

Visualizing and Formalizing Risk Information: An Experiment

Jyrki Kontio^{*}, Jani-Pekka Jokinen^{*}, and Esa Rosendahl⁺
jyrki.kontio@hut.fi, jani-pekka.jokinen@hut.fi, ejrosend@surfeu.fi

^{*} Helsinki University of Technology
Software Business and Engineering Institute
P.O.Box 9600, 02015 HUT, Finland
<http://www.soberit.hut.fi/SWBiz/>

⁺ Finnish Institute of Occupational Health
Rautkallionkatu 4 A 9
01360 Vantaa, Finland

Abstract

An essential element of software engineering risk management is the conceptualization of potential risks to a project. It is the basis of risk analysis and, even more importantly, it strongly influences how risks are communicated and understood by participants in a project. This paper reports the results of a study where different risk visualization and documentation methods were compared in a controlled experiment with students. The study indicated that a defined and sufficiently expressive visualization approach can help capture more of the risk information than less formal methods. At the same time, participants felt that the more formal approaches were not more difficult to either learn or use than less formal ones. The SEI risks statements turned out to be inferior to other methods in most comparisons.

1 Introduction

Risk management is an integral part of software project management, yet most practitioners rely on intuition when actually managing risks: software engineering risk management does not use well-defined methods or processes in dealing with project uncertainty [Chittister et al., 1992; Kontio, 1995; Ropponen, 1999]. Yet, the unplanned events and “surprises” often cause projects substantial problems or even project failures [Flowers, 1996; Glass, 1997].

Risk management requires that risks are identified and understood as well as possible. As risks deal with future outcomes, risk identification and analysis is a highly speculative and subjective activity, especially when performed in domains where historical trends and data provide little basis for estimating the future.

Furthermore, risk identification and analysis often takes place in a meeting or involves teamwork, requiring

communication and discussions between different participants. Hence, how risks are modeled and documented in this process is a central part of the risk management process. Implicit risk management makes it difficult to share and communicate risk information.

The traditional model for risk documentation has been based on the use of detailed forms, containing possibly dozens of items. These forms have often become practically impossible to maintain. At the other extreme, many practitioners document their risks using just a few words. This is cost efficient, but leaves the true meaning of risk vague and open to different interpretations.

We believe that the above problems can be substantially mitigated by developing and using better methods for modeling risk information. More specifically, we should (i) provide practitioners with the conceptual tools to capture risk information in such terms that the project situation can be understood, (ii) find ways to document risks more accurately, (iii) understand the challenges in effective risk communication, and (iv) find the most cost-effective methods for modeling risks.

We have conducted several qualitative studies in risk management and these studies have yielded valuable insights to the issues in risk management process and techniques [Englund, 1997; Freimut et al., 2001; Getto et al., 1999; Kontio et al., 1996b; Kontio et al., 1996a; Kontio et al., 1997; Kontio et al., 1998]. The experiment reported in this paper complements these studies by providing specific information about a particular, critical step in the risk management process: how risk information can be captured and modeled.

In this paper we report the results of an experiment that evaluated three different risk modeling methods: the Riskit Analysis Graphs, the SEI risk statements, and risk forms. The motivation for the study was two-fold. First, we wanted to evaluate how the three main approaches for documenting risks compared to each other with respect to their accuracy, effectiveness, and ease of use. Second,

we also wanted to compare the three different implementation techniques for the Riskit Analysis Graphs – use of blank whiteboard, use of software, or use of laminated risk element symbols.

In the following, we will first present the different modeling approaches (chapter 2). We will then present the experiment design and arrangement (chapter 3) and the experiment results on the main methods (chapter 4), and Riskit-based methods (chapter 5) leading to our conclusions in chapter 6.

2 The Risk Modeling Methods Used

2.1 The Riskit Analysis Graphs

The Riskit method is a comprehensive and fully documented approach for risk identification, analysis, and control [Kontio, 1997; Kontio, 2001]. Compared to other risk management approaches in software engineering [Boehm, 1991; Charette, 1989; Dorofee *et al.*, 1996; Hall, 1998], its main novel characteristics include the following:

- it includes stakeholder perspectives in all aspects of risk management cycle,
- it uses more sound, i.e., less biased, methods in risk prioritization, and
- it uses a well-defined conceptual model to document risk information graphically.

This experiment focuses on the last point. The Riskit Analysis Graph can be seen both as a conceptual framework for structuring risk information, as well as a visual notation for documenting risk scenarios. We suggest that the Riskit Analysis Graphs may be one effective method for mitigating the risk modeling problems listed in the previous chapter. As the Riskit Analysis Graphs are well defined in other reports

[Kontio, 1997], we will only present a summary of them here.

As can be seen in Figure 1, the Riskit Analysis Graph contains different symbols to model different aspects of risk. A *risk factor* is a characteristic that affects the probability of a risk event occurring. A risk factor describes the characteristics of an environment. Consequently, in the Riskit Analysis Graph, a risk factor does not have probability associated with it, it describes a relevant environment characteristic as it is or will be. Risk factors that are documented typically increase the probability of risk events occurring, but they may also reduce them, i.e., they are *success factors* for a project.

