate that using the re-engineered web-based CORAS tool have dominant most of the time. But, for every task, the time spent by using the re-engineered web-based CORAS tool to complete the tasks are not always less. For instance, in tasks 4, 5, 8, and 10, the time spent is less when using the existing web-based CORAS. However, on the whole, compared with using the existing web-based CORAS tool, using the re-engineered web-based CORAS tool spent an average of 26.6 seconds faster on each task. These results about the time spent on each task indicate that using the re-engineered web-based CORAS tool helps participants efficiently solve the task, and does not reduce participants' comprehensibility of the tool.

Question	**Answer**
I can quickly understand how to use the existing web-based CORAS tool, when I see its interface.	Neutral
I can quickly understand how to use the re-engineered web-based CORAS tool, when I see its interface.	Half agree and half strongly agree
Compared with the existing web-based CORAS tool, I prefer the interface of the re-engineered web-based CORAS tool.	Agree
I think the tip feature is useful, when I first used this tool.	Strongly agree
I think the yellow "Likelihood Reset" button is useful, when I should modify one of the elements' likelihood value.	Agree
I think the dark cyan "Likelihood Reset" button is useful, when I should modify all elements' likelihood value.	Agree
The Likelihood Calculation feature is helpful for me.	Strongly agree
Compared with the existing web-based CORAS tool, my satisfaction with the re-engineered web-based CORAS tool is.	Strongly agree

Table 5.12: Feedback survey questionnaire answers of the first eight questions

5.5.5 Findings from the feedback survey

From the feedback survey questionnaire, I obtain the Likert values of the first eight questions in the questionnaire. From these Likert values, I found the overall degree of consent to each statement from all participants by using mode for all the values. Also, from the feedback questionnaire, I obtain the feedback answers to the last four questions. Table 5.12 shows the answers to the first eight questions and the answers to the last four questions are shown in Appendix F.

From the feedback answers to the previous eight questions, I notice that participants generally hold a neutral attitude towards the comprehensibility of the user interface of the existing web-based CORAS tool. They think they can not quickly but also not difficult to understand how to use the tool when they first time saw its user interface. However, for the comprehensibility of the user interface of the re-engineered web-based CORAS tool, participants generally think that they can quickly understand how to use this tool. For this reason, most participants prefer the user interface of the re-engineered web-based CORAS tool, and compare with the existing web-based CORAS tool, they are therefore more satisfied with the re-engineered web-based CORAS tool. Moreover, they think the calculation feature and "Likelihood Reset" button for reset all likelihood values in the re-engineered web-based CORAS tool are really useful and helpful. For the tip feature and another "Likelihood Reset" button for reset likelihood value of the current element, they think it is also quite useful.

From the feedback answers to the last four questions, I know that 50% of the participants prefer to use the mouse wheel to zoom the diagram in the drawing panel, while the other 50% of the participants prefer to use the button to zoom. They gave reasons for their preference. The reason they like the scroll wheel is that the adjustment operation on the diagram can zoom in and out quite quickly. The reason they like the button is that it is easier to control with button, and the button is more intuitive since one participant gave the feedback that "it took me a while to realize that it was possible to zoom in the existing version.". In addition, some participants reported that the scroll wheel zoom is very useful for desktop computers using a mouse, but it is not necessarily for the touchpad operation of laptop computers. For the toolbar position adjustment and added the hidden feature to it in the re-engineered web-based CORAS tool, there are 75% of the participants think it is good. The toolbar position adjustment makes it easier for them to see the toolbar and have it in view without having to scroll. The hiding feature reduces unnecessary clutter and distractions during the calculation, so that they have more convenient and a better view to see the whole diagram.

Furthermore, at the end of the feedback survey questionnaire, the participants also put forward many valuable opinions and suggestions for the re-engineered web-based CORAS tool. For instance, they think the error messages are not good enough. They want to have more detail with error messages and more warnings. They also want to have some more features on the tool, for instance, undo and redo features, an option to clear likelihood values that are generated by the tool, and so on. See these opinions and suggestions for details in Appendix F.

5.5.6 Validity concerns of results

Before presenting the results, it is important to assess the validity of the results. Therefore, in this subsection, I discuss the validity concerns of my controlled experiment and the threat of validity concerns. There are basically four categories of validity concerns in the context of software engineering that pose a threat to the validity of the results and need to be discussed [49]:

- **Internal validity** Without the knowledge of the researcher, certain factors have occurred that make the observed outcome is not the result of the given treatment, thus affecting the dependent variable. These factors are called confounding factors and have been introduced in Section 5.3.2.

- **External validity** Whether the problems the participants are engaged in are representative and whether the participants represent the target population.

- **Conclusion validity** Whether the relationship between treatments and the experimental outcome can draw a correct conclusion. Hence, it is necessary to consider the issues regarding whether the statistical power of the hypothesis test is reasonable and the reliability of the measurements.

- **Construct validity** The validity of the construction is related to the relationship between theories and observations. For example, whether the concepts are defined clearly enough before defining the measurements.

Internal validity

Regarding internal validity, the following threats are involved. First, the threat to internal validity involves introductory material. Since participants have to read it by themselves, I cannot control the extent to which participants learn the given material. This uncertainty can lead to two different situations. Participants spend more or less time learning the material than others. There is also a threat to internal validity that is the participants' proficiency and skill in security risk analysis, the CORAS approach, and the use of CORAS tools. From the demographic survey, we know that 5 participants have experience within security risk analysis, 3 participants know the CORAS approach, and 2 participants have used the CORAS tool. Although they have given a self-assessment of which level their proficiency and skills in these areas are, I cannot guarantee that someone will underestimate or overestimate themselves. Then, another threat to internal validity I face is the fatigue level of the participants. Participants can answer the questionnaire at any time they want, such as in the morning, late at night, or after work. Therefore, I cannot determine how fatigue they are and whether their fatigue levels have a great influence on the dependent variable. In addition, I also noticed the possibility of information exchange between participants. However, I have avoided this situation by keeping the participants unaware of other participants' information. Finally, in order to avoid the participants have excessive proficiency in performing the same task for the second time with a similar tool, I slightly changed the content of the drawn diagram and calculated values in the task but the difficulty factor remained the same. Moreover, let the participant use treatments in different orders is to rule out the effects of the order.

External validity

The participants in this experiment are some undergraduates and graduate students, but not everyone has experience in security risk analysis, hence such samples cannot fully represent my target population. My target population is those who will be involved in the field of security risk analysis, whether they are professionals or novices, anyway, it is the stakeholders who may eventually use CORAS tools. However, the focus of the study is on comprehensibility and efficiency in the use of tools. Therefore, this study has nothing to do with security risk analysis, and my sample does not represent people from the field of security risk analysis. The sample is composed of developers at different levels and a large part of them have work experience in IT, and some have experience in User Experience and User Interface, all of these factors are important. It can be said that they can probably provide some good feedback on the tool experience.

Conclusion validity

The comprehensibility and efficiency of theoretical construction introduce a threat to construct validity through measurement. In order to mitigate the threat, a variety of methods are used to analyze from different perspectives, such as task scores assessing, average time measuring, and feedback collecting. Moreover, there are open-ended and closed-ended questionnaires for information retrieval that used to measure these constructs. The CORAS diagrams drawing and likelihood calculation of security risk analysis in the task almost correspond to the actual situation. Finally, I did not make the subjects to perform the calculation of likelihood probability, that may introduce a threat. However, in actual situations, the calculation of likelihood frequency will be more commonly used, because the frequency is easier to measure than probability, and work best in practice [35, 41]. Therefore, my focus is on the use of the calculation feature of likelihood frequency.

Construct validity

For my hypothesis testing, I used a parametric paired samples t-test. There are four main assumptions for this testing:

- The dependent variable must be continuous.
- The observations are independent of each other.
- The dependent variable should be approximately normally distributed.

- The dependent variable should not contain any outliers.

If the dependent variable is not approximately normally distributed, a non-parametric test will be available. Furthermore, in order to mitigate an erroneous conclusion made when accepting or rejecting my null hypothesis, my t-test is conducted from multiple perspectives. In addition, I have to admit that by increasing the sample size, my conclusion will be more reliable.

5.5.7 Analysis summary

In summary, This empirical study is conducted through a controlled experiment.

In my analysis, the collected data is visualized, represented by descriptive statistics, and finally using the t-test to do the hypothesis testing. All t-tests reject my null hypothesis, which shows that the re-engineered web-based CORAS tool has a certain difference in comprehensibility and efficiency compared with the existing web-based CORAS tool. This answers the research questions I formulated in section 5.2.1. Moreover, by comparing the average time spent on each task, Research question 1 is further answered, that is, there is a noticeable difference in efficiency between the two tools. The time that participants use the re-engineered web-based CORAS tool to complete tasks is much less than use the existing web-based CORAS tool to complete tasks. Finally, the feedback survey questionnaire provides feedback on the comprehensibility and satisfaction of the tools, which can be used for future tool improvements.

After discussing the validity concerns of results, the threat to validity, and the research questions for this empirical study, I continue to discuss whether my artifact meets the success criteria in the following chapter.

Chapter 6

Discussion

In this chapter, I will re-examine the success criteria established in Section 2.3 of Chapter 2 based on our thesis project. This is part of the technology research method used in this thesis introduced in Chapter 3. After an artifact has been developed, the researcher must discuss whether the artifact meets the requirements established during the problem analysis process. Therefore, the purpose of this chapter is to discuss whether our thesis work has fulfilled the success criteria and to what extent.

6.1 Success criterion 1.

The first success criterion states: *The tool should correctly calculate the risk likelihood based on the drawn CORAS diagram.*

As mentioned in Section 4.2, I strictly follow the Likelihood calculation rules introduced in the CORAS Approach [35] to implement the likelihood calculating feature. Moreover, to ensure that the calculation, in any case, can be accurate, I used different CORAS diagrams to test and verify the calculation feature. First of all, I tested whether the calculation feature can accurately calculate the likelihood based on the initial rule. The test was conducted in two situations in a CORAS diagram, single path and multiple paths. Next is to ensure that the calculation feature can accurately calculate the leads-to rule in the single path case, and then ensure that the calculation feature can use different calculation formulas to calculate the correct value according to the different relations between the elements in the multiple paths. After testing and ensuring that both the initiates and leads-to rules are properly implemented, introduce the vulnerability element to the calculation feature, and make it so that it does not interfere with any likelihood calculation. Finally, the likelihood calculating feature was tested and verified through multiple complete CORAS diagrams with different structures.

All in all, the calculation feature so far is accurate for calculating the likelihood value in the CORAS diagram.

6.2 Success criterion 2.

The second success criterion states: *The re-engineered tool must be more comprehensible for users.*

To evaluate the comprehensibility of the artifact, an empirical study was conducted to assess whether the comprehensibility of the artifact is different from the existing web-based CORAS tool. The study was not conducted with security risk analysts as participants, but involved target groups with a background in computer science. Participants were undergraduates and graduate students, some of whom have experience in security risk analysis.

After a controlled experiment was conducted, I analyzed the collected data from multiple perspectives. All perspectives rejected the null hypothesis with a 95% confidence interval, instead accepted the alternative hypothesis. Hence, this study has concluded that there is a difference in comprehensibility between the re-engineered web-based CORAS tool and the existing web-based CORAS tool. Furthermore, through the feedback from participants in Appendix F, I found the re-engineered web-based CORAS tool can make participants better understand how to use it. Therefore, the re-engineered web-based CORAS tool has better comprehensibility then the existing web-based CORAS tool. However, the message in the re-engineered web-based CORAS tool's error warning feature is a bit difficult to understand. It would be better if the error message could be more detailed and understandable.

6.3 Success criterion 3.

The third success criterion states: *The re-engineered tool must improve the efficiency of carrying out a security risk analysis.*

The empirical study is also conducted to evaluate the efficiency of the artifact, to assess whether the efficiency of the artifact is different from the existing web-based CORAS tools. Similarly, after conducting a controlled experiment in the empirical study, I analyzed the collected data from multiple perspectives. All perspectives also reject the null hypothesis with a 95% confidence interval but accept the alternative hypothesis. Hence, the study of the artifact efficiency concluded that the efficiency between the redesigned Web-based CORAS tool and the existing Web-based CORAS tool is different.

Furthermore, the findings related to the efficiency, the average time spent report, further proves that there is a difference in efficiency between the two tools. From this report, I know that the overall average time spent when using the re-engineered web-based CORAS tool to complete tasks is 266 seconds less than using the existing web-based CORAS tool. But for some individual tasks, such as tasks 4, 5, 8, and 10, it is better to use the existing web-based CORAS tool. Tasks 4, 5, and 10 take a little longer time when using the re-engineered web-based CORAS tool. This may be that when a tool has more important features in the user interface, participants need to spend a little more time to find the corresponding features they need. From the feedback survey I know, task 10 take a longer time when using the re-engineered web-based CORAS tool, is because the error messages are not good enough for the participants to understand. However, on the whole, the re-engineered web-based CORAS tool has better efficiency.

6.4 Success criterion 4.

The fourth success criterion states: *The re-engineered tool must improve users' satisfaction with using the tool.*

At the end of the empirical study, I conducted a feedback survey to assess the satisfaction of participants with the artifact. From the participants' answers, I know that the participants were satisfied with the re-engineered web-based CORAS tool. 87.5% of the participants thought the likelihood calculation feature is extremely helpful for them, and 12.5% of the participants thought it was only quite helpful. Participants were quite satisfied with the tips feature and error reporting feature, and thought that the features can be better if the text content of the descriptions in these two features can be more detailed and easier to understand. However, with regard to the zoom feature of the drawing panel, half of the participants like using the buttons in the re-engineered version to zoom, and half like the mouse wheel in the existing version. The reasons they like buttons are because they are more intuitive and easier to control. While the reason they like the mouse wheel is the convenience and the ease with which to make big adjustments. Further, some participants suggested that it would be better if both features could be available at the same time. For the two "Likelihood Reset" buttons, participants thought that a button that can reset all likelihood values at once was more useful than the one that only resets the likelihood value of the current element. Compared with the existing web-based CORAS tool, the participants also like my modified user interface more.

All in all, I can draw a conclusion from the report of my questionnaire, by re-engineering the existing version to have more usability, I have implemented new functions and modified the user interface, which greatly improved the satisfaction of the tool.

Chapter 7

Conclusion and further work

7.1 Conclusions

With the advancement in science and technology today, computers and even computer software are playing a more and more important role in our society. A truly usable software product can be able to help us improve the quality of work and life.

In order to make the existing web-based CORAS tool easier to use, I conducted a usability-oriented re-engineering for the existing web-based CORAS tool. I added features such as calculation of likelihood value, reset of likelihood value, toolbar hiding, and tool-using prompts. In addition, I also designed and modified the user interface according to design principles. After completing the re-engineering, I conducted an empirical study. This study was aimed at evaluating the existing web-based CORAS tool and the web-based CORAS tool I have re-engineered, thereby illustrating the effects of usability measures on the user performance of tool-supported security risk analysis.

This empirical study was a controlled experiment conducted by 8 participants using the existing web-based CORAS tool and the re-engineered web-based CORAS tool to complete the given tasks. I analyzed the collected data from three different perspectives to verify the following hypotheses I established:

Null hypothesis (H_0):

The efficiency and comprehensibility of using the re-engineered web-based CORAS tool for security risk analysis are the same as the efficiency of using the existing web-based CORAS tool for security risk analysis.

