Empirical Software Engineering
https://doi.org/10.1007/s10664-023-10321-y

A new, evidence-based, theory for knowledge reuse in security risk analysis

Katsiaryna Labunets[1] · Fabio Massacci[2,3] · Federica Paci[4] · Katja Tuma[2]

Accepted: 15 March 2023

Abstract
Security risk analysis (SRA) is a key activity in software engineering but requires heavy manual effort. Community knowledge in the form of security patterns or security catalogs can be used to support the identification of threats and security controls. However, no evidence-based theory exists about the effectiveness of security catalogs when used for security risk analysis. We adopt a grounded theory approach to propose a conceptual, revised and refined theory of SRA knowledge reuse. The theory refinement is backed by evidence gathered from conducting interviews with experts (20) and controlled experiments with both experts (15) and novice analysts (18). We conclude the paper by providing insights into the use of catalogs and managerial implications.

Keywords Information security · Risk assessment · Empirical study · Knowledge reuse

Communicated by: Raula Kula

The list of authors is ordered alphabetically.

✉ Katja Tuma
k.tuma@vu.nl

Katsiaryna Labunets
k.labunets@uu.nl

Fabio Massacci
f.massacci@ieee.org

Federica Paci
federicamariafrancesca.paci@univr.it

[1] Utrecht University, Utrecht, The Netherlands

[2] Vrije Universteit Amsterdam, Amsterdam, The Netherlands

[3] University of Trento, Trento, Italy

[4] University of Verona, Verona, Italy

Published online: 25 May 2023

1 Introduction

Systematic knowledge reuse improves organizational effectiveness and competitiveness (O'Dell and Grayson 1998; Markus 2001; Dixon 2002). This is particularly critical for knowledge that is difficult to acquire such as security knowledge (Souag et al. 2015).

In this respect, several approaches exists towards reuse of existing security knowledge (e.g., security patterns Schumacher et al. (2006), CAPEC Barnum (2008), weaknesses CWE MITRE (2020b), automated program repair for vulnerabilities MITRE (2020a), ISO27x security controls, to name a few). In the design space, several studies have proposed security risk and threat analysis of software design representations (Abe et al. 2013; Almorsy et al. 2013; Berger et al. 2016; Shostack 2014; Deng et al. 2011) for example by using security knowledge bases with *catalogues* of security patterns to address specific threats (Riaz et al. 2017, 2016). In the professional practice, security knowledge reuse for security risk analysis takes the form of detailed structured methods (e.g. (Publicas 2012; of Standards and Technology 2012)) coupled with catalogues of security patterns matching threats to security controls (e.g. (Group 2021; of Standards and Technology 2020)). Yet due to the challenges of automating security risk analysis procedures Tuma et al. (2018) security experts still leverage knowledge bases during *manual* analyses Tuma et al. (2021).

A key aspect that is still missing is a general, empirically validated theory on which aspects are important in reusing security knowledge and for which kind of re-users.

For example, in the general domain of software patterns, it is taken for granted that software patterns help Gamma et al. (1995) but empirical evidence showed that the "Gang of Four" patterns have limited usability and do not help novices to learn about design Zhang and Budgen (2012). For security patterns, empirical studies were not able to demonstrate that security patterns improve the productivity of the software designers or the security of the design Yskout et al. (2015). In general, security patters trying to capture tacit knowledge exhibit varying degree of success Riaz et al. (2017).

The goal of this work is to propose and empirically evaluate a conceptual theory of knowledge reuse for *security risk analysis* (SRA for short). An important dimension of the work is to understand whether such knowledge can be equally used by novices and domain experts alike.

Towards a principled approach, Markus Markus (2001) has studied the theory of knowledge reuse and found that effective knowledge reuse requires both the transfer of explicit (i.e., captured, structured and disseminated) knowledge and the internalisation of tacit (experts' personal knowledge that has not yet been captured) knowledge. Yet, such knowledge bases have meaning only when they are processed by an "interpreter" Newell (1982) which must share the knowledge base (i.e. the security catalogue) but also the mental process that has been used to build it (i.e. the methodology).

Therefore capturing and packaging SRA knowledge bases requires careful consideration and is by no means a trivial process. In security, if the structure of a security catalog is not of high quality, it may end up introducing new security problems rather than eliminating them Markus (2001).