A *risk event* represents an occurrence of a negative incident – or a discovery of information that reveals negative circumstances. Risk event is a stochastic phenomenon, i.e., it is not known for certain whether it will happen or not. This uncertainty can be characterized by a probability estimate associated to the risk event. Each risk event can be influenced by many risk factors but a risk event does not have to have a risk factor associated with it.

The next element in Figure 1 is called *risk outcome*. It represents the situation in a project after the risk event has occurred but before any corrective action is taken to reduce the effects of a risk event. The purpose of the concept of risk outcome is to document the immediate results and situation after the risk occurs. Based on the risk outcome description, different reactions can sometimes be considered more objectively and creatively than directly from a risk event.

When a risk event occurs, the resulting risk outcome is rarely accepted as such. Instead, organizations react to the situation to reduce the negative impact of the risk event. These corrective reactions are an important part of understanding what is the overall impact of the risk

event to the project domain. Thus, each risk outcome is associated with one or more risk reactions: a *risk reaction* describes a possible action that can be taken as a response to risk event and resulting risk outcome. If only one risk reaction is described, it is deterministic: it will be taken if the event occurs. If more than one reaction is described, they represent alternative lines of actions.

The *risk effect set* represents the final impact of a risk event to the project. In other words, it

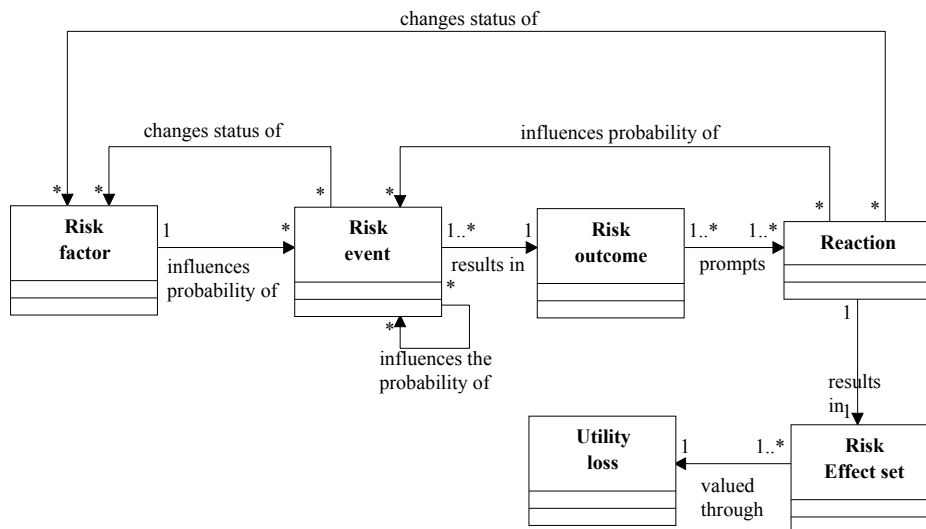


Figure 1: The full Riskit Analysis Graph

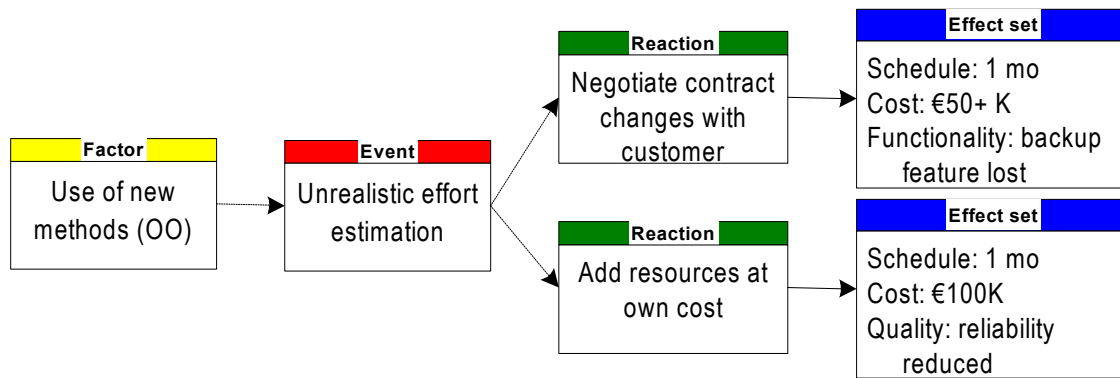


Figure 2: Example of the Riskit Analysis Graph

documents what characteristics of the project were affected, taking into account the impact of reactions. Effects are described through the explicitly stated goals for the project.

While the risk effect represents the impact the risk had on each project goal, the concept of *utility loss* captures how severe the overall impact of effects is. The concept of utility loss is based on the utility theory, a widely used theory in economics and management science [Von Neumann *et al.*, 1944; French, 1989]. The use of utility theory allows the simultaneous consideration of multiple criteria and consideration of several stakeholders. Furthermore, it is likely to result in more realistic evaluation of the losses as the utility functions of stakeholders are generally believed to be non-linear [Friedman *et al.*, 1948; Boehm, 1981] and there may be points of discontinuity in them.

In this study we used three different ways to model risks, based on the Riskit Analysis Graphs. First, we had implemented the Riskit Analysis Graph symbols on a commercial drawing tool called VISIO. This tool gives users the possibility to “drag and drop” symbols quickly and easily on a computer. Figure 2 shows an example of such a diagram.