Alternative hypothesis (H_1):

> There is a difference between using the re-engineered web-based CORAS tool and using the existing web-based CORAS tool with respect to efficiency and comprehensibility for doing security risk analysis.

From all perspectives, the null hypothesis I established was rejected and the alternative hypothesis was accepted. The alternative hypothesis was accepted, because this study showed that there is a clear difference between the existing web-based CORAS tool and the re-engineered web-based CORAS tool in terms of comprehensibility and efficiency.

Furthermore, this empirical study indicates that compared with the existing version, the re-engineered web-based CORAS tool is more effective and easier to understand from an usability perspective. Participants' task scores are higher when using the re-engineered web-based CORAS tool than using the existing web-based CORAS tool. When using the re-engineered web-based CORAS tool, the overall time spent in completing the task is reduced by an average of 266 seconds, which is 11.77% of a total 2260 seconds for the complete task. But for a small number of individual tasks, using the re-engineered web-based CORAS tool took a little more time. This is probably because when the tool has more features in the user interface, the time to find the corresponding function you need will increase. Moreover, parts of the features need to be improved so that users will easier understand them.

From these findings, I arrived at the following conclusions:

- Some of the usability measures will affect the user performance of the tool-supported security risk analysis, but different usability measures have different effects. For the likelihood calculation feature, it greatly improves the speed of people doing security risk analysis, while ensuring the accuracy of security risk analysis, and especially for the novice crowds. Because the novice crowds have little or no experience with the security risk analysis. Therefore, when the tool does not support likelihood calculation, the efficiency of the user to calculate manually is low, as the collected data show. Furthermore, for the security risk analysis CORAS tool, such a likelihood calculation feature is the most basic and important feature [35], and should be implemented first. In other words, no matter what software program is used, the most important usability measures are to implement and perfected the basic and important features first. Only in this way can a software program achieve the most basic usability. But not all impacts of usability measures are good. For instance, following the design principles of Visibility and Feedback, I designed the likelihood consistency checking feature and added two auxiliary features, error warning and error display features, to help the user know what went wrong. However, this did not increase the users performance, but brought more problems, and made the users confused. No one likes mistakes. Thus, mistakes are big frustration points for users and will make users deviate from their intended goals [3]. Therefore, when designing the error warning feature, it is important to consider the content of error messages to make them easy to understand. The usability of such an auxiliary feature should be tested several times before being implemented. If you cannot be sure that it is good enough, it is better not to implement it in the software program.

- Some of the usability measures will not affect user performance of tool-supported security risk analysis. For instance, I replaced the mouse wheel with buttons to make a zoom function in the diagram in the drawing panel. Some participants reacted well to this, buttons make it easier for them to understand how they can zoom the diagram. This indicates that using buttons to zoom improves the user's understanding of the tool, that is, the tool's comprehensibility is improved. Some participants reacted well to this, the mouse wheel is easier to make big adjustments with and is convenient to use. This seems to be understood as the scroll wheel is better than the buttons, especially for zooming the diagram from very small to very big. According to experimental data, it can be seen that the zooming buttons improve comprehensibility and reduce efficiency, while the mouse wheel improves efficiency and reduces comprehensibility. Therefore, for situations like this, sometimes it is best to combine two similar but different features together.

- As the users' feedback says, the usability-oriented re-engineered user interface makes them feel more intuitive and it is easier to understand how to use the tool. It also gives the users a better view of the tool and makes it easier for them to focus on security risk analysis. Therefore, in addition to implementing the most basic and important features first, it is also important to have a user interface where users easily can recognize what each feature does and how to use it.

- This findings indicate that for the target group of this empirical study, the usability-oriented, re-engineered web-based CORAS tool has been improved in terms of comprehensibility and efficiency. It is to be expected that the application of the approach described in this thesis also will improve the results for users of risk analysis programs. In order to expand the results of these findings to the field of risk analyis, an empirical study of a user group of risk analysis programs is recommended to further test the hypothesis in Section 5.3.1.

7.2 Directions for Future Work

I have identified several directions that may be of interest in future work:

- The re-engineered web-based CORAS tool can be further modified according to the user feedback I obtained in this research, and make it better and easier to use for users.

- Investigate the usability of the CORAS tool by conducting a usability study with the professional CORAS research and development staff. Such a usability study may find out the behaviors required to realize the full potential of the tool. Then further provide more valuable advice on how to perfect the tool to meet the needs of security risk testers.

- Implement a new feature, according to the translation method described in The CORAS Approach [35]. Let the tool translate the diagrams drawn in the drawing panel into English prose with respect to the already defined formal semantics of the CORAS language. This will simplify the communication of the CORAS threat model between security risk analysts.

- Software program code and architecture optimization. Optimization of code means to make the code clear and easy to understand for every developer that should work with the program. Optimization of code means to make the architecture suitable to the program. Because good and suitable architecture can reduce the maintenance costs of the program [2]. In addition, the optimization of software program code and architecture can further provide convenience and efficiency in future development and modification.

- Further needed is an investigation in what kinds of usability measures for the web-based CORAS tool are more important, and an evaluation of which of them need to be implemented first.

- Investigation on whether the user's education level and the discipline they learn will affect the comprehensibility of the tool and the efficiency of using the tool.

Acronyms

API Application Programming Interface. 7

CSS Cascading Style Sheets. 6, 7, 9

HTML Hyper Text Markup Language. 5, 6, 7, 9

IP Internet Protocola. 59

JS JavaScript. 6

JSON JavaScript Object Notation. 12, 13, 63

NSD Norwegian Center for Research Data. 59

PC Personal Computer. 7

SVG Scalable Vector Graphics. 12, 13, 62, 63

UI User Interface. 61

UX User Experience. 61

W3C World Wide Web Consortium. 6, 7

WHATG Web Hypertext Application Technology Working Group. 6

Bibliography

[1] Atomic web design. http://bradfrost.com/blog/post/atomic-web-design/. Accessed Frebruary 25, 2020.

[2] The benefits of software architecting. https://www.ibm.com/developerworks/rational/library/may06/eeles/index.html?S_TACT=105AGX52&S_CMP=cn-a-r. Accessed Jul 5, 2020.

[3] Best error messages: 5 tips for a user-friendly experience. https://freshsparks.com/user-experience-tips-best-error-messages/. Accessed Jul 4, 2020.

[4] Css(cascading style sheet). https://en.wikipedia.org/wiki/Cascading_Style_Sheets. Accessed Frebruary 25, 2020.

[5] Definition of research. http://www.merriam-webster.com/dictionary/research. Accessed Frebruary 28, 2020.

[6] Design principles. https://www.interaction-design.org/literature/topics/design-principles. Accessed March 15, 2020.

[7] Eval&go. http://www.evalandgo.com/. Accessed April 20, 2020.

[8] Google forms. https://www.google.com/forms/about/. Accessed April 20, 2020.

[9] Guidance on usability. https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-1:v1:en. Accessed Frebruary 23, 2020.

[10] In5130 – unassailable it-systems. https://www.uio.no/studier/emner/matnat/ifi/IN5130/index-eng.html. Accessed April 6, 2019.

[11] Indirectly identifiable personal data. http://www.nsd.uib.no/personvernombud/en/help/vocabulary.html. Accessed April 30, 2020.

[12] Jointjs. https://www.jointjs.com/. Accessed August 06, 2019.

[13] Limesurvey. https://www.limesurvey.org/. Accessed April 20, 2020.

[14] Microsoft excel. https://www.microsoft.com/en-us/microsoft-365/excel. Accessed May 22, 2020.

[15] Modeling language. https://en.wikipedia.org/wiki/Modeling_language. Accessed Frebruary 22, 2020.

[16] Nettskjema. https://nettskjema.no/. Accessed April 20, 2020.

[17] Norwegian centre for research data. http://www.nsd.uib.no/nsd/english. Accessed April 30, 2020.

[18] Norwegian centre for research data - will you be processing personal data? https://nsd.no/personvernombud/en/notify/notification_test.html. Accessed April 30, 2020.

[19] Online surveys. http://www.nsd.uib.no/personvernombud/en/help/research_methods/online_surveys.html. Accessed April 26, 2020.

[20] Paired sample t-test. https://www.statisticssolutions.com/manova-analysis-paired-sample-t-test/. Accessed May 28, 2020.

[21] React. https://reactjs.org/. Accessed August 06, 2019.

[22] Redux. https://redux.js.org/. Accessed August 06, 2019.

[23] Surveyhero. https://www.surveyhero.com/. Accessed April 20, 2020.

[24] Surveymethods. https://app.surveymethods.com/. Accessed April 20, 2020.

[25] Surveymonkey. https://www.surveymonkey.com/. Accessed April 20, 2020.

[26] Usability 101: Introduction to usability. http://www.useit.com/alertbox/20030825.html. Accessed Frebruary 23, 2020.

[27] Web-based coras tool. https://coras-explorer.firebaseapp.com/try-it. Accessed November 6, 2019.

[28] Kazim Ali. A study of software development life cycle process models. *International Journal of Advanced Research in Computer Science*, 8(1), 2017.

[29] R. Balzer, F. Belz, R. Dewar, D. Fisher, R. Gabriel, J. Guttag, P. Hudak, and M. Wand. Prototyping. *Annual Review of Computer Science*, 4:453–465, 1990.

[30] Anis Bey, Patrick Jermann, and Pierre Dillenbourg. A comparison between two automatic assessment approaches for programming: An empirical study on moocs. *Journal of Educational Technology & Society*, 21(2):259–272, 2018.

[31] Wilco J. Bonestroo and Ton de Jong. Effects of planning on task load, knowledge, and tool preference: a comparison of two tools. *Interactive Learning Environments*, 20(2):141–153, 2012.

[32] R. Conradi and A. I. Wang. *Empirical methods and studies in software engineering: Experiences from ESERNET.* Springer, 2003.

[33] F.E. Grubbs. 'procedures for detecting outlying observations in samples'. *Technometrics*, 11(1), 1969.

[34] Maurizio Leotta, Diego Clerissi, Filippo Ricca, and Paolo Tonella. Visual vs. dom-based web locators: An empirical study. In Sven Casteleyn, Gustavo Rossi, and Marco Winckler, editors, *Web Engineering*, pages 322–340, Cham, 2014. Springer International Publishing.

[35] M.S. Lund and K. Stølen B. Solhaug. *Model-Driven Risk Analysis - The CORAS Approach.* Springer-Verlag Berlin Heidelberg, 2011.

[36] Birger Madsen. *Statistics for Non-Statisticians.* Springer, 2011.

[37] Robert Mcgill, John W. Tukey, and Wayne A. Larsen. Variations of box plots. *The American Statistician*, 32(1):12–16, 1978.

[38] J.E. McGrath. *Groups: Interaction and Performance.* Prentice-Hall, 1984.

[39] A. Perini, A. Susi, F. Ricca, and C. Bazzanella. An empirical study to compare the accuracy of ahp and cbranking techniques for requirements prioritization. In *2007 Fifth International Workshop on Comparative Evaluation in Requirements Engineering*, pages 23–35, 2007.

[40] J. Preece, H. Sharp, and Y. Rogers. *Interation Design beyond human-computer interaction.* John Wiley and Sons, 2011.

[41] A. Refsdal, B. Solhaug, and K. Stølen. *Cyber-Risk Management.* Springer, 2015.

[42] Safdar Aqeel Safdar, Muhammad Zohaib Iqbal, and Muhammad Uzair Khan. Empirical evaluation of uml modeling tools–a controlled experiment. In Gabriele Taentzer and Francis Bordeleau, editors, *Modelling Foundations and Applications*, pages 33–44, Cham, 2015. Springer International Publishing.

[43] T. Sandelin and M. Vierimaa. *'Empirical studies in esernet'. In: Empirical Methods and Studies in Software Engineering.* Springer, 2003.

[44] F. Shull, J. Singer, and D.I.K. Sjøberg. *Guide to Advanced Empirical Software Engineering.* Springer, 2007.

[45] 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'. In: IEEE transactions on soft- ware engineering.* IEEE Computer Society, 2005.

[46] I. Solheim and K. Stølen. *Technology research explained.* SINTEF Information and Communication Technology, 2007.

[47] R. van Solingen, V. Basili, G. Caldiera, and H. D. Rombach. *'Goal Question Metric (GQM) Approach'. In: Encyclopedia of Software Engineering.* John Wiley and Sons, 2002.

[48] R.J. Wieringa. *Design Science Methodology for Information Systems and Software Engineering.* Springer Berlin Heidelberg, 2014.

[49] C. Wohlin, M. Höst, and K. Henningsson. *'Empirical research methods in software engineering' In: Empirical Methods and Studies in Software Engineering.* Springer, 2003.

Appendices

Appendix A

Presentation - Group A

Experiment – Assessment of web-based CORAS tools

In this experiment, we need your help to evaluate web-based CORAS tools for security risk analysis.

➢ STAGE 1:

You will be given a survey to gather background information. This survey will be used for subsequent data analysis of the personal usage of different tools. Furthermore, the letter of consent must be signed and emailed back to fangronf@ifi.uio.no

➢ STAGE 2:

You will get two questionnaires with tasks to be solved.
There are 10 tasks to be solved in each questionnaire.

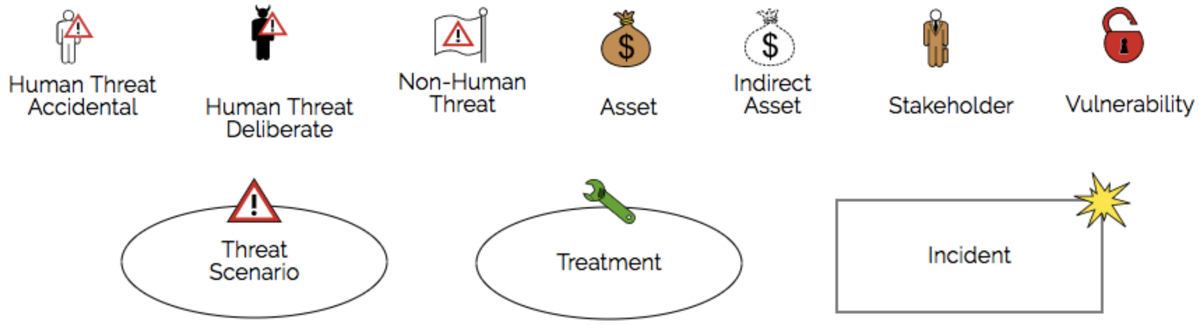
Please complete the surveys according to the following requirements:
(**NOTE**: When you are doing any one of the questionnaires, please do not take a break before submitting)

1. First please use the existing web-based CORAS tool (https://coras-explorer.firebaseapp.com/try-it) to complete the tasks for the existing CORAS tool (https://existingcorastool.evalandgo.com/s/?id=JTk3byU5QWslOUQlQjE=&a=JTk2byU5N28lOUElQjE=)
2. Can take a break before you start to the next survey.
3. Please use the re-engineered web-based CORAS tool (https://re-coras.web.app/try-it) to complete the tasks for the re-engineered CORAS tool (https://recorastool.evalandgo.com/s/?id=JTk3cCU5MW8lOTYlQUI=&a=JTk2byU5N28lOUElQjE=)

➢ STAGE 3:

At the end, please complete the following feedback survey questionnaire. This questionnaire is to get your feedback on the experience of using the above two tools.
https://app.evalandgo.com/s/?id=JTk3byU5MmwlOTclQUE=&a=JTk2byU5N28lOUElQjE=