To achieve the goal of this study, we first aim to identify the key success factors:

RQ1: *What are the success factors in SRA knowledge reuse?*

Markus' theory on knowledge reuse Markus (2001) goes a step further and investigates the relation between knowledge repositories' features and different types of knowledge reusers. In particular, Markus emphasizes that different knowledge reusers (shared work producers, shared work practitioners, expertise seeking novices, and secondary knowledge miners) face different challenges in using knowledge repositories and therefore need different support.

For instance, shared work producers work in teams to produce knowledge for their own later reuse. The main difficulty that they experience is looking for information that is specific to their context (especially, when it is not available). In contrast, expertise seeking novices are people with an occasional need for expert knowledge (with no need to retain it in the long term). They have great difficulties in locating the knowledge they need, selecting the relevant information and apply them to their context.

In our scenario, an additional challenge is the cross-cutting nature of security, which means that different profiles of employees may need to be involved in both creating and using the catalogue of security patterns: one could be security expert but not domain expert or viceversa. Having both domain and security expertise might be hard and costly to get. Despite the common belief that security catalogs can facilitate the job of practitioners with little or no security expertise, there is little if none research about the *types of SRA knowledge-base reusers* and which knowledge base features support them in *effective reuse of security* knowledge. We measure effectiveness of SRA knowledge reuse with measures of *actual efficacy* and *perceived efficacy* (for details see Section 6.2). To achieve the main goal of the study, we propose to empirically measure the performance of different knowledge reusers (novices and domain experts):

RQ2: *What are the differences (in terms of use and effectiveness) between different types of security knowledge reusers?*

Past studies have shown that threat analysis requires a high manual effort Scandariato et al. (2015), demands the involvement of security and domain experts Cruzes et al. (2018), and has been proven time and again difficult to automate Tuma et al. (2020). As the landscape of security threats evolves, there is an increasing need to perform security risk analysis (e.g., latest BSIMM study reports an increased investment by more than 65% Group (2021)) but threat analysis practices are set back by a globally recorded shortage of the security workforce (CyberSeek 2019; Blažič 2021).

A question with a significant practical impact is whether a *satisficing solution* to this void of security expertise could be having domain experts that are *not* also security experts, performing a security assessment with a reasonable levels of quality when supported by security knowledge reuse such as a cataologue. For instance, in Tuma et al. (2021), security novices with no prior knowledge of threat analysis performed it with 70-80% precision, where as security experts achieved nearly perfect precision. However, compared to security experts, the intellectual demands for security novices (albeit potentially domain expert) to navigate and effectively use a very large catalog of security threats is surely greater. This may even be true for security experts that have never used a similar catalog. Intellectual demands are defined as normal cognitive load perceived by individuals in performing their work Gray and Meister (2004). Consistent with Knowles' Knowles (1970) ideas, highly demanding work produces a greater need for knowledge and triggers learning behaviors. Gray et al. Gray and Meister (2004) hypothesise (hyp. 3) that the impact on knowledge sourcing is stronger under condition of higher intellectual demands, which our study puts to the test.

To achieve the main goal of the study, we propose to empirically compare the performance of domain experts reusers on a challenging scenarios that they ahve not seen before, possibly supported or not supported by a security catalogue:

RQ3: *What is the impact of security knowledge reuse under various conditions of intellectual demands?*

To this aim we have adopted a grounded theory approach (Jedlitschka et al. 2014) to gather key success factors of security knowledge reuse from the trenches and validated their effects with controlled experiments.

1.1 Contributions

After introducing some general background to security risk analysis and general knowledge reuse (Section 2) we propose a first theory of knowledge reuse for security risk analysis (Section 3). The theory is based on three key ideas: knowledge reuse can be re-used in three dimension: for finding information, for validating one's own findings, and finally for identifying a common terminology for communicating findings between analyses.

Building upon evidence gathered in the EMFASE project Labunets et al. (2014a), we report a qualitative study (interviews and focus groups) with 20 experts that shows that these three claims are supported (Section 4).