The second method we used was based on drawing Riskit Analysis Graphs on a blank flipchart. Obviously, drawing on a blank sheet requires that participants remember the symbols without being prompted and there are less possibilities for editing changes and for automatic storing of such drawing.

The third Riskit based approach was to use magnetic, laminated plastic frames on a whiteboard. The different symbols were color-coded and named, with space to write the description with a marker. This technique is referred to as the RiskitFrames technique later in this paper.

2.2 Risk Forms

We used a paper-based form as the fourth technique for capturing risk information. An example of such a form is given in appendix A. The form was intentionally

made semantically compatible with the Riskit Analysis Graphs as we were interested in discovering differences in graphical vs. form-based risk documentation. The form is similar to the ones presented in the literature [Boehm, 1991].

2.3 Risk Statements

The SEI method documents risks using risk statements [Dorofee *et al.*, 1996; Gluch, 1994; Sisti *et al.*, 1994]. Risk statements document risks in condition – consequence pairs. The condition attribute contains a sentence describing the situation and the consequence attribute describes the outcome of the current condition if a risk occurs. The risk statements can be used visually, as shown in Figure 3, or on textual basis, as presented below:

(Given the condition that)
 we must use Java and we have little experience in it;
(then there is a concern that)
 the implementation phase may last longer than planned.

The SEI risk statements, actually, have three elements [Gluch, 1994]: condition, transition, and consequence. However, the transition is often not documented. The SEI risk statements elements have some correspondence to Riskit Analysis Graphs: condition corresponds to risk factor, transition to risk event, and the consequence either to outcome or to effect set.

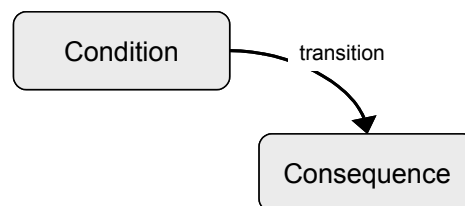


Figure 3: Visual format for the SEI risk statements

3 Study Objectives, Design and Arrangements

3.1 Objectives and Metrics

We used the GQM method [Basili *et al.*, 1984; van Solingen *et al.*, 1999] to formulate the experiment goals more accurately and to derive metrics for the study, as shown in Table 1. The metrics defined based on the GQM goals are listed in appendix B, and they included the measurement of time to complete a risk scenario, number and type of elements documented in risk scenarios, number of discussed and but undocumented issues, as well as user feedback on usability, ease of use, effectiveness, and amount of overhead.

We define effectiveness as how well a technique supports all essential aspects of risk documentation without overhead or unnecessary activities. By accuracy we mean how well a technique captures the information produced in a session. Ease of use is defined as the participant perception of the ease of use.

Table 1: GQM statement for the study

<i>Analyze</i>	different methods for modeling risks in a risk analysis meeting, first by using 1a) Riskit Analysis Graph –based techniques, 1b) using forms to capture risk information, and 1c) using the SEI Risk statements; and second, using the three Riskit-based variants 2a) software with video projection, 2b) blank whiteboard, and 2c) laminated Riskit Analysis Graph frames (RiskitFrames);
<i>In order to</i>	comparative analysis on them;
<i>with respect to</i>	their effectiveness, accuracy and ease of use;
<i>From perspective of</i>	risk management method developer and risk management process owner.
<i>in context of</i>	university software project course

Note that the magnitude of risk is not a controlled variable in this study as (i) risk prioritization is meant to be done only when risks are understood, (ii) given the large number of brainstormed risks, the selected five scenarios are likely to be sufficiently similar from the point of view of modeling, and (iii) any changes in risk magnitude are randomly assigned to methods, as far as the participants are concerned, due to the predefined order of method use. Furthermore, it is practically impossible to normalize risk magnitude between different projects and stakeholders in the first place.

3.2 Instrumentation

Several methods were used to obtain this metric data: interview questionnaires were used to capture participant opinions, postmortem artifact analysis was used to analyze how well information was documented, and

video recordings were used to measure the time as well as to capture the undocumented issues.

The information from the video recordings was obtained during two successive viewings. Time information was obtained during the first viewing using three clocks. The purpose of the second viewing was to capture and compare the number of unclear issues and to calculate and compare the total number of “points” made during the discussion of a particular risk scenario with the number of “points” documented in that scenario. The “points” here refer to individual and separate pieces of risk information regarding any aspect of a risk scenario. For example, a risk scenario could have the following consequences: 1) “x” amount of extra work is required, 2) a deliverable will be delivered 2 weeks late, 3) the client will lose confidence in us, etc. These three points or pieces of information would be counted separately.

We also used a simple questionnaire to obtain participant perceptions on the usability, ease of use, and effectiveness of the methods. We did not provide definitions for “usability” and “ease of use” in the questionnaire, i.e., we used these slightly overlapping concepts, essentially using triangulation to obtain two points of view on usability. However, the effectiveness was defined in the questionnaire.