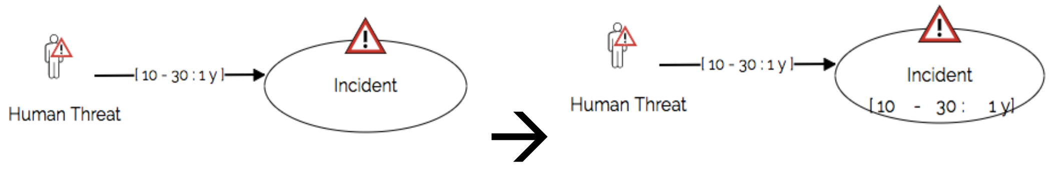
Symbols of the CORAS risk modelling language:



How the frequency likelihood is calculated:

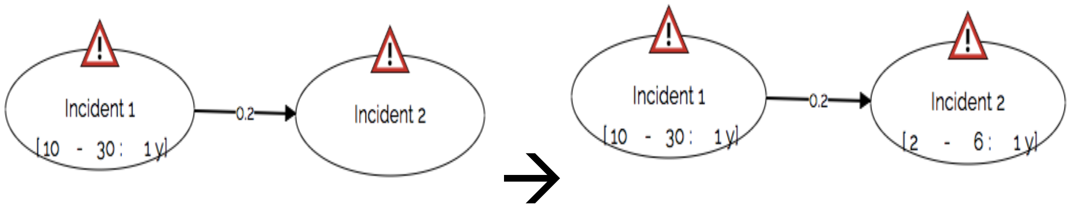
1. **Initiates relation:**
   The probability *p'* of the occurrences of *scenarios/incidents* due to threat is equal to the probability *p* with which initiates:

$$p' = p$$
$$= [10 - 30 : 1y]$$



2. **Leads-to relation:**
   The probability *p'* of the occurrences of *scenarios/incident 2* that are due to *scenarios*/incident 1 is equal to the probability *p* of *scenarios/incident 1* multiplied with the conditional likelihood *l*:

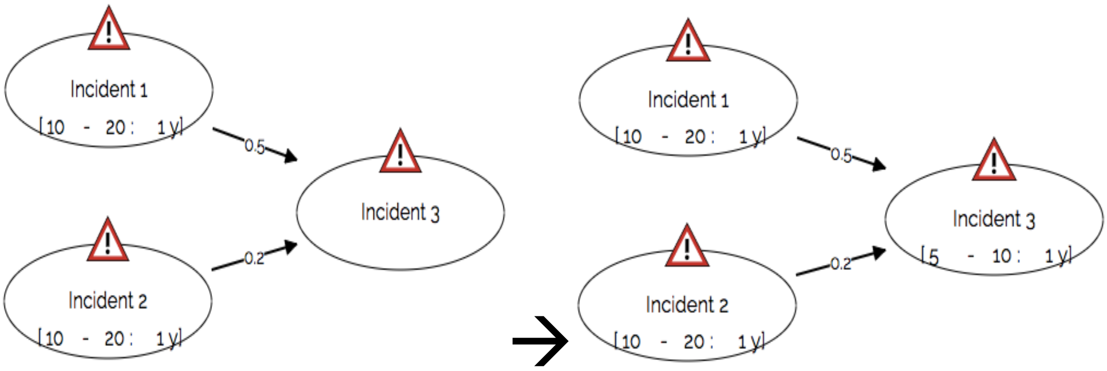
$$p' = p * l$$
$$= [10 - 30 : 1y]* 0.2$$
$$= [2 - 6 : 1y]$$



3. **Mutually exclusive relation:**
   (Two incidents are mutually exclusive means that either incident 1 or incident 2 occurs, but not both.)

   The probability *p3* of the occurrences of *scenarios/incident 3* that are due to *scenarios/incident 1* or *scenarios/incident 2* is equal the maximum probability value from *scenarios/incident 1* and *scenarios/incident 2* 's probabilities：

$$p3 = MAX(p1 * l_1 , p2 * l_2)$$
$$= MAX([10 - 20 : 1y]* 0.5, [10 - 30 : 1y]* 0.2)$$
$$=[5 – 10 : 1y]$$

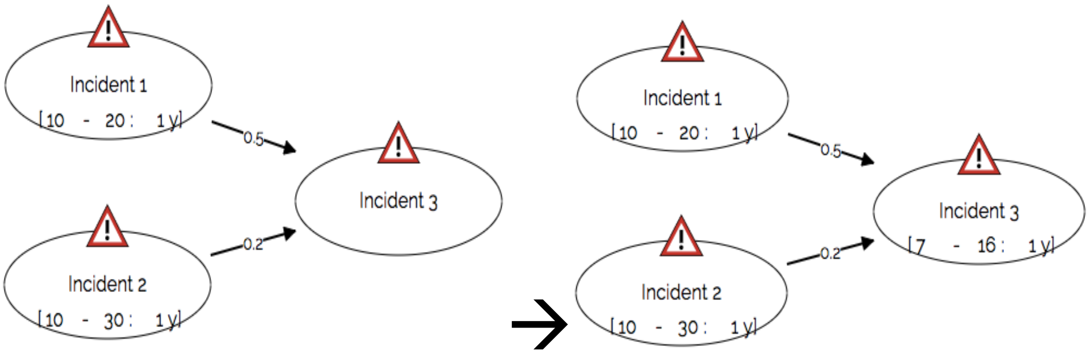


4. Statistically independent relation:

(Two incidents are statistically independent means two events can occur simultaneously and do not affect each other.)

The probability *p3* of the occurrences of *scenarios/incident 3* that are due to *scenarios/incident 1* or *scenarios/incident 2* is equal to the sum of *scenarios/incident 1* and *scenarios/incident 2* 's respective probabilities:

$$p3 = p1 * l_1 + p2 * l_2$$
$$= [10 - 20 : 1y]* 0.5 + [10 - 30 : 1y]* 0.2$$
$$= [7 - 16 : 1y]$$

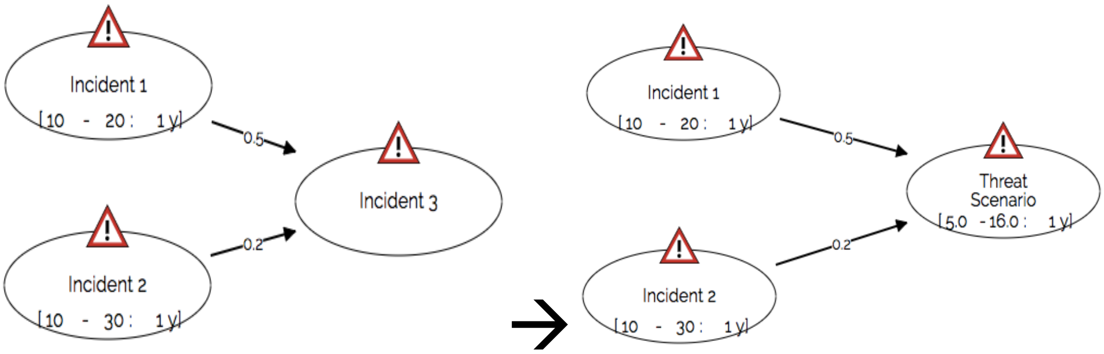


5. Neither mutually exclusive nor statistically independent relation:

The probability *p3*'s minimum is equal the maximum probability value from *scenarios/incident 1* and *scenarios/incident 2* 's minimum probabilities.

The probability *p3*'s maximum is equal to the sum of *scenarios/incident 1* and *scenarios/incident 2* 's respective maximum probabilities:

$$p3 = [p3\_min , p3\_max]$$
$$= [MAX(p1\_min * l1 , p2\_min * l2) , (p1\_max * l1 + p2\_max * l2)]$$
$$= [MAX(10 * 0.5 , 10 * 0.2) - (20 * 0.5 + 30 * 0.2) :1y]$$
$$= [5 – 16 : 1y]$$



How to check consistency:

Assigned interval: *e([min , max])*
Calculated interval: *e([min' , max'])*
Consistency check: *[min' , max']* ⊆ *[min , max]* or, equivalently *min* ≤ *min'* and *max* ≥ *max'*

Appendix B

Presentation - Group B

Experiment – Assessment of web-based CORAS tools

In this experiment, we need your help to evaluate web-based CORAS tools for security risk analysis.

➢ STAGE 1:

You will be given a survey to gather background information. This survey will be used for subsequent data analysis of the personal usage of different tools. Furthermore, the letter of consent must be signed and emailed back to fangronf@ifi.uio.no

➢ STAGE 2:

You will get two questionnaires with tasks to be solved.
There are 10 tasks to be solved in each questionnaire.

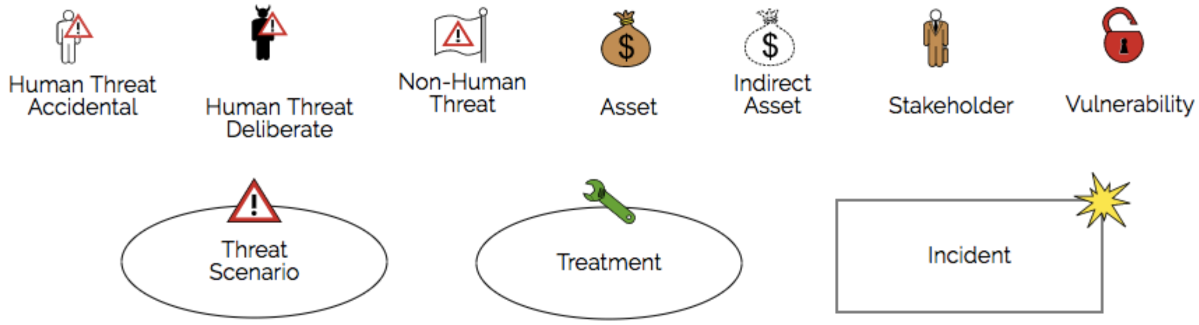
Please complete the surveys according to the following requirements:
(**NOTE**: When you are doing any one of the questionnaires, please do not take a break before submitting)

1. Please use the re-engineered web-based CORAS tool (https://re-coras.web.app/try-it) to complete the tasks for the re-engineered CORAS tool (https://recorastool.evalandgo.com/s/?id=JTk3cCU5MW8lOTYlQUI=&a=JTk2byU5N28lOUElQjE=)
2. Can take a break before you start to the next survey.
3. First please use the existing web-based CORAS tool (https://coras-explorer.firebaseapp.com/try-it) to complete the tasks for the existing CORAS tool (https://existingcorastool.evalandgo.com/s/?id=JTk3byU5QWslOUQlQjE=&a=JTk2byU5N28lOUElQjE=)

➢ STAGE 3:

At the end, please complete the following feedback survey questionnaire. This questionnaire is to get your feedback on the experience of using the above two tools.
https://app.evalandgo.com/s/?id=JTk3byU5MmwlOTclQUE=&a=JTk2byU5N28lOUElQjE=

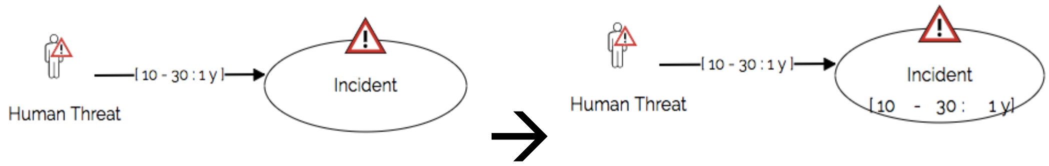
Symbols of the CORAS risk modelling language:



How the frequency likelihood is calculated:

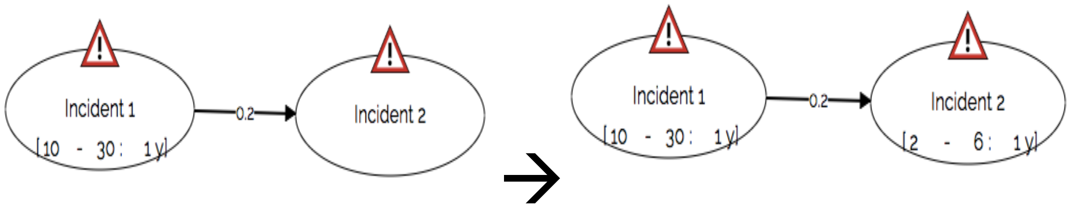
1. **Initiates relation:**
   The probability *p'* of the occurrences of *scenarios/incidents* due to threat is equal to the probability *p* with which initiates:

$$p' = p$$
$$= [10 - 30 : 1y]$$



2. **Leads-to relation:**
   The probability *p'* of the occurrences of *scenarios/incident 2* that are due to *scenarios*/incident 1 is equal to the probability *p* of *scenarios/incident 1* multiplied with the conditional likelihood *l*:

$$p' = p * l$$
$$= [10 - 30 : 1y] * 0.2$$
$$= [2 - 6 : 1y]$$



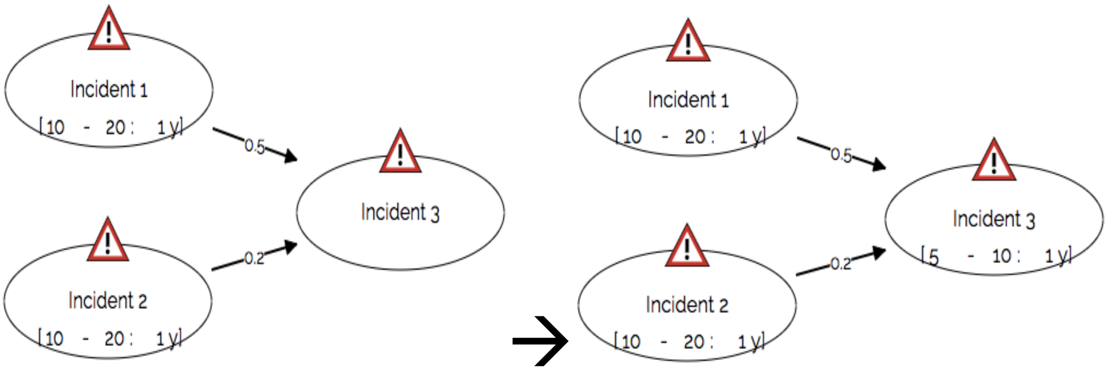
3. **Mutually exclusive relation:**
   (Two incidents are mutually exclusive means that either incident 1 or incident 2 occurs, but not both.)

   The probability *p3* of the occurrences of *scenarios/incident 3* that are due to *scenarios/incident 1* or *scenarios/incident 2* is equal the maximum probability value from *scenarios/incident 1* and *scenarios/incident 2* 's probabilities：

$$p3 = MAX(p1 * l1 , p2 * l2)$$

$$= MAX([10 - 20 : 1y]* 0.5, [10 - 30 : 1y]* 0.2)$$

$$=[5 – 10 : 1y]$$



4. Statistically independent relation:

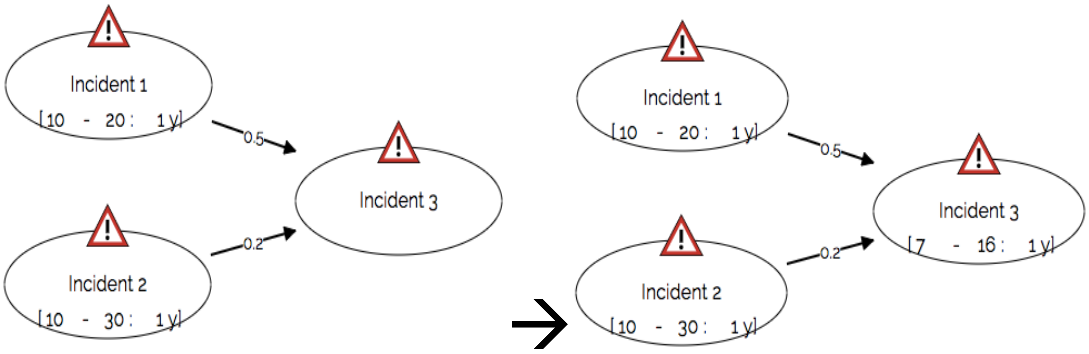
(Two incidents are statistically independent means two events can occur simultaneously and do not affect each other.)