However, the deep analysis of the qualitative study also shows that not all types of re-users seek the same reuse. Therefore we propose a revised theory (Section 5) in which we explicitly distinguish between novice users of codified SRA knowledge (who mostly benefit for finding information and for communicating information) and expert users (who mostly use it to communicate information in a consistent form and to validate their ideas that nothing is forgotten). In this respect, the use of domain specific catalogues vs general catalogues would seem to be more favorable as it would diminish the level of intellectual demands by the reusers.

A further empirical analysis at first with 18 Msc students as representatives of novices and 15 industry experts in the Air Traffic Management Domain is the reported to validate the revised theory (Section 6). While the claim for novices is at least partly supported the claim for domain experts are not supported and more analysis is needed. We have thus found some evidence supporting hyp. 3 of Gray and Meister (2004) that the impact on knowledge sourcing will be stronger under conditions of higher intellectual demands.

Further qualitative analysis of the feedback by the industry experts particpating to the experiments helped us to come with a final refined theory that we propose for further studies (Section 7) and a discussion of possible implications for research and practice (Section 8).

The remaining sections discuss threats to validity (Section 9) and related works (Section 10), and conclude the paper (Section 11).

2 Background

In this section we provide an introduction to security risk assessment and the Markus' theory on knowledge reuse that are the building blocks of the proposed conceptual theory on knowledge reuse in security risk assessment.

2.1 Security Risk Assessment

Risk management is the means for an organization to define a security strategy that addresses the threats to which the organization is exposed. The cornerstone of risk management is security risk assessment (SRA) that allows to identify, analyze and evaluate security risks and select security controls that mitigate the risks to an acceptable level.

Several methods and standards for security risk assessment are available, for example EBIOS la Sécurité Des Systèmes D'information (ANSSI) (2019), MAGERIT Publicas (2012), OCTAVE Caralli et al. (2007), IT-Grundschutz BSI (2017), CORAS Fredriksen et al. (2002) and NIST 800-30 of Standards and Technology (2012) are some of the widely adopted methods. They differ in one or more of the following components: the *risk model* specifies the

factors to measure risks and the relationships among the factors; the risk *assessment approach* specifies the range of values those risk factors can assume and how they can be functionally combined to evaluate risk; the *analysis approach* describes how combinations of risk factors are identified/analyzed. However, the process to conduct a risk assessment is similar for all the methods and standards for security risk assessment and consists of the following phases: framing/scoping the process, identifying threats sources, vulnerabilities, and threat events, determining likelihood of occurrence of threat events, determine the impact of threat events, determine their risk level, communicate the risks, and maintain the assessment (Gritzalis et al. 2018). The process is conducted by a team which includes the risk analyst who guides the analysis and employees that have knowledge of the organization but may not have security knowledge like IT-staff who knows the network and infrastructure, and executives who have knowledge of the core business functions and processes. Each step of the risk assessment process requires the team members to identify relevant information for the assessment, being able to share this information with each other and validate the timeliness, the specificity, relevance and completeness of the identified information for the system under analysis. To complete these three key tasks, the team relies upon different knowledge bases that can be internal or external to the organization. Internal knowledge bases can include previous risk assessment reports, incident reports, and security logs. External knowledge bases may include cross-organizations and sector advisories on the latest threats and vulnerabilities like the one published by CISA Agency (2023), national CERTs, the SANS Institute SANS (2011), and CVE MITRE (2020a), catalogs of threats like the OWASP Top Ten for web applications OWASP (2021), the IT-Grundschutz catalog BSI (2017), MITRE Att&ck matrix of attack tactics and techniques (MITRE 2022), and catalogs of security controls like the SP NIST 800-53 of Standards and Technology (2020), the CIS Critical Security Controls for Internet Security (2023), the NIST Cyber Security Framework of Standards and Technologies (2023), and the UK NCSC's 10 Steps to Cyber Security Center (2021).

For example, during the threat identification step, the team members can start by identifying from a catalog of threats, generic threat categories and then identify concrete threat scenarios using the MITRE Enterprise Att&ck matrix, recent security advisories, and security incidents reports, then present the scenarios to the other members of the team and determine whether the threats are relevant for the organization.

In the next section we will introduce a theory of knowledge reuse that links the key tasks performed by a security risk assessment team - identifying information, sharing information and validating information - with the features that the knowledge base used by the team should have to produce effective risk assessment results.