3.3 Arrangements

Participants in the study were students who were taking the class “Software Project”. This course is a two-term (ca. nine months), five-credit course, where students work through major software projects in groups of six to seven people [HUT, 2001]. Each project contains all typical software design and implementation phases, such as requirements analysis, conceptual and detailed design, coding, testing, documentation, and delivery to the customer. Each student is expected to contribute five working weeks of effort to their projects, i.e., usually projects plan their work to contain approximately seven to ten person months. However, quite often these initial effort estimates are exceeded. The course provides a basic process, templates and milestones for the main activities of projects. As most students at this point in their studies already have industry work experience, we believe that the student participants, as well as the projects they were involved, were not substantially different from small projects in the industry.

We asked volunteer project teams to take an additional course along with the software project course. This additional course was a risk management course and contained performing systematic risk management in the projects of the software project class. Each project team received training on the basic concepts of risk management and on the methods needed to perform risk management and to participate in this study. Each

participating student received two credits for this course. Four teams and a total of 18 students signed up for the course and participated in the experiment. However, only 14 participants returned the questionnaire about method preferences. In order to avoid the author bias, the training was not given by the Riskit method developer, but by a course assistant.

Each project had prepared a risk management plan and documented its goals and risk management mandate before the risk identification and analysis session. We held separate sessions for each team and the sessions had similar structure. At the beginning of the session, the participants were given a short briefing on the five documentation methods to be used and on the agenda of the session. Risk identification session was held to identify potential new threats to their projects. Based on the identification results, the five risks to be documented and used as scenarios in the experiment were selected from these newly identified raw risks or from raw risks that had been identified on earlier risk management sessions. Selection was not random, it was based on what the team as a whole considered to be the most important risks to be analyzed in more detail.

Table 2: Order of using the modeling techniques in the group sessions

	Group A	Group B	Group C	Group D
Scen. 1	SEI risk statements	SEI risk statements	Riskit-Frames	Risk forms
Scen. 2	Riskit Software	Riskit Drawing	Risk forms	Riskit Software
Scen. 3	Riskit Drawing	Riskit-Frames	Riskit Software	Riskit Drawing
Scen. 4	Riskit-Frames	Risk forms	Riskit Drawing	Riskit-Frames
Scen. 5	Risk forms	Riskit Software	SEI risk statements	SEI risk statements

After this, the groups analyzed and discussed each of the five risks in turn. Participants were told that they had approximately two hours of time available for the whole session and decided themselves when their analysis was complete for each scenario. The groups used different methods in different order to avoid systematic learning bias in the study results, as shown in Table 2. While this design results in unequal number of “treatments” per method, we used this design to obtain additional data points for the techniques in the available time. The items in bold mark the Riskit based method that was used in the first part of the analysis, i.e., to compare the Riskit based approach to other approaches. Note that in the comparison of different main methods we used the first Riskit-based method as the representative Riskit-based method to avoid maturity bias in favor of Riskit. All three different variants of the Riskit method are based on the same principles and we believe that their relative

differences were less significant than the differences between the three main methods.

3.4 Data Analysis Methods

We used parametric and non-parametric statistical techniques in the analysis of the study data, in addition to qualitative, contextual analysis we performed. The limited number of student groups produced essentially only four data points for each “treatment”. We used Wilcoxon’s signed rank test [Conover, 1999] to compare the techniques pairwise to test the hypothesis that techniques are not from the same sample. We also used paired-samples t-test to test the same hypothesis once again for the sake of credibility (if both variables were measured on interval-scale).

First we analyzed differences of three main methods. The results are presented in section 4. The differences of Riskit based methods (3) are presented in section 5.

Riskit based methods were used three times in each group and, therefore, we analyzed the influence of the instance of use by one-way variance analysis. Because of small data set, we did not find any statistically significant differences. However the data indicates that those differences are quite small. In order to eliminate the bias, which repeated use of Riskit based methods may cause, we select only observations of the first instance of use to the comparison of the main methods.

4 Comparison of the Main Methods

4.1 Effectiveness

The teams spent different amounts of time discussing their scenarios, team C being the fastest, and team B taking the most time, as shown in Table 3. We initially had prepared to monitor the amount of time spent on discussing the method usage but as the method related questions only accounted for a small amount of time in sessions (one and a half minutes on the average), this data was not used in the study. Also, we measured the time spent on general or non-related discussions and, as a result, produced the data on effective time, as shown in Table 3.

Table 3: Effective time per scenario (in minutes)

Methods	Group A	Group B	Group C	Group D	Average
Riskit	6.5	13.3	3.7	6.5	7.5
Risk forms	5.0	6.6	2.7	6.1	5.1
SEI risk statements	7.9	3.1	1.9	5.3	4.5
<i>Total (average for the last column)</i>	19.4	23.0	8.3	17.9	5.7

The average times indicate that the SEI statements were fastest and Riskit slowest to complete. However,

statistical analysis did not reveal these differences to be significant. However, using the Wilcoxon test, the difference between Riskit and risk forms was significant with a p value of 0.068, the corresponding p value with a t test was 0.197. It should be noted, however, that the Riskit scenario by group B resulted in more than twice the time of any other scenario and it is possibly a more complex risk scenario, i.e., an outlier, and this influenced the method averages.

We studied the efficiency of the methods by calculating the ratio between the number of information entities produced and the effective time used for each risk scenario. This data is presented in Table 4.

The analysis in Table 4 concerns only the effective time used in constructing the scenarios. It does not include any information about what was produced during analysis. We calculated the ratio between effective time and number of risk items produced during the analysis.