The probability *p3* of the occurrences of *scenarios/incident 3* that are due to *scenarios/incident 1* or *scenarios/incident 2* is equal to the sum of *scenarios/incident 1* and *scenarios/incident 2* 's respective probabilities:

$$p3 = p1 * l1 + p2 * l2$$

$$= [10 - 20 : 1y]* 0.5 + [10 - 30 : 1y]* 0.2$$

$$= [7 - 16 : 1y]$$



5. Neither mutually exclusive nor statistically independent relation:

The probability *p3*'s minimum is equal the maximum probability value from *scenarios/incident 1* and *scenarios/incident 2* 's minimum probabilities.

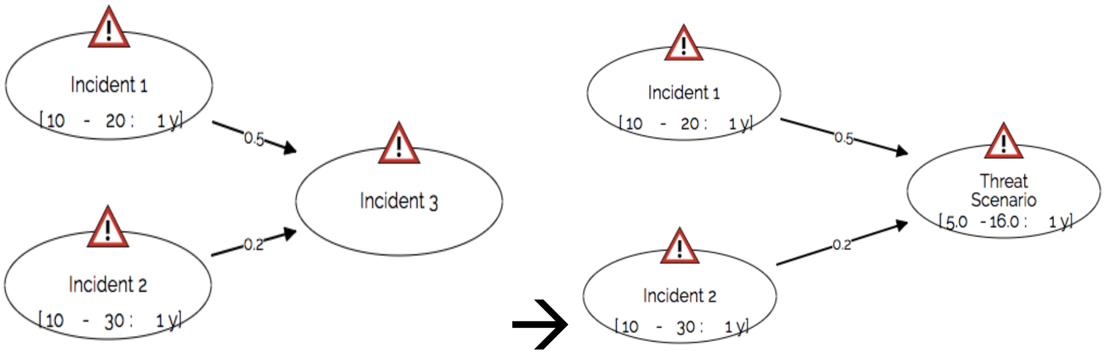
The probability *p3*'s maximum is equal to the sum of *scenarios/incident 1* and *scenarios/incident 2* 's respective maximum probabilities:

$$p3 = [p3\_min , p3\_max]$$

$$= [MAX(p1\_min * l1 , p2\_min * l2) , (p1\_max * l1 + p2\_max * l2)]$$

$$= [MAX(10 * 0.5 , 10 * 0.2) - (20 * 0.5 + 30 * 0.2) :1y]$$

$$= [5 – 16 : 1y]$$



How to check consistency:

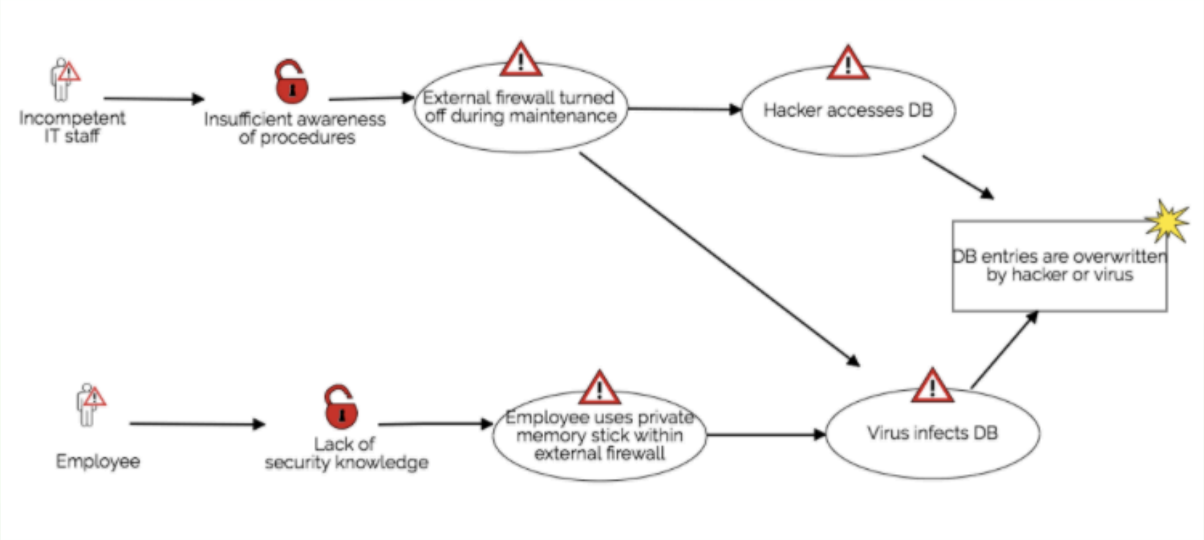
Assigned interval: $e([min\ ,\ max])$
Calculated interval: $e([min'\ ,\ max'])$
Consistency check: $[min'\ ,\ max'] \subseteq [min\ ,\ max]$ or, equivalently $min \leq min'$ and $max \geq max'$

Appendix C

Task Questionnaire - Tasks for the existing CORAS tool

Task 1:

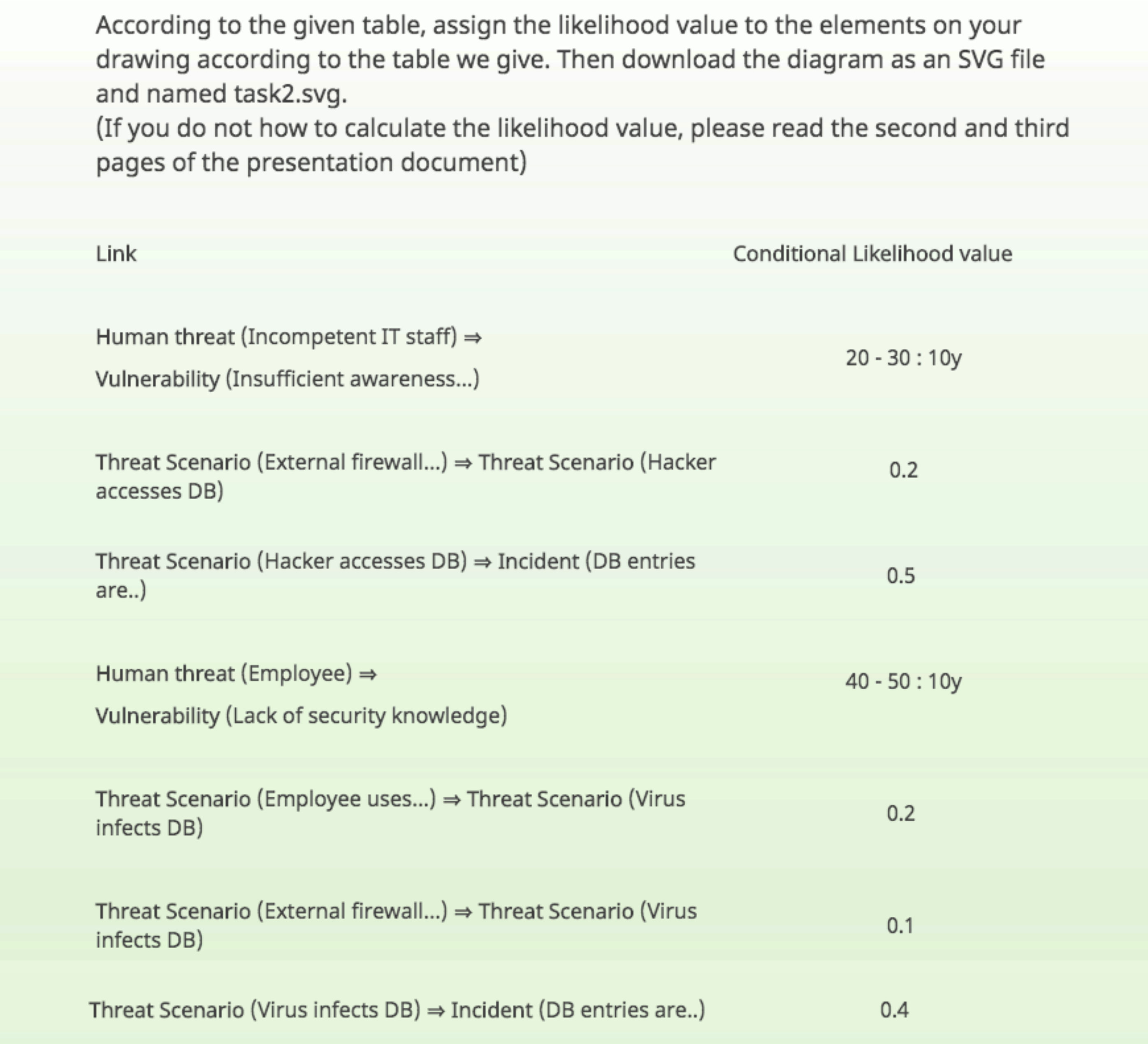
According to the given figure, draw an exactly the same diagram on the CORAS tool. Then download the diagram as an SVG file and named task1.svg.


Incompetent IT staff
Insufficient awareness of procedures
External firewall turned off during maintenance
Hacker accesses DB
DB entries are overwritten by hacker or virus
Employee
Lack of security knowledge
Employee uses private memory stick within external firewall
Virus infects DB

Name of elements that you can copy:

Incompetent IT staff
Insufficient awareness of procedures
External firewall turned off during maintenance
Hacker accesses DB
Employee
Lack of security knowledge
Employee uses private memory stick within external firewall
Virus infects DB
DB entries are overwritten by hacker or virus

Task 2:

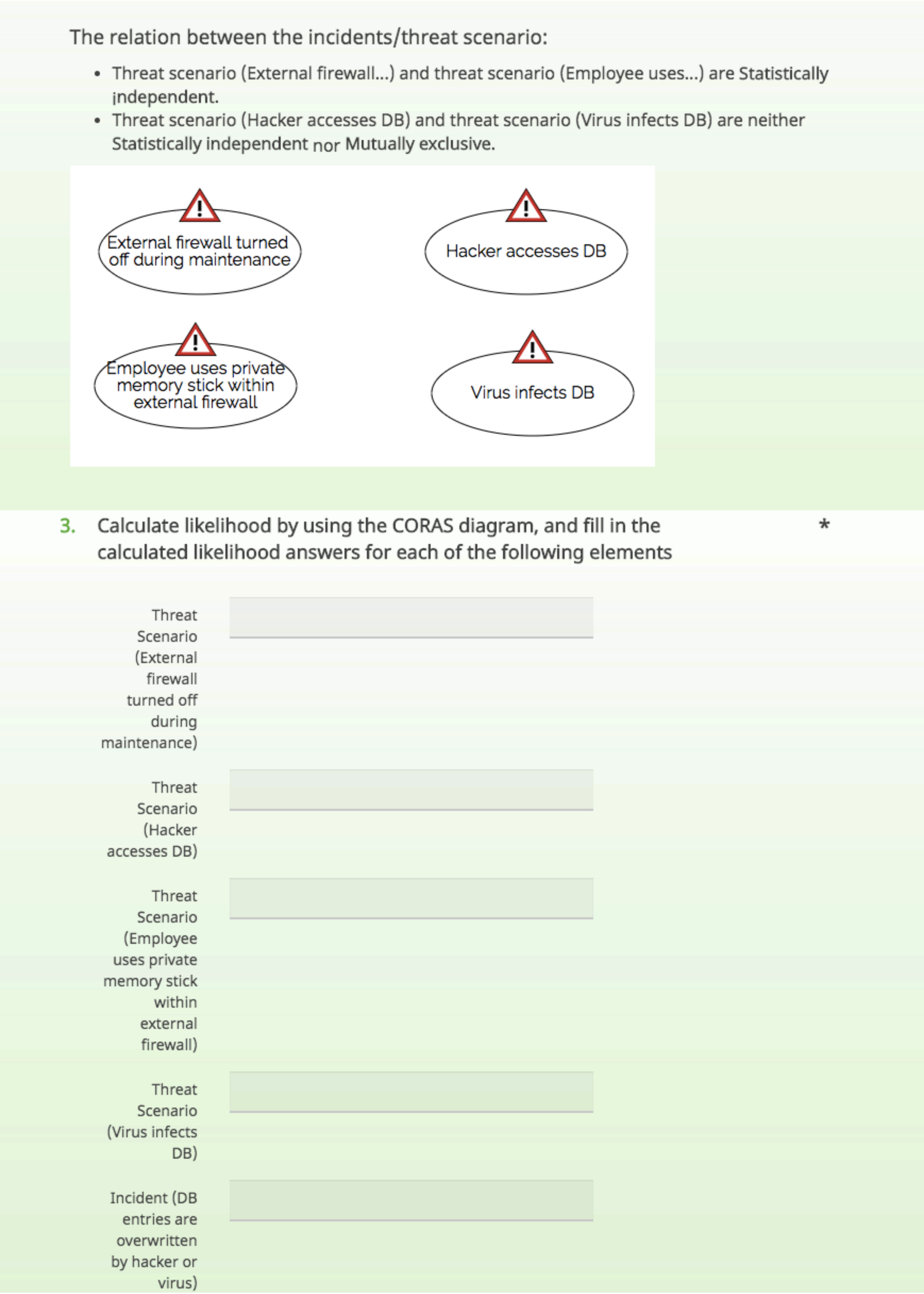


According to the given table, assign the likelihood value to the elements on your drawing according to the table we give. Then download the diagram as an SVG file and named task2.svg.

(If you do not how to calculate the likelihood value, please read the second and third pages of the presentation document)

Link	Conditional Likelihood value
Human threat (Incompetent IT staff) ⇒ Vulnerability (Insufficient awareness...)	20 - 30 : 10y
Threat Scenario (External firewall...) ⇒ Threat Scenario (Hacker accesses DB)	0.2
Threat Scenario (Hacker accesses DB) ⇒ Incident (DB entries are..)	0.5
Human threat (Employee) ⇒ Vulnerability (Lack of security knowledge)	40 - 50 : 10y
Threat Scenario (Employee uses...) ⇒ Threat Scenario (Virus infects DB)	0.2
Threat Scenario (External firewall...) ⇒ Threat Scenario (Virus infects DB)	0.1
Threat Scenario (Virus infects DB) ⇒ Incident (DB entries are..)	0.4

Task 3:



The relation between the incidents/threat scenario:

- Threat scenario (External firewall...) and threat scenario (Employee uses...) are Statistically independent.
- Threat scenario (Hacker accesses DB) and threat scenario (Virus infects DB) are neither Statistically independent nor Mutually exclusive.