2.2 Theory of Knowledge Reuse

The importance of knowledge and its management is well understood in traditional Information Systems (IS) from both theoretical and empirical perspectives. For example, the survey by Schultze and Leidner (2002) analyzed 94 knowledge management papers published between 1990 and 2000 in six IS journals and showed that knowledge in organizations is generally considered an asset.

Knowledge can be divided into *personal* and *community* knowledge (Wasko and Faraj 2000). Personal knowledge is tacit knowledge that people create by themselves or learn from their own experience. Based on personal knowledge people make decisions in their future projects. If people lack necessary knowledge they turn to community knowledge. Community knowledge is "personal knowledge" shared between members—for example, in

documented form (the catalogue being just one of such form). A theory of knowledge reuse by Markus (2001) suggests that the work of experts addressing problems in a new context can be facilitated by providing knowledge about proven solutions or best practices.

Schultze and Stabell (2004) extended the previous work and proposed a theoretical framework whereby knowledge can be considered as an asset that can be owned and transferred with the purpose of advancing both individuals and organizations. Garicano and Wu (2012) also argued that knowledge is a fundamental component for organizational processes, and that organization structure should be designed to facilitate knowledge communication between workers. This idea is also supported in Software Engineering by Rus and Lindvall (2002) and by Pilat and Kaindl (2011) who emphasized the importance of knowledge-sharing practice.

Boh (2008) proposed the research model explaining the factors affecting how individuals benefit from reuse of knowledge assets. which are "a key aspect of firm's intangible resources" as admitted by Bharadwaj (2000). The model shows that combination of electronic knowledge repositories and communication with the author of the knowledge asset can facilitate knowledge sharing process. Indeed, sharing knowledge and transferring best practice between teams and organizations is important for improving their performance. As shown by Raman and Bharadwaj (2010), the deviations from the standard process based on objective evidences (i.e. knowledge-based performative deviations) lead to positive outcomes from practice transfer in comparison to the agent-based performative deviations where the changes are made due to personal preferences of participants of the process. Also user motivation, perceived value of knowledge (Kankanhalli et al. 2011), and perceived cost of knowledge source adoption Boh (2014) are important aspects which impact the knowledge reuse. Recently, Arora et al. (2015) investigated knowledge sharing practice in alliances where competitors may be affected from knowledge misappropriation and asymmetric learning problems. The authors extended learning model of alliances with a knowledge sharing component and demonstrated its importance for alliance management and partner selection processes.

Raman and Bharadwaj (2010) empirically investigated the practice transfer routines in evidence-based practice transfer in healthcare. They found out that the deviations from the standard process based on objective evidences (i.e. knowledge-based performative deviations) lead to positive outcomes from practice transfer in comparison to the agent-based performative deviations where the changes are made due to personal preferences of participants of the process. Later, Kankanhalli et al. (2011) investigated how user motivation and perception of the value of knowledge repositories impact knowledge reuse and user's performance. Similarly, Boh (2014) studied how perceived cost of using different sources of knowledge (knowledge repositories and online discussion forums) affect knowledge sharing practice in professional communities. Arora et al. (2015) investigated knowledge sharing practice in alliances where competitors may be affected from knowledge misappropriation and asymmetric learning problems. This work incorporated a knowledge sharing component into a learning model of alliances and demonstrated its importance for alliance management and partner selection processes.

Another example of knowledge re-use is medical evidence-based practice (EBP) where "most forms of knowledge could be considered evidence […] the evidence used to guide practice should be '…subjected to historic or scientific evaluation'." Leach (2006). See Raman and Bharadwaj (2010) for a discussion. SRA is similar to EBP as a security analyst also needs to evaluate practical relevance of threats and effectiveness of security controls and can benefit from reusing security knowledge in form of catalogues.