Table 4: Efficiency of methods: number of information entities / effective time

Methods	Group A	Group B	Group C	Group D	Average
Riskit	1.4	1.5	1.1	1.2	1.3
Risk forms	1.6	1.5	1.5	1.7	1.6
SEI risk statements	1.4	1.0	1.6	1.1	1.3
<i>Total (average for the last column)</i>	4.4	4.0	4.2	4.0	1.4

The averages in Table 4 show that risk forms were most effective. However, the lowest p value of 0.109 using Wilcoxon ($p=0.089$ with t test) was between Riskit and risk forms, indicating lack of statistical significance.

We also assessed the effectiveness of the different method by asking participants opinions about effectiveness. The results of this data are shown in Table 5, where the number in each cell indicates how many respondents replied with that value or preference in the questionnaire. The column “weighted average” contains the weighted average of the rank numbers and corresponding frequencies.

Table 5: Effectiveness of techniques

Methods	Rank frequency			Average
	1	2	3	
Riskit	7	5	2	1.64
Risk forms	5	7	2	1.79
SEI risk statements	2	3	9	2.50

The rankings of the methods were significantly different from each other (with a p value of 0.042 on Friedman’s test). Kendall’s coefficient of concordance was 0.227, which means that the participants gave very different rankings for the methods. We also used Friedman’s test in paired comparisons between the

methods. This is a recommended method if the null hypothesis that the rankings are equal is rejected (as it is above with a p value of 0.042), but such p values should be interpreted cautiously [Conover, 1999]. This comment applies also to the statistical analysis of tables 8, 9, 12, 15 and 16, where preferences of participants are presented and analyzed.

The statistical analysis showed that the participants considered Riskit to be more effective than the SEI statements (with a p value of 0.033 on Friedman test). The risk forms were also considered more effective than the SEI statements, but the corresponding p value was 0.052.

4.2 Accuracy

We also recorded data on how well the information that was raised during the discussion was actually recorded in the documentation of the scenarios. We performed two statistical analyses on this data. First, we wanted to compare whether techniques possibly prompted more information to be discussed during the session. This is reflected by the item “information produced”. Second, we want to study how well different techniques succeed in capturing the information that was raised. This was analyzed using the “information captured / produced” ratio as the metric. The rationale for this metric was that we wanted to compare the methods with respect to how accurately they capture the content of the discussion in the documentation. Any information not documented at a session is likely to be ignored in future discussions about the risks. Therefore, methods that capture all or most of the discussion content were considered better methods. Table 6 presents the data on what information was produced and Table 7 presents the data on how well the information was captured.

Table 6: Information produced

Methods	Group A	Group B	Group C	Group D	Average
Riskit	9	20	4	8	10.3
Risk forms	8	10	4	10	8.0
SEI risk statements	11	3	3	6	5.8
<i>Total (average for the last column)</i>	28	33	11	24	8.0

The simple mathematical averages indicate that Riskit method prompted the most information and the SEI statements the least. However, we found no statistical significance in these differences.

Table 7: Information captured / produced

Methods	Group A	Group B	Group C	Group D	Average
Riskit	0.78	0.90	1.00	0.88	0.89
Risk forms	0.88	0.90	1.00	1.00	0.94
SEI risk statements	0.45	1.00	0.67	0.83	0.74
<i>Total (average for the last column)</i>	2.11	2.80	2.67	2.71	0.86

The average data in Table 7 indicates that risk forms were most successful in capturing the points raised during the discussion, followed by the Riskit method and SEI statements. However, these differences were not statistically significant.

4.3 Ease of Use

Finally, we also surveyed how the participants perceived the various five different methods with respect to usability and ease of use. Usability referred to how convenient and practical the method is to use and ease of use specifically attempted to clarify how easy and intuitive the method was perceived. Participants were asked to rank the methods in order of preference, i.e., a low number indicated a preferred method. The frequency data on usability rankings is presented in Table 8.

Table 8: Usability of techniques

Methods	Rank frequency			Weighted Average
	1	2	3	
Riskit	11	2	1	1.29
Risk forms	2	12	0	1.86
SEI risk statements	1	0	13	2.86

The rankings of the methods were significantly different from each others (with p value of 0.0001 on Friedman's test). Kendal's coefficient of concordance was 0.633, which means that the participants had quite a high agreement in rankings.

The data in Table 8 shows that that Riskit was preferred over risk forms (with Friedman's p value of 0.033), and the SEI statements (p=0.001). The risk forms were preferred over SEI statements with p value of 0.001.

The ease of use preferences were obtained with a similar question and the resulting data is presented in Table 9.

Table 9: Ease of use of techniques

Methods	Rank frequency			Weighted Average
	1	2	3	
Riskit	7	7	0	1.50
Risk forms	1	6	7	2.43
SEI risk statements	6	1	7	2.07

The rankings of the methods were significantly different from each others (with p value of 0.046 on Friedman's test). Kendal's coefficient of concordance was 0.219, which means that the participants gave very different rankings for the methods.

Data in Table 9 indicates that the Riskit method was preferred over risk forms with p value of 0.001. The preference of Riskit over SEI statements only gave a p value of 0.59, i.e., that preference is not statistically significant.