External firewall turned off during maintenance
Hacker accesses DB
Employee uses private memory stick within external firewall
Virus infects DB

3. Calculate likelihood by using the CORAS diagram, and fill in the calculated likelihood answers for each of the following elements *

Threat Scenario (External firewall turned off during maintenance)

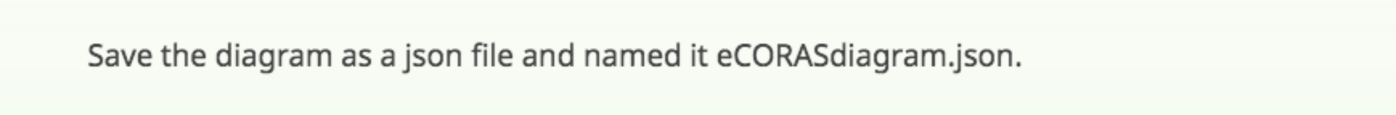
Threat Scenario (Hacker accesses DB)

Threat Scenario (Employee uses private memory stick within external firewall)

Threat Scenario (Virus infects DB)

Incident (DB entries are overwritten by hacker or virus)

Task 4:

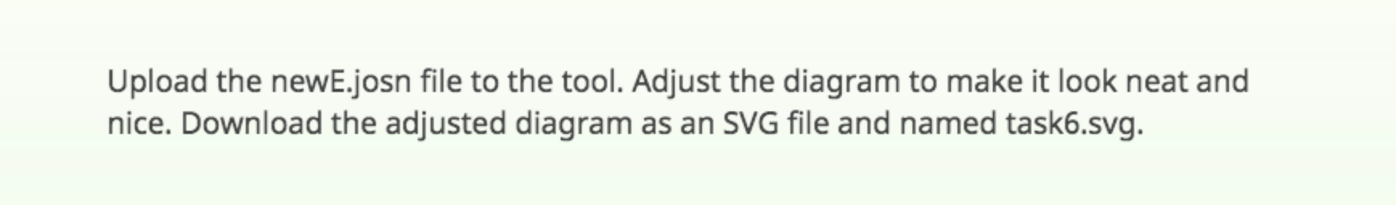


Save the diagram as a json file and named it eCORASdiagram.json.

Task 5:

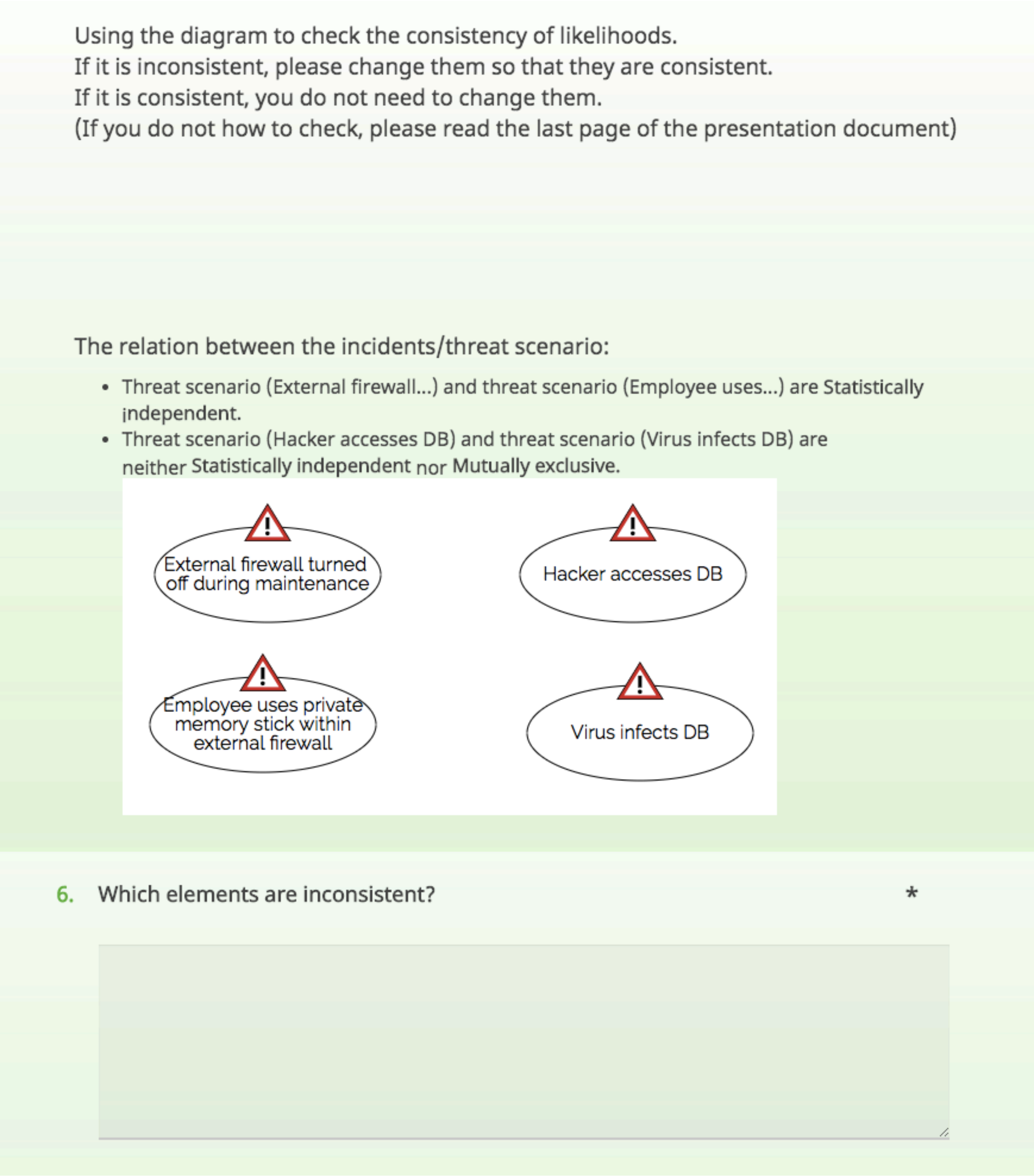
Clear up the drawing panel

Task 6:



Upload the newE.josn file to the tool. Adjust the diagram to make it look neat and nice. Download the adjusted diagram as an SVG file and named task6.svg.

Task 7:



Using the diagram to check the consistency of likelihoods.
If it is inconsistent, please change them so that they are consistent.
If it is consistent, you do not need to change them.
(If you do not how to check, please read the last page of the presentation document)

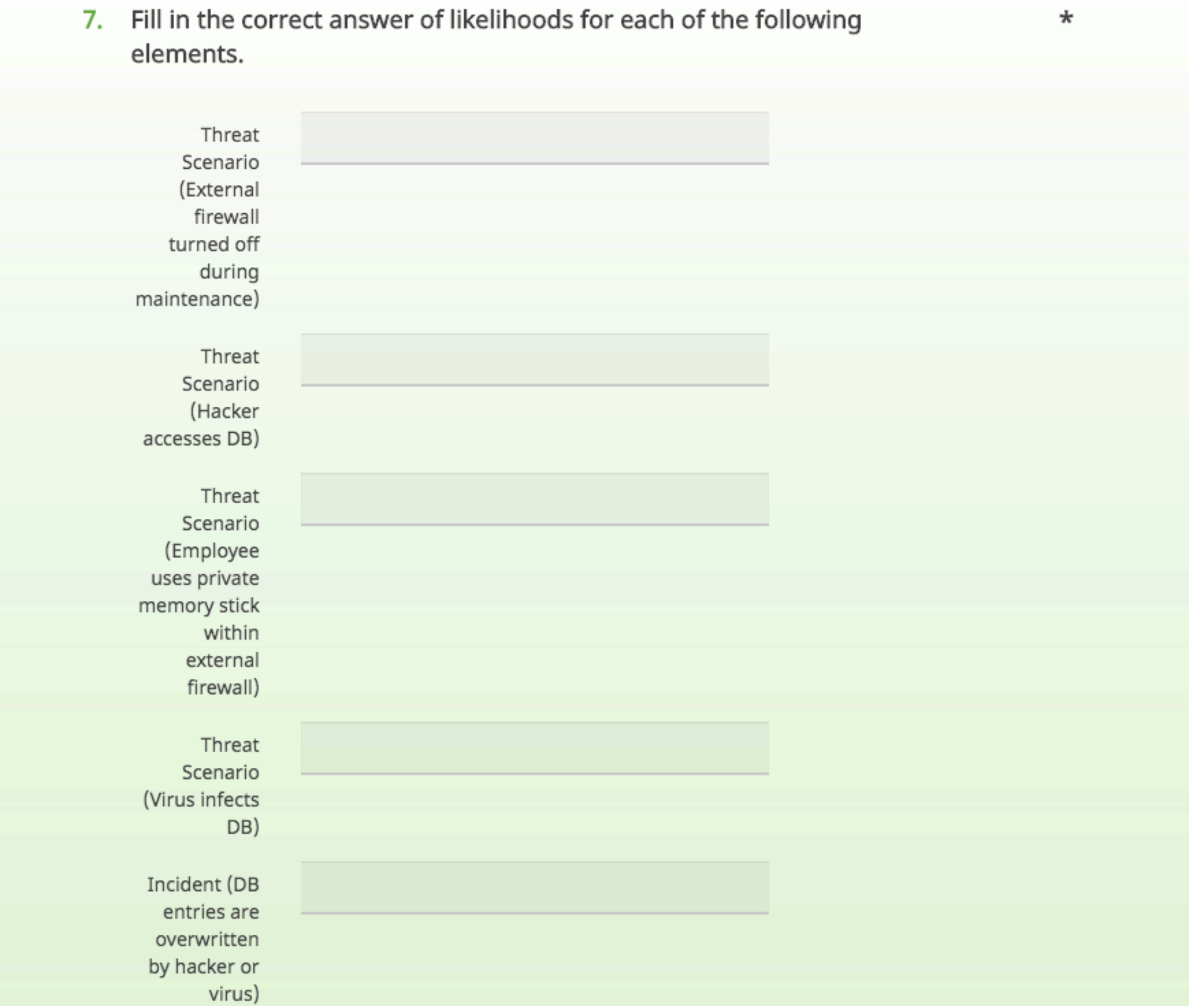
The relation between the incidents/threat scenario:

- Threat scenario (External firewall...) and threat scenario (Employee uses...) are Statistically independent.
- Threat scenario (Hacker accesses DB) and threat scenario (Virus infects DB) are neither Statistically independent nor Mutually exclusive.

External firewall turned off during maintenance
Hacker accesses DB
Employee uses private memory stick within external firewall
Virus infects DB

6. Which elements are inconsistent? *

Task 8:



7. Fill in the correct answer of likelihoods for each of the following elements. *

Threat Scenario (External firewall turned off during maintenance)

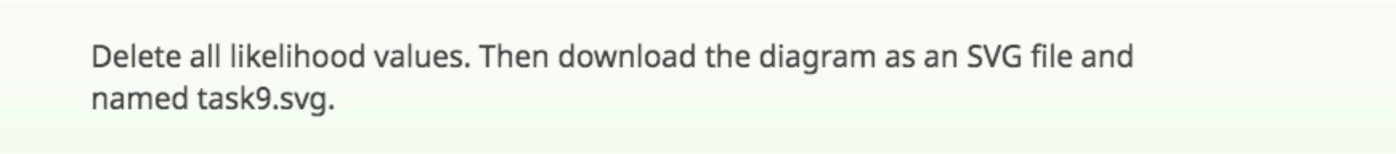
Threat Scenario (Hacker accesses DB)

Threat Scenario (Employee uses private memory stick within external firewall)

Threat Scenario (Virus infects DB)

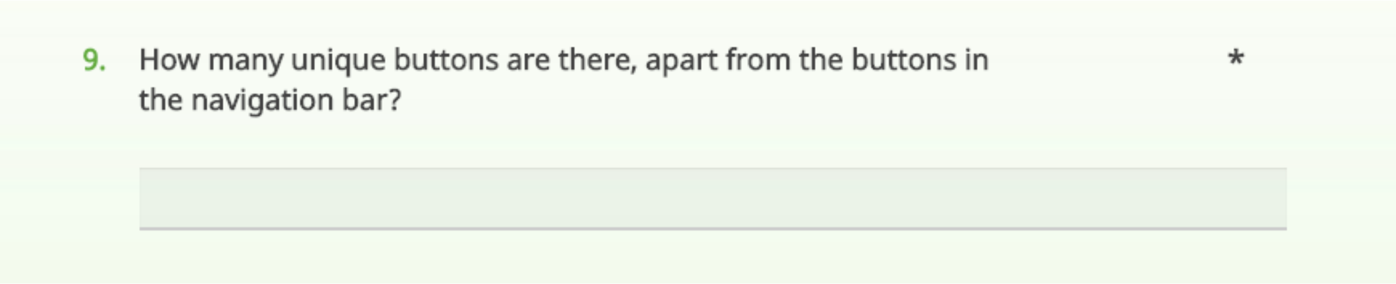
Incident (DB entries are overwritten by hacker or virus)

Task 9:



Delete all likelihood values. Then download the diagram as an SVG file and named task9.svg.

Task 10:



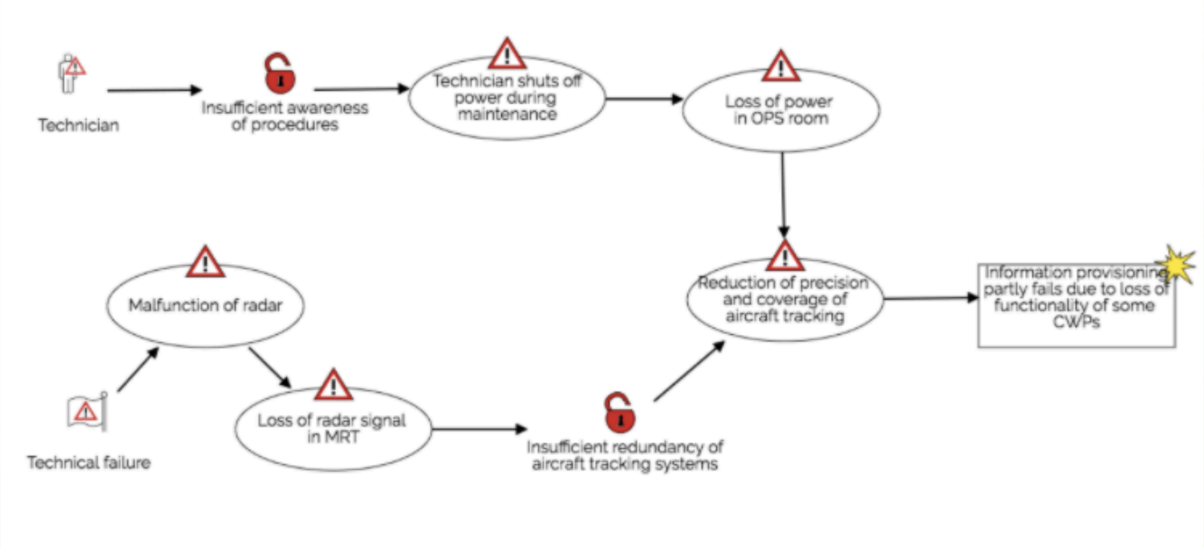
9. How many unique buttons are there, apart from the buttons in the navigation bar? *

Appendix D

Task Questionnaire - Tasks for the re-engineered CORAS tool

Task 1:

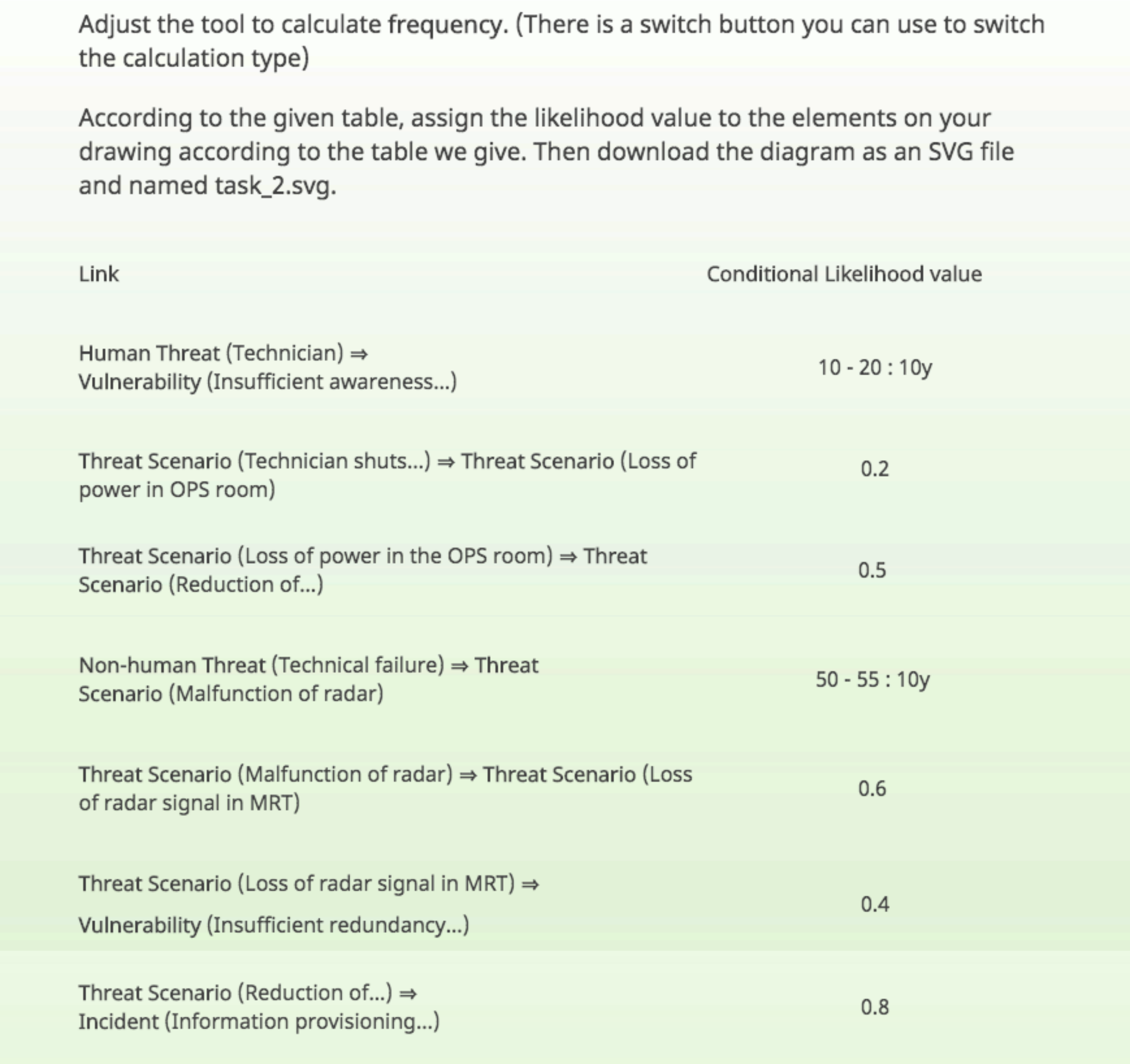
According to the given figure, draw an exactly the same diagram on the CORAS tool. Then download the diagram as an SVG file and named task_1.svg.


Technician
Insufficient awareness of procedures
Technician shuts off power during maintenance
Loss of power in OPS room
Reduction of precision and coverage of aircraft tracking
Information provisioning partly fails due to loss of functionality of some CWPs
Malfunction of radar
Technical failure
Loss of radar signal in MRT
Insufficient redundancy of aircraft tracking systems

Name of elements that you can copy:

Technician
Insufficient awareness of procedures
Technician shuts off power during maintenance
Loss of power in OPS room
Technical failure
Malfunction of radar
Loss of radar signal in MRT
Insufficient redundancy of aircraft tracking systems
Reduction of precision and coverage of aircraft tracking
Information provisioning partly fails due to loss of functionality of some CWPs

Task 2:

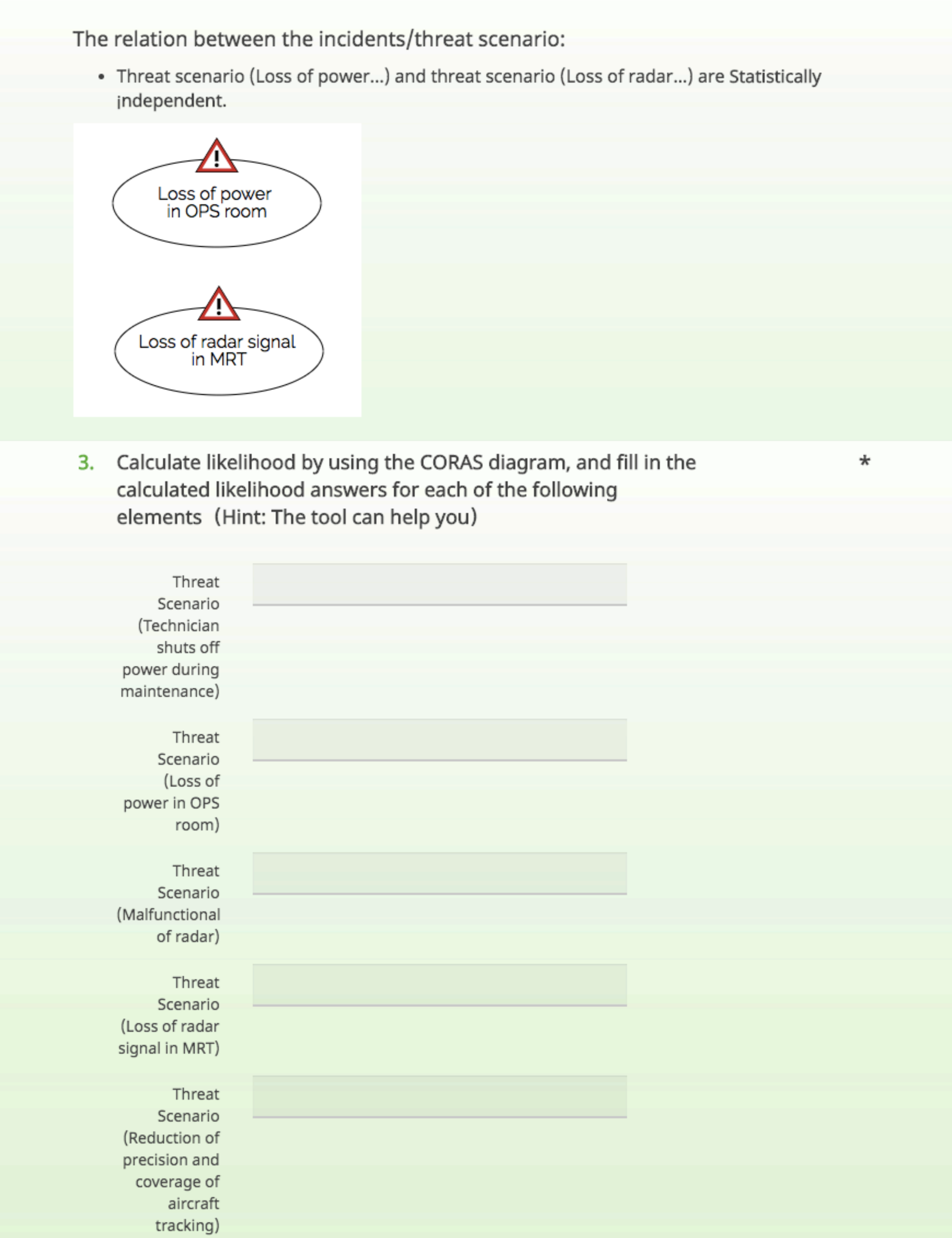


Adjust the tool to calculate frequency. (There is a switch button you can use to switch the calculation type)

According to the given table, assign the likelihood value to the elements on your drawing according to the table we give. Then download the diagram as an SVG file and named task_2.svg.

Link	Conditional Likelihood value
Human Threat (Technician) ⇒ Vulnerability (Insufficient awareness...)	10 - 20 : 10y
Threat Scenario (Technician shuts...) ⇒ Threat Scenario (Loss of power in OPS room)	0.2
Threat Scenario (Loss of power in the OPS room) ⇒ Threat Scenario (Reduction of...)	0.5
Non-human Threat (Technical failure) ⇒ Threat Scenario (Malfunction of radar)	50 - 55 : 10y
Threat Scenario (Malfunction of radar) ⇒ Threat Scenario (Loss of radar signal in MRT)	0.6
Threat Scenario (Loss of radar signal in MRT) ⇒ Vulnerability (Insufficient redundancy...)	0.4
Threat Scenario (Reduction of...) ⇒ Incident (Information provisioning...)	0.8

Task 3:



The relation between the incidents/threat scenario:

- Threat scenario (Loss of power...) and threat scenario (Loss of radar...) are Statistically independent.

3. Calculate likelihood by using the CORAS diagram, and fill in the calculated likelihood answers for each of the following elements (Hint: The tool can help you) *

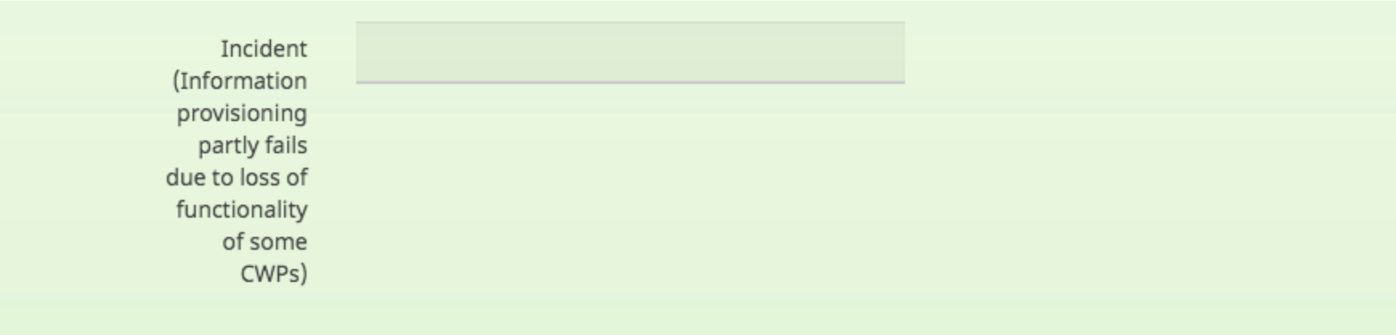
Threat Scenario (Technician shuts off power during maintenance)

Threat Scenario (Loss of power in OPS room)

Threat Scenario (Malfunctional of radar)

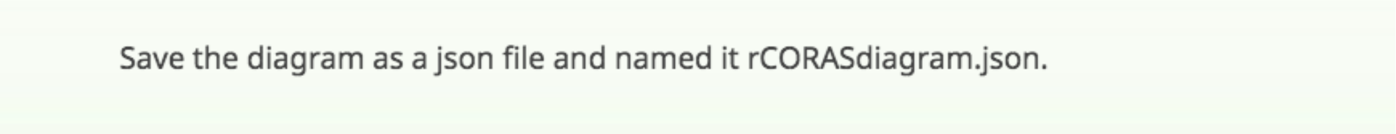
Threat Scenario (Loss of radar signal in MRT)

Threat Scenario (Reduction of precision and coverage of aircraft tracking)



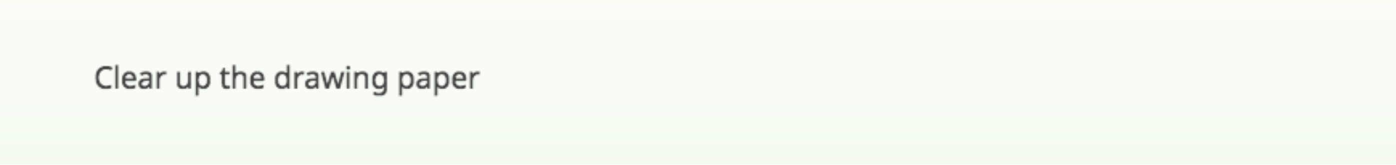
Incident (Information provisioning partly fails due to loss of functionality of some CWPs)

Task 4:



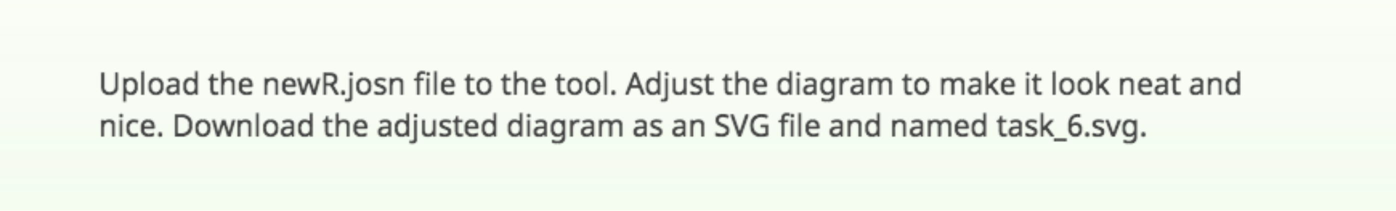
Save the diagram as a json file and named it rCORASdiagram.json.

Task 5:



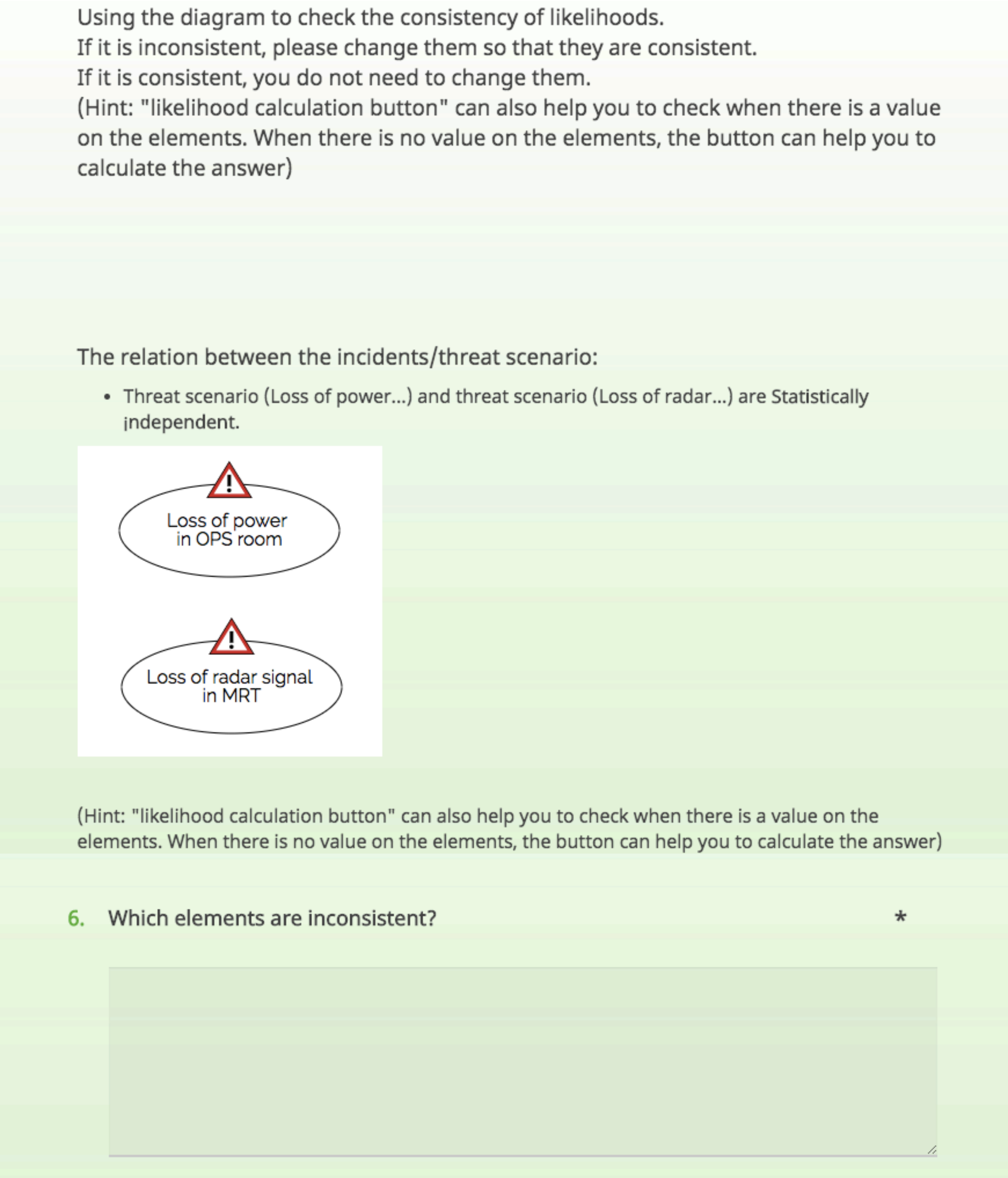
Clear up the drawing paper

Task 6:



Upload the newR.josn file to the tool. Adjust the diagram to make it look neat and nice. Download the adjusted diagram as an SVG file and named task_6.svg.