3 Conceptual Theory of Knowledge Reuse in Security Risk Analysis

To investigate ***RQ1*** we observe the success factors for the key tasks of security risk analysis. Figure 1 shows the conceptual theory of knowledge reuse in security risk analysis. Based on the theory by Markus (2001) we assume that security analysts leverage both personal knowledge and community knowledge. Second, we assume three key tasks that are at the core of security risk analysis. Finally, we postulate there exist certain characteristics (or features) of the catalog that have a measurable effect on the success of a security risk analysis outcomes. Following the existing literature on SRA (see related work in Section 10), we consider the quality of analysis outcomes as measurable by *actual and perceived efficacy* of threats and security controls produced by the subjects. To measure the actual efficacy, independent experts from industry can assess the quality of threats and security controls (as 'Bad', 'Poof', 'Fair', 'Good', and 'Excellent') by following prescribed guidelines. *Perceived efficacy* is instead typically measured with a post-task questionnaire including 5-point Likert scale questions (see Section 6.2 for details). In the following we present our hypotheses regarding the key success factors for the core tasks of SRA.

Finding Information Finding the right information during SRA is very challenging, especially in multinational organization operating globally, with extremely complex systems (Tuma and Widman 2021). Analysts must at first find relevant information about how system component actually work (see domain expertise in Fig. 2), which are the actual security threats (security expertise is needed), and which is the risk (likelihood and impact) of the threats. Then, to mitigate the highest risks, appropriate security controls have to be selected (specific controls). Therefore, finding information is a core task of the security risk analysis process because it is essential for activities that go beyond the identification of specific threats and controls.

*H*1*a: A key factor for the success of security knowledge reuse, in particular finding information, is a knowledge representation that facilitate the retrieval of information.*

Presenting Information Organizations perform risk analysis in teams (of different experts) for particular subsystems or components of the global product. For example, in the automotive sector, a risk analysis session may investigate a single Electronic Computing Unit (ECU) type. The outcomes of such analyses must be communicated to the expert teams assessing the interacting subsystems (because the in-vehicle architecture is composed of several hundred ECUs). Therefore, comparability of the results *between analyses* is surely an important factor of success. Within large organizations it is very difficult to establish common terminology, which can be a time-waster during an SRA. For instance, a previous study has

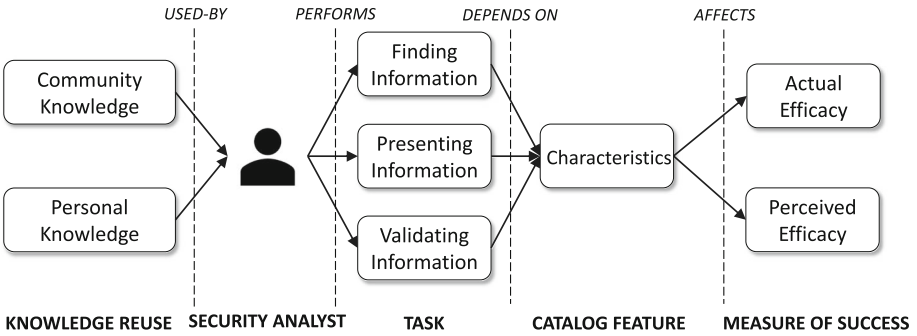


Fig. 1 Conceptual theory evaluated with interview study

found that terminology and threat feasibility discussions were actually detouring the analysis process, making the entire analysis session less productive (Tuma et al. 2021).

*H*1*b: A key factor for the success of security knowledge reuse, in particular presenting information, is the use of domain specific terminology.*

Validating Information Compliance to standards is in many organizations the driving force for performing security risk assessment in the first place. Therefore, comparability of the results *across analyses*, but also with respect to the standards the domain is subject to comply with, is surely an important factor of success. However, existing SRA techniques do not necessarily produce the same type of outcomes (e.g., consider the outcomes of STRIDE (Cruzes et al. 2018; Karahasanovic et al. 2017), which are a list of security threats expressed with natural language vs CORAS Fredriksen et al. (2002), which include specialized mitigation diagrams). Neither do they require the same type of information as input to the analysis procedure Tuma et al. (2018). At the very least, the outcomes must be shared in a standardized form (e.g., the same type (and amount) of outcomes). For compliance to standards, organizations need to care about systematically performing all the required analysis steps, as well as, the completeness of the analysis. Analysts need to make sure that all the security threats have been considered, which is more difficult to assure when using a short list of threats. Possibly, for companies that need to comply to standards, a knowledge base that is more comprehensive could provide more certainty that the experts have not missed a particular threat.