5 Comparison of Riskit-based Methods

The second research question in our study was to compare the three Riskit based methods to find out whether the software-based, drawing on a blank sheet or use of hardcopy templates resulted in differences between the methods. This section presents this data and corresponding analyses.

Data was obtained from the same sessions and questionnaires, leaving out the other methods from the analysis. Consequently, all metrics and instrumentation was the same as in the previous section.

5.1 Effectiveness

The effectiveness data is presented in Table 10 and the averages indicate that the software was fastest to use while the RiskitFrames were slowest. However, these differences were not significant. It should also be noted that the group B spent the most time in total in the analysis.

Table 10: Effective time per scenario (in minutes)

Methods	Group A	Group B	Group C	Group D	Average
Software	6.5	9.1	9.3	6.5	7.8
Drawing	11.7	13.3	5.7	7.5	9.5
RiskitFrames	5.9	21.3	3.7	13.1	11.0
<i>Total (average for the last column)</i>	24.1	43.7	18.7	27.1	9.5

The efficiency data is presented in Table 11 and, while the averages indicate that the drawing method was the most effective, the pairwise comparison over RiskitFrames resulted in p value of 0.109, i.e., the differences are not statistically significant.

Table 11: Efficiency of methods: number of information entities / effective time

Methods	Group A	Group B	Group C	Group D	Average
Software	1.4	0.7	0.8	1.2	1.0
Drawing	1.1	1.5	1.1	1.3	1.3
RiskitFrames	1.0	0.8	1.1	0.8	0.9
<i>Average</i>	1.2	1.0	1.0	1.1	1.1

The participant preference data on effectiveness is presented in Table 12. The rankings of the methods were significantly different from each others (with p value of 0.001 on Friedman’s test). Kendal’s coefficient of concordance was 0.474, which means that the participants had halfway agreement in rankings. The data indicates that the software method was preferred over drawing (p=0.007) and RiskitFrames (p=0.035). The RiskitFrames was preferred over drawing (p=0.007)

Table 12: Effectiveness of techniques

Methods	Rank frequency			Weighted Average
	1	2	3	
Software	12	1	1	1.21
Drawing	4	0	10	2.43
RiskitFrames	4	10	0	1.71

5.2 Accuracy

The accuracy data analysis did not produce any statistically significant differences between the methods, but the resulting data is presented in Table 13 and Table 14. In terms of averages, the drawing method seemed to produce most information and the software least; and the software also seemed to be most effective in capturing most of the points in the discussion, while drawing was least successful in this.

Table 13: Information produced

Methods	Group A	Group B	Group C	Group D	Average
Software	9	6	7	8	7.5
Drawing	13	20	6	10	12.3
RiskitFrames	6	18	4	11	9.8
<i>Total (average for the last column)</i>	28	44	17	29	9.8

Table 14: Information produced / captured

Methods	Group A	Group B	Group C	Group D	Average
Software	0.78	1.00	1.00	0.88	0.91
Drawing	1.00	0.90	0.67	1.00	0.89
RiskitFrames	1.00	0.94	1.00	0.91	0.96
<i>Total (average for the last column)</i>	2.78	2.84	2.67	2.79	0.92

5.3 Ease of Use

As in the comparing the main method, the Riskit based methods usability and ease of use preferences were asked from participants.

The usability data is presented in Table 15. The rankings of the methods were significantly different from each others (with p value of 0.003 on Friedman’s test). Kendal’s coefficient of concordance was 0.418, which means that the participants had almost halfway agreement in rankings. The data indicates that software is preferred over drawing (p=0.013) and RiskitFrames is preferred over drawing (p=0.002).

Table 15: Usability of techniques

Methods	Rank frequency			Weighted Average
	1	2	3	
Software	9	3	2	1.50
Drawing	2	1	11	2.64
RiskitFrames	5	9	0	1.64

The ease of use data is presented in Table 16. The rankings of the methods were significantly different from each others (with p value of 0.025 on Friedman’s test). Kendal’s coefficient of concordance was 0.264, which means that the participants gave very different rankings for the methods. The data indicates that software is preferred over drawing (p=0.052) and RiskitFrames over drawing (p=0.013).

Table 16: Ease of use of techniques

Methods	Rank frequency			Average
	1	2	3	
Software	8	3	3	1.64
Drawing	2	3	9	2.50
RiskitFrames	6	7	1	1.64

6 Conclusions

In this concluding section we present the main findings, discuss the generalizability of the results, and indicate some areas for further research.

6.1 Differences between Main Methods

This study discovered some differences between the methods that were used, despite the small number of data points available.

The relative strengths of the Riskit method seemed to be ease of use, tendency to prompt and capture more points during the discussion, and perceived effectiveness. It seemed to be more time-consuming, though.

The risk forms seemed to be accurate in capturing points made during the discussion.

The SEI statements were not performing better in any of the evaluations we made, and performed worst in several areas: perceived effectiveness, usability, and ability to capture information from the discussion.

The results indicate that in a situation where more detailed discussion about risks is needed, the Riskit method should be considered, as it seems to encourage more detailed discussion about risks. High-risk projects are likely to benefit from such approach.

However, if there is a need to be cost-effective and the project is not particularly risky, the risk forms should be considered. They seem to produce results fast and capture relevant information accurately.