Task 7:



Using the diagram to check the consistency of likelihoods.
If it is inconsistent, please change them so that they are consistent.
If it is consistent, you do not need to change them.
(Hint: "likelihood calculation button" can also help you to check when there is a value on the elements. When there is no value on the elements, the button can help you to calculate the answer)

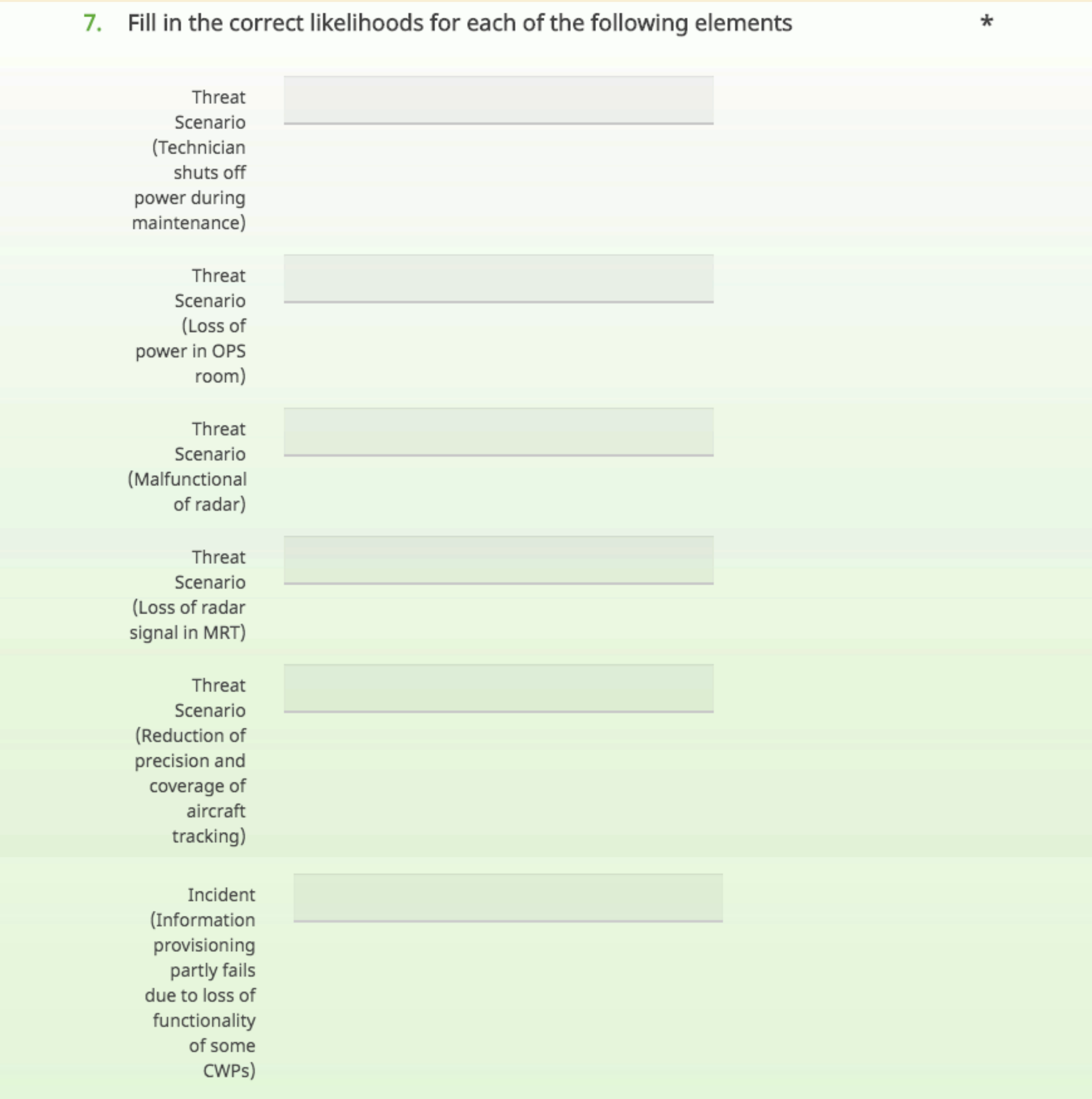
The relation between the incidents/threat scenario:

- Threat scenario (Loss of power...) and threat scenario (Loss of radar...) are Statistically independent.

(Hint: "likelihood calculation button" can also help you to check when there is a value on the elements. When there is no value on the elements, the button can help you to calculate the answer)

6. Which elements are inconsistent? *

Task 8:



7. Fill in the correct likelihoods for each of the following elements *

Threat Scenario (Technician shuts off power during maintenance)

Threat Scenario (Loss of power in OPS room)

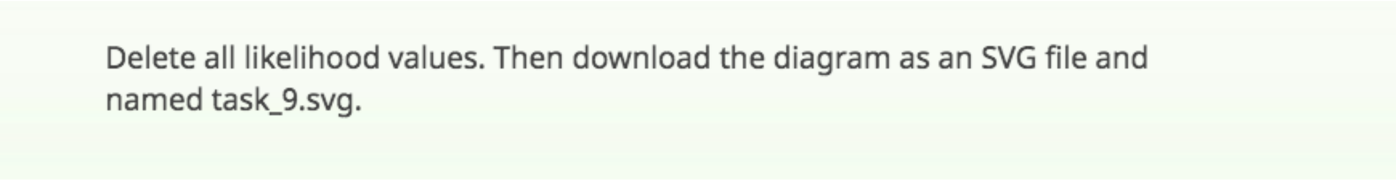
Threat Scenario (Malfunctional of radar)

Threat Scenario (Loss of radar signal in MRT)

Threat Scenario (Reduction of precision and coverage of aircraft tracking)

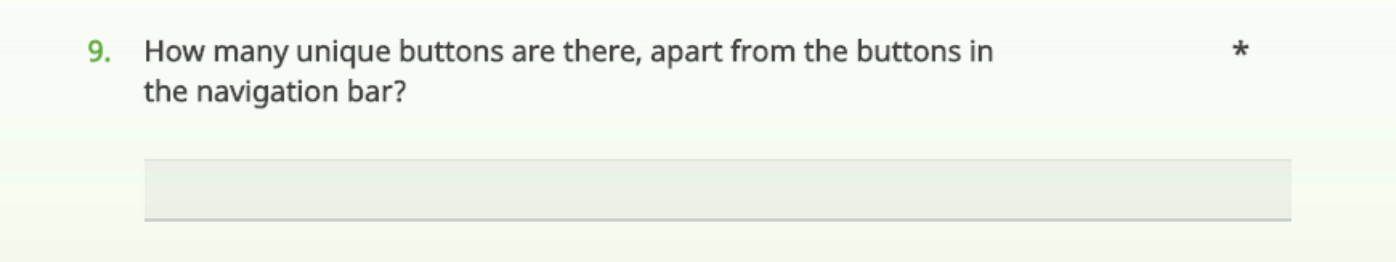
Incident (Information provisioning partly fails due to loss of functionality of some CWPs)

Task 9:



Delete all likelihood values. Then download the diagram as an SVG file and named task_9.svg.

Task 10:



9. How many unique buttons are there, apart from the buttons in the navigation bar? *

Appendix E

Task Scores from the Experiment

Group A					
Task for the existing version	**Person a_1**	**Person a_2**	**Person a_3**	**Person a_4**	**Avg.**
Task 1	20	20	20	19	19.75
Task 2	18	10	17	20	16.25
Task 3	2	5	0	4	2.75
Task 4	1	1	0	1	0.75
Task 5	1	1	1	1	1
Task 6	2	2	2	2	2
Task 7	3	5	3	0	2.75
Task 8	3	5	4	4	4
Task 9	1	1	1	1	1
Task 10	2	2	4	5	3.25
Total	53	52	52	57	53.5

Group B					
Task for the existing version	**Person b_1**	**Person b_2**	**Person b_3**	**Person b_4**	**Avg.**
Task 1	20	20	20	20	20
Task 2	20	20	20	20	20
Task 3	4	3	5	3	3.75
Task 4	0	1	0	0	0.25
Task 5	1	1	1	1	1
Task 6	2	2	2	2	2
Task 7	3	4	3	3	3.25
Task 8	3	5	5	3	4
Task 9	1	1	1	1	1
Task 10	4	2	2	4	3
Total	58	59	59	57	58.25

Group A					
Task for the re-engineered version	**Person a_1**	**Person a_2**	**Person a_3**	**Person a_4**	**Avg.**
Task_1	20	20	20	20	20
Task_2	20	20	20	20	20
Taks_3	5	3	3	5	4
Task_4	1	1	1	1	1
Task_5	1	1	1	1	1
Task_6	2	2	2	2	2
Taks_7	4	4	5	4	4.25
Task_8	5	5	4	5	4.75
Task_9	1	1	1	1	1
Task_10	5	2	3	5	3.75
Total	64	59	60	64	61.75

Group B					
Task for the re-engineered version	**Person b_1**	**Person b_2**	**Person b_3**	**Person b_4**	**Avg.**
Task_1	20	20	20	20	20
Task_2	20	20	20	20	20
Taks_3	3	5	5	3	4
Task_4	1	1	1	0	0.75
Task_5	1	1	1	1	1
Task_6	2	2	2	2	2
Taks_7	4	5	5	4	4.5
Task_8	5	5	5	3	4.5
Task_9	1	1	1	1	1
Task_10	4	2	3	5	3.5
Total	61	62	63	59	61.25

Appendix F

Feedback survey

Feedback Survey

I can quickly understand how to use the existing web-based CORAS tool, when I see its interface.

#	Question	No.	Min.	Average	Max.
1	I can quickly understand how to use the existing web-based CORAS tool, when I see its interface.	8	1	1.88	3

#	Question	Detail No.(%)
1	I can quickly understand how to use the existing web-based CORAS tool, when I see its interface.	8 (100%)
	0	0 (0%)
	1	2 (25%)
	2	5 (62.5%)
	3	1 (12.5%)
	4	0 (0%)

Feedback Survey

I can quickly understand how to use the re-engineered web-based CORAS tool, when I see its interface.

#	Question	No.	Min.	Average	Max.
2	I can quickly understand how to use the re-engineered web-based CORAS tool, when I see its interface.	8	3	3.5	4

#	Question	Detail No.(%)
2	I can quickly understand how to use the re-engineered web-based CORAS tool, when I see its interface.	8 (100%)
	0	0 (0%)
	1	0 (0%)
	2	0 (0%)
	3	4 (50%)
	4	4 (50%)

Feedback Survey

Compared with the existing web-basedCORAS tool, I prefer the interface of the re-engineered web-based CORAS tool.

#	Question	No.	Min.	Average	Max.
3	Compared with the existing web-basedCORAS tool, I prefer the interface of the re-engineered web-based CORAS tool.	8	3	3.88	4

#	Question	Detail No.(%)
3	Compared with the existing web-basedCORAS tool, I prefer the interface of the re-engineered web-based CORAS tool.	8 (100%)
	0	0 (0%)
	1	0 (0%)
	2	0 (0%)
	3	1 (12.5%)
	4	7 (87.5%)

Feedback Survey

I think the tip feature is useful, when I first used this tool.

#	Question	No.	Min.	Average	Max.
4	I think the tip feature is useful, when I first used this tool.	8	1	3.13	4

#	Question	Detail No.(%)
4	I think the tip feature is useful, when I first used this tool.	8 (100%)
	0	0 (0%)
	1	1 (12.5%)
	2	1 (12.5%)
	3	2 (25%)
	4	4 (50%)

Feedback Survey

I think the yellow "Likelihood Reset"button is useful, when I should modify one of the elements' likelihood value.

#	Question	No.	Min.	Average	Max.
5	I think the yellow "Likelihood Reset"button is useful, when I should modify one of the elements' likelihood value.	8	1	3.13	4

#	Question	Detail No.(%)
5	I think the yellow "Likelihood Reset"button is useful, when I should modify one of the elements' likelihood value.	8 (100%)
	0	0 (0%)
	1	1 (12.5%)
	2	2 (25%)
	3	0 (0%)
	4	5 (62.5%)

Feedback Survey

I think the dark cyan "Likelihood Reset" button is useful, when I should modify all elements' likelihood value.

#	Question	No.	Min.	Average	Max.
6	I think the dark cyan "Likelihood Reset" button is useful, when I should modify all elements' likelihood value.	8	4	4	4

#	Question	Detail No.(%)
6	I think the dark cyan "Likelihood Reset" button is useful, when I should modify all elements' likelihood value.	8 (100%)
	0	0 (0%)
	1	0 (0%)
	2	0 (0%)
	3	0 (0%)
	4	8 (100%)

Feedback Survey

The Likelihood Calculation feature is helpful for me.

#	Question	No.	Min.	Average	Max.
7	The Likelihood Calculation feature is helpful for me.	8	3	3.88	4

#	Question	Detail No.(%)
7	The Likelihood Calculation feature is helpful for me.	8 (100%)
	0	0 (0%)
	1	0 (0%)
	2	0 (0%)
	3	1 (12.5%)
	4	7 (87.5%)

Feedback Survey

Compared with the existing web-based CORAS tool, my satisfaction with the re-engineered web-based CORAS tool is

#	Question	No.	Min.	Average	Max.
8	Compared with the existing web-based CORAS tool, my satisfaction with the re-engineered web-based CORAS tool is	8	3	3.63	4

#	Question	Detail No.(%)
8	Compared with the existing web-based CORAS tool, my satisfaction with the re-engineered web-based CORAS tool is	8 (100%)
	0	0 (0%)
	1	0 (0%)
	2	0 (0%)
	3	3 (37.5%)
	4	5 (62.5%)

Feedback Survey

Comparing the zoom feature of the two tools, which one do you prefer, and why?