*H*1*c: A key factor for the success of security knowledge reuse, in particular validating information, is inclusion of comprehensive amount of content in the knowledge representation.*

The success factors (i.e., knowledge representation, use of domain specific terminology, and type and amount of information) are not mutually exclusive. In fact, it is desirable to report SRA outcomes that are both organized into a standardized form (e.g., with a specified csv), include a comprehensive amount of threats and their mitigations, *and* are described using domain specific language.

4 Qualitative Study: Success Factors for SRA

This section presents the semi-structured interview study conducted with 20 practitioners. The goal of this qualitative study is to gather evidence about success factors (i.e., testing hypotheses $H1a - H1c$ and answering ***RQ1***) of re-using knowledge in terms of structured methodologies and catalogues for security risk assessment.

4.1 Study Design

This study has been conducted as a part of SESAR EMFASE research project. SESAR (Single European Sky Air Traffic Management Research) is a public–private partnership which includes a total of 70 organizations. SESAR coordinates and concentrates all EU industrial R&D activities for future Air Traffic Management (ATM), including the development of its operational concepts (estimated at €2.1 billion). We used purposive expert sampling and got access to experts with relevant expertise in security risk assessment by obtaining permission to conduct our study as a part of one of a SESAR Working Group regular security meetings.

The format and content of the interviews was discussed and brainstormed with senior researchers from SINTEF, and in particular the group in charge of CORAS, a security risk

assessment methodology (Lund et al. 2010). The group had an extensive experience on professional risk analysis. The questions were also discussed with an Eurocontrol expert on security risk assessment and the Eurocontrol chair of the working group where the focus groups would be held.

The format of small focus groups and semi-structured interviews was chosen to allow flexibility in topics discussed and capture the success factors more precisely (Yin 2010). It also allowed to limit the time needed by interviewees and the number of interviewers who had to travel to the meeting.

4.2 Participants and Interviews Collection

We interviewed 20 ATM experts from the working group in charge of the security analysis (for all developed solutions within the SESAR Joint Undertaking program) and who participated in the meeting. The subjects were industry professionals with on average 17.5 years of working experience and, in particular, 7 years of experience in risk assessment. Almost all experts worked in industry, but for one who worked in academia and one who worked in a research center. As current occupation they were security experts (40%), security consultants (30%), system engineers (15%) and researchers (10%), from different national and international organizations and companies involved in the ATM domain at different levels. For confidentiality reasons, participants or their companies can not be disclosed.

We collected the data by holding four parallel focus group sessions with 5 subjects (FG1, FG2, FG3, and FG4) lasting approximately 30 minutes each. Two interview groups were held by two authors researchers, one by a anthropologist contracted by the project, and one by an industry expert from SESAR who who had previous experience with qualitative studies.

4.3 Data Analysis with Coding

We adopt a combination of *open and selective coding* as a method of qualitative analysis of the interview transcriptions. Coding is a qualitative technique for marking chunks of the transcribed text with short and concise descriptions of the key expressed points. The first **coding hierarchy** was iteratively developed and discussed with SINTEF following the best practices of conducting qualitative analysis of data gathered from interviews (Labunets et al. 2014a, b; Tuma et al. 2021). It consisted of 3 categories and 15 codes[1]. Additional codes emerged to avoid constraining the coding process and limiting the observations. The final set of codes includes 27 codes of success criteria for SRA methodology (they are all listed on the left side of Fig. 2).

To better understand the emerging issues, we also coded the particular *task code* (see the right side of Fig. 2) the transcription statements were referring to (i.e., *selective coding*). Particularly, we coded three tasks: finding, presenting/sharing, and validating information.

The coding was performed with *Atlas.ti*[2] by the anthropologist contracted by the project and the final results was checked by the two researchers who participated to the interviews. Table 1 illustrates an example of coding.

[1] Well-defined process, Time effective, Holistic process, Compliant with ISO/IEC standards, Visualization, Comprehensibility, Evolution support, Well-defined terminology, Specific controls, Coverage, Comparability, Catalogs, Documentation templates, Practical guidelines, Tool support

[2] https://atlasti.com

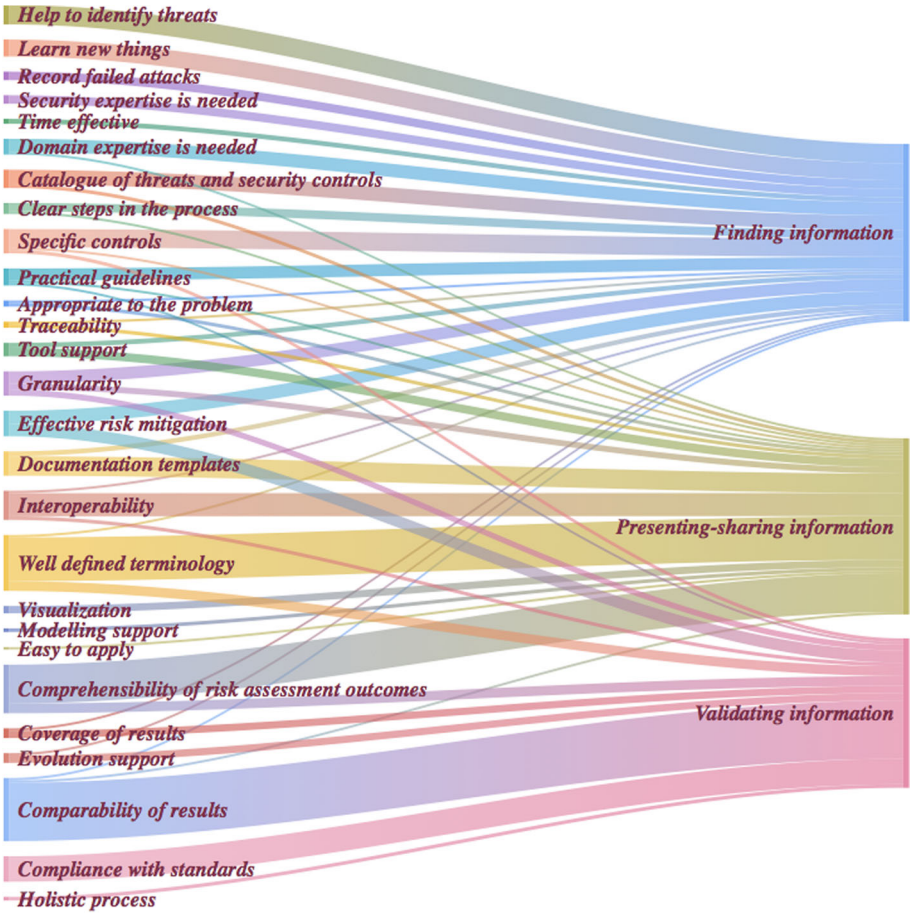


Fig. 2 Co-occurrence between codes for success criteria and tasks

As a proxy for salience Guest et al. (2011), in addition to estimating the relevance of each criterion in terms of frequency in the interviews, we also calculated the frequency of their co-occurrence in the same statement.

Figure 2 shows the co-occurrence between the success criteria and the tasks coded in the transcriptions.

Table 1 Sample of coding choices

Transcription Snippet	Applicable Codes
"[The methodology] has to support the risk analysts in achieving the results they want, of course. So either identification of threats or estimation of likelihood or identification of security controls or whatever…"	**Success Factor Code**: Help to identifying threats; **Task Code**: Finding Information

4.4 Results: Answering RQ1

Finding Information Our analysis shows that finding information is supported by the highest number of different success criteria identified by the subjects. In particular, the methodology fulfills its own purpose when it allows to acquire some knowledge previously unknown (i.e., learn new things), to quote one of our participant:

> *"The best methodology leads to [...] [a] specific solution that is not covered by best practice".* (FG4; St. 1)

> *"The methodology should have a comprehensive threats catalog so people may start from a base catalog and then eventually add other threats."* (FG1; St. 1)

Overall, we observe the complexity of the task of finding information is larger when compared to the task of presenting-sharing or validating information. However, related work does not measure the effect of leveraging structured catalogs on the analysis (perceived and actual) outcomes. Therefore, we identify a need for more empirical evidence about the effect of having **structured catalog** on the efficacy of