While the SEI statements seemed to perform worse than the other methods, it may be a useful approach in situations where ease of use and cost-efficiency need to be given special consideration.

6.2 Differences between The Riskit-based Methods

The comparisons between the three different variants of the Riskit method revealed small differences between the methods.

The software supported variant seemed to be most effective and most usable.

It is also interesting to note that the RiskitFrames were preferred over drawing on a blank sheet in terms of usability.

Not surprisingly, the drawing method seemed to promote the most points during the discussion and be least effective in capturing the discussion content. This, of course, could be compensated by having a scribe, perhaps using a software version of the method, to support the capture of the rich set of information surfacing in a discussion.

6.3 Discussion of Validity and Generalizeability

The main limitation of the experiment described in this paper is that there were quite few data points available to result in statistically meaningful results. Lack of such significance reduces the generalizability of results.

We were initially also somewhat concerned about the potential bias for the Riskit method, as participants may have unconsciously given preferable ratings to a “local” method or the researchers involved may have interpreted or measured some data favorably for Riskit. We believe that the several methods taken to control such biases: facilitation and measurements were done by a non-Riskit author, objectivity was emphasized to participants, participants grades were not dependent on their answers, answers were anonymous, and participants were

conducting risk management on projects that were real and concrete for them. Nevertheless, we cannot rule out the existence of such bias.

The problems associated with the SEI risk statements may be partially due to this set up. However, some of the metrics were independent of the underlying method. For instance, the information produced versus captured data recorded all information that was presented in the meeting, regardless of whether it was Riskit or SEI “compatible”. Another potential bias in the student group may have been that students may have received more exposure to Riskit method than to the SEI method. Assuming that this exposure would have resulted in favorable bias for the Riskit method, the student feedback on the perception question might have been affected by this. We did not observe any such impact but cannot rule such an effect out, either.

Based on our experience, the overall arrangements and design of this study are reasonably close to situations in young and small software companies and small projects with relatively young personnel. As our experience also indicates that even experienced software professionals rarely have received any formal risk management training and that the use of formal risk management methods is relatively rare [Ropponen, 1999], it seems that the situation and context in the study may also be applicable to larger projects and more experienced personnel.

However, the study did not address the issues of dealing with large number of risks over long periods of time. We hypothesize that the form-based approaches may provide benefits in terms of standardizing the risk documentation but we also think that the low values of ease of use, lack of visualization, and perceived appearance of bureaucracy may create an overhead that de-motivates project personnel to use forms. Correspondingly, the SEI statements have reportedly been used in large projects and, thus, they may have benefits that are more apparent in such larger contexts.

6.4 Future Research

This experiment suggested that the risk modeling methods seem to matter: even with small sample size, differences between the methods emerged. Further research on the types of differences is needed to better understand how to uncover, model and manage risks more effectively.

As this study suggested some potentially important differences between the methods, it is also necessary to replicate this experiment with a larger number of students to increase the significance level of the findings. In fact, we are planning and executing such a study in the near future.

The software-based risk documentation and management support, in particular, offers a potential high payoff, since students were only considering benefits from the point of view of simple drawing aids. The automatic capture of the risk information during the graphical editing function offers substantial additional elements for sharing information, tracking and updating information, and using the information for process improvement purposes. Further research towards such a tool would also be of value to risk management practitioners in the industry.

7 Acknowledgements

We would like to thank Mr. Jari Vanhanen for allowing this experiment to take place in the course he was teaching. We are particularly grateful to numerous anonymous referees that have pointed out errors and omissions in earlier versions of this paper, thus helping us improve this paper.

References

- Basili, V.R., Weiss, D.M., A Methodology for Collecting Valid Software Engineering Data, *IEEE Transactions on Software Engineering*, 10(6):728-738, 1984.
- Boehm, B.W., *Software Engineering Economics*, Prentice Hall, 1981.
- Boehm, B.W., *Software Risk Management: Principles and Practices*, *IEEE Software*, 8(1):32-41, 1991.
- Charette, R.N., *Software Engineering Risk Analysis and Management*, McGraw-Hill, 1989.
- Chittister, C., Kirkpatrick, R.J., Van Scoy, R.L., *Risk Management in Practice*, *American Programmer*, 5(September):30-35, 1992.
- Conover, W.J., *Practical Nonparametric Statistics*, John Wiley and Sons, 1999.
- Dorofee, A.J., Walker, J.A., Alberts, C.J., et al, *Continuous Risk Management Guidebook*, Software Engineering Institute, 1996.
- Englund, H., *A Case Study to Explore Risk Management Methods*.(1997), Kungliga Tekniska Högskolan, Stockholm, Sweden.
- Flowers, S., *Software Failure : Management Failure : Amazing Stories and Cautionary Tales*, John Wiley & Son Ltd, 1996.
- Freimut, B., Hartkopf, S., Kaiser, P., Kontio, J., Kobitzsch, W., *An Industrial Case Study of Implementing Software Risk Management*. In *Proceedings of the European Software Engineering Conference*, 2001.
- French, S., *Readings in Decision Analysis*, Chapman and Hall, 1989.
- Friedman, M., Savage, L.J., *The Utility Analysis of Choices Involving Risk*, *Journal of Political Economy*, 56:279-304, 1948.
- Getto, G., Landes, D., *Risk Management in Complex Project Organizations: A Godfather-driven Approach*. In *Proceedings of the Project Management Institute (PMI) Conference 99*, 1999.
- Glass, R.A., *The Software Runaways*, Prentice Hall, 1997.
- Gluch, D.P., *A Construct for Describing Software Development Risks*. (Software Engineering Institute, 1994) CMU/SEI-94-TR-14, SEI Technical Reports.
- Hall, E.M., *Managing Risk: Methods for Software Systems Development*, Addison-Wesley Pub Co., 1998.
- HUT, Tik-76.115 Software Project, <http://mordor.cs.hut.fi/tik-76.115/98-99/projects.htm>, 2001
- Kontio, Jyrki, IWSED-95 Web pages, <http://www.cs.umd.edu/projects/SoftEng/ESEG/iwsed/iwsed95/>, 1995
- Kontio, J., *The Riskit Method for Software Risk Management, version 1.00*. (College Park, MD, University of Maryland, 1997) CS-TR-3782 / UMIACS-TR-97-38, Computer Science Technical Reports.
- Kontio, J., *Software Engineering Risk Management: A Method, Improvement Framework, and Empirical Evaluation*. Doctoral dissertation. (2001), Helsinki University of Technology, publisher: Center of Excellence, ISBN: 952-5136-22-1.
- Kontio, J., Basili, V.R., *Risk Knowledge Capture in the Riskit Method*. In *Proceedings of the 21st Software Engineering Workshop*, 1996a. NASA.
- Kontio, J., Basili, V.R., *Empirical Evaluation of a Risk Management Method*. In *Proceedings of the SEI Conference on Risk Management*, 1997. Software Engineering Institute.
- Kontio, J., Englund, H., Basili, V.R., *Experiences from an Exploratory Case Study with a Software Risk Management Method*. (College Park, Maryland, University of Maryland, 1996b) CS-TR-3705, Computer Science Technical Reports.
- Kontio, J., Getto, G., Landes, D., *Experiences in improving risk management processes using the concepts of the Riskit method*. In *Proceedings of the Sixth International Symposium on the Foundations of Software Engineering (FSE-6)*, pages 163-174, 1998.
- Ropponen, J., *Software Risk Management - Foundations, Principles and Empirical Findings*.(1999), University of Jyväskylä.
- Sisti, F.J., Joseph, S., *Software Risk Evaluation Method Version 1.0*. (Pittsburgh, Software Engineering Institute, 1994) CMU/SEI-94-TR-19, SEI Technical Report.
- van Solingen, R., Berghout, E., *Goal/Question/Metric Method -- A Practical Guide for Quality Improvement of Software Development*, McGraw-Hill, 1999.
- Von Neumann, J., Morgenstern, O., *Theory of Games and Economic Behavior*, Princeton University Press, 1944.

APPENDIX A: Risk form used in the study

ID:		Project:	
Owner:		Date reviewed:	
Priority:		Timeframe:	
Probability:		Loss:	
Description:			
Context:			
Factors:			
	Reaction:	Effect set:	
	Reaction:	Effect set:	
	Reaction:	Effect set:	
Risk mitigation strategy:			
Potential risk controlling actions:			
Selected risk controlling actions:	<i>Action</i>		<i>Responsible</i>
Risk metrics:	<i>Metric/indicator</i>		<i>Status/Value</i>
History:			
Closing date:	Closing Rationale:		

APPENDIX B: Metrics used in the risk documentation study

Metric and Description	Measurement
Elapsed time to complete a scenario: The time it takes to create a risk scenario for each scenario	Measured with a stopwatch from videotape, elapsed minutes
General discussion time in a scenario: Time during scenario discussion when the discussion strayed away from the scenario at hand.	Time within elapsed time that was spent on general discussion
Time spent on method questions: Time spent on discussing "what should we do now"	Measured from videotape
Effective time: Elapsed time minus the time spent on method questions and general discussion	Measured from videotape
Number of elements in scenario: Number of Riskit Analysis Graph elements that are produced in each scenario	Counted from produced risk scenarios (not done for SEI risk statements)
Time per element: The average time spent per Riskit Analysis Graph element	Number of elements / Effective time (Not done for SEI risk statements)
Information captured / information produced: Average of the amount of the information that is captured on risk scenario divided by the risk information produced (what the risk means, its effects, etc.). The "amount" of produced information is calculated by counting the "points" made during the discussion including the points that were forgotten to add to scenarios and the points that do not have a placeholder in them, but not including the points that were deliberately left out. The "amount" of captured information is calculated by counting the "points" recorded in risk scenarios or statements.	Measured from videotape by comparing discussion and produced risk scenarios
Number of goals in effect sets: The number of goals in effect sets of the scenarios	Counted from produced risk scenarios
Number of unclear issues: The number of unclear issues that arise, when using different documentation methods. Count two metrics: issues related to the use of the method and issues related to subject matter	Counted from the videotape by following the discussion
Usability: Open ended question.	Questionnaire
Ease of use: Open ended question	Questionnaire
Effectiveness: Open ended question	Questionnaire
Overhead: Open ended question	Questionnaire
Recommendation for further use: Open ended question	Questionnaire
Improvement suggestions: Open ended question, Improvement suggestions from the participants	Questionnaire