#	Question	Text
9	Comparing the zoom feature of the two tools, which one do you prefer, and why?	- If you can combine both, it will be better. Personally, I prefer the previous one. - I liked the old one better, since it was easier to make big adjustments. However, I would like it better if both options were included. - I like to old one. Scroll wheel is convenient. But you should probably add the + and - regardless, as some people may not have access to a scroll wheel (Like in a laptop scenario) - the second one, much easier - I prefer the re-engineered web-based CORAS tool. Easier and clearer. - the new one, because it has the buttons, and is more intuitive. It took me a while to realise that it was possible to zoom in the existing version. - the re-engineered web-based CORAS tool much easier to zoom in and zoom out. - First tool, since I can zoom in and out with the mouse scroll.

Feedback Survey

Do you think it makes sense of the toolbar position modification and the toolbar can hide in the re-engineered web-based CORAS tool? why?

#	Question	Text
10	Do you think it makes sense of the toolbar position modification and the toolbar can hide in the re-engineered web-based CORAS tool? why?	- Yes. When we deal with a complicated model, it will give us more space for the vertical direction. The re-engineered tool also has a better view when we hide the toolbar. Instead of a fixed toolbar, it's a good design that we can hide it whenever we want. - Yes. Because it was easier to see the toolbar when editing as it was on the side. - Yes. Having it in view, without having to scroll is better. Kind of like how most tool-heavy software does it. Like for instance Adobe Photoshop etc. - not much since I need to share two screens for doing tasks, then the buttons go back to the bottom of the screen, which is same as before - Yes. It's more convenient and user-friendly. - It didn't make much of a difference for me. - Yes, useful when I need to see the whole image. - Yes, less clutter and distractions these objects that are not needed when calculating likelihood and cleaning up the diagram.

Feedback Survey

What do you dislike in the re-engineered tool, and how could they be better?

#	Question	Text
11	What do you dislike in the re-engineered tool, and how could they be better?	- The zoom feature. I prefer using the mouse wheel. The switch button makes me a little confused the first time I use it. Maybe it can be located on the pop-up box when we click on the element. - When entering the likelihood values the input field rejected commas, without warning. I would have liked to receive a warning, or even better, the tool should accept both , and . - The elements should auto scale to the screen size if possible. In my 1080 resoultion, there was some overlap. Zooming out resulted in a very little horiztonal scaling. So people with very large screens cannot use their screen real estate for a larger workspace. - It could be more instructions on how people should use it. The button of frequency and probability could be moved into the box where we fill the info. It would be nice the error message can tell why there is an error in the likelihood calculation - The likelihood reset function. It would be more helpful if a single likelihood value could be reset freely. - when an element is modified, it would be saved by clicking 'enter' rather than having to click on the save button. - I like all, and in error-message part may give more advice about how to fix the errors. - Nothing I can think about now.

Feedback Survey

What additional features in the re-engineered tool do you want to have?

#	Question	Text
12	What additional features in the re-engineered tool do you want to have?	- I think it will be much better if we can undo or redo our change to the model graph. - I would like an option to clear likelihood values that are generated by the tool, in addition to the button that clears all likelihood values. - Possibility of increasing the size of the bubbles that houses text. Sometimes the amount of text is simply too much and overflow. - different colors or formats of arrows to show different relations. description text can be automatically adjusted in the box field. - Choose the type of different relations when calculating. - no idea - Scrolling support.

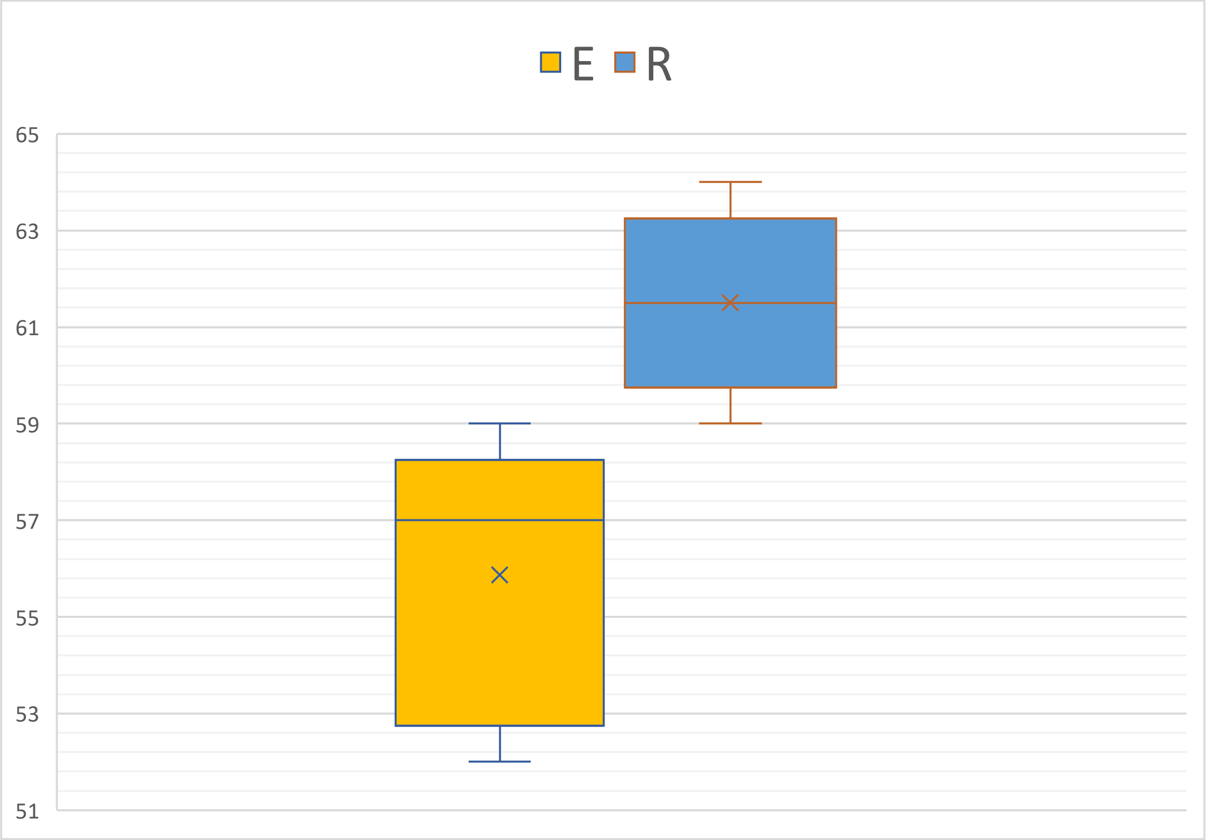
Appendix G

Statistics Calculations

G.1 Total score

	E	R
Mean	55,875	61,5
Median	57	61,5
Vanriance	9,268	4,286
SD	3,044	2,070
SE	1,076	0,732
Skewness	-0,450	0,000
Kurtosis	-1,972	-1,792

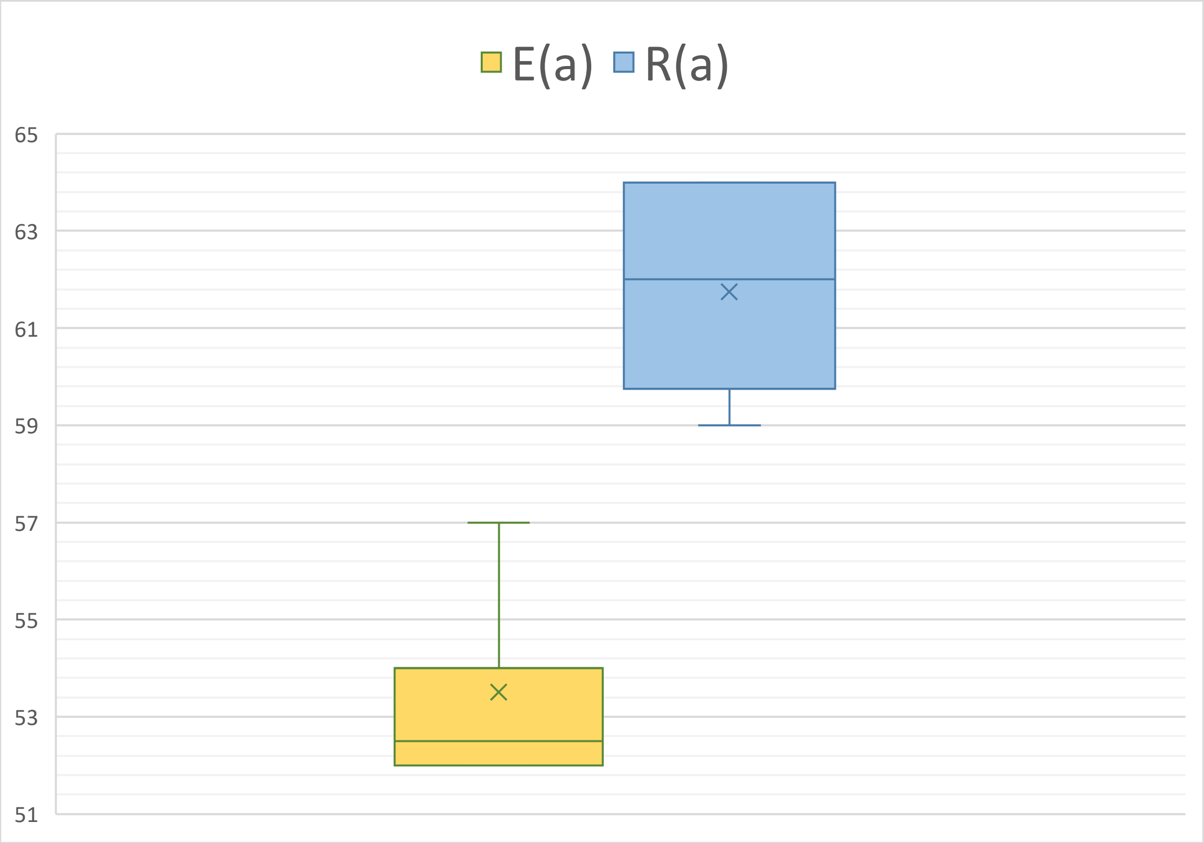
t-Test: Paired Two Sample for Means		
	E	*R*
Mean	55,875	61,5
Variance	9,268	4,286
Observations	8	8
Pearson Correlation	0,30600918	
Hypothesized Mean Difference	0	
df	7	
t Stat	-5,109	
P(T<=t) one-tail	0,001	
t Critical one-tail	1,895	
P(T<=t) two-tail	0,001	
t Critical two-tail	2,365	


E
R
65
63
61
59
57
55
53
51

G.2 Total score – Group A

	E(a)	R(a)
Mean	53,5	61,75
Median	52,5	62
Vanriance	5,667	6,917
SD	2,380	2,630
SE	1,190	1,315
Skewness	1,779	-0,124
Kurtosis	3,135	-5,290

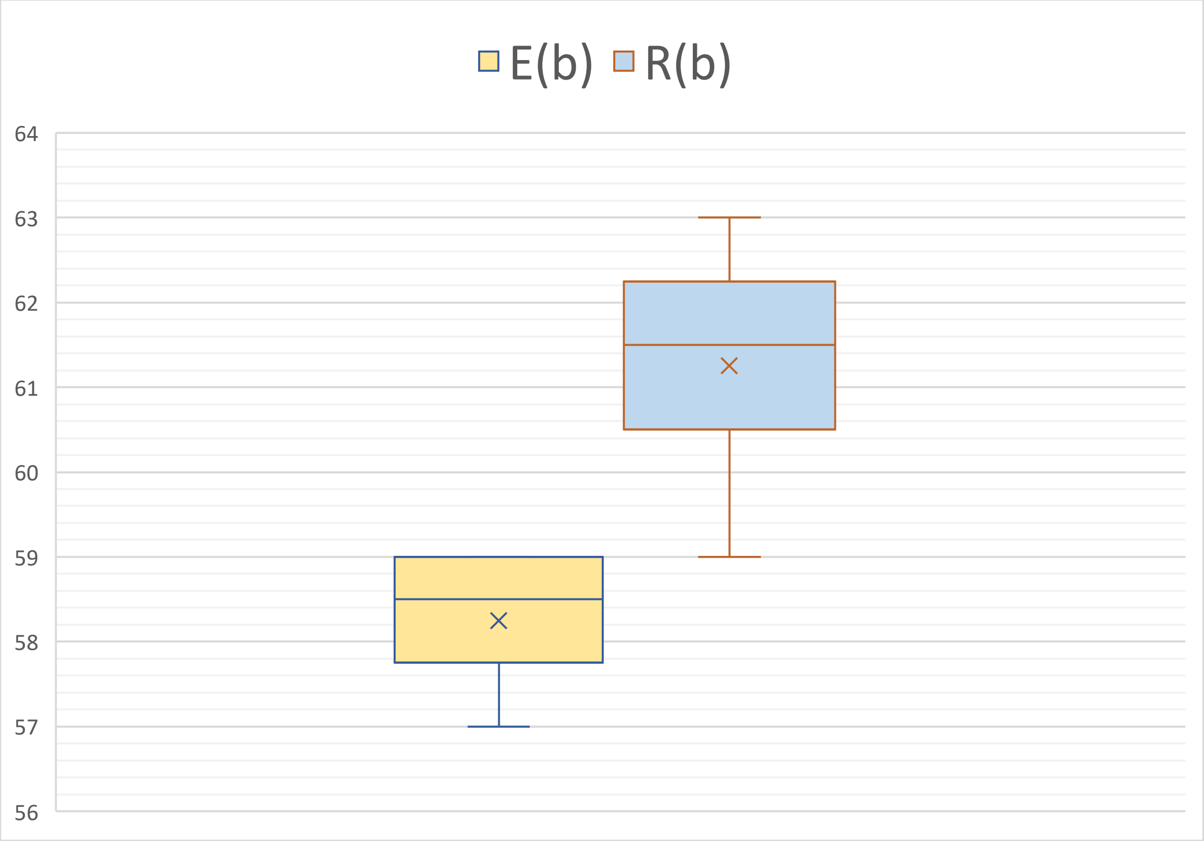
t-Test: Paired Two Sample for Means		
	E(a)	*R(a)*
Mean	53,5	61,75
Variance	5,667	6,917
Observations	4	4
Pearson Correlation	0,718787072	
Hypothesized Mean Difference	0	
df	3	
t Stat	-8,716	
P(T<=t) one-tail	0,002	
t Critical one-tail	2,353	
P(T<=t) two-tail	0,003	
t Critical two-tail	3,182	


E(a)
R(a)
65
63
61
59
57
55
53
51

G.3 Total score – Group B

	E(b)	R(b)
Mean	58,25	61,25
Median	58,5	61,5
vanriance	0,917	2,917
SD	0,957	1,708
SE	0,479	0,854
Skewness	-0,855	-0,753
Kurtosis	-1,289	0,343

t-Test: Paired Two Sample for Means		
	E(b)	*R(b)*
Mean	58,25	61,25
Variance	0,917	2,917
Observations	4	4
Pearson Correlation	0,968329664	
Hypothesized Mean Difference	0	
df	3	
t Stat	-7,348	
P(T<=t) one-tail	0,003	
t Critical one-tail	2,353	
P(T<=t) two-tail	0,005	
t Critical two-tail	3,182	


E(b)
R(b)
64
63
62
61
60
59
58
57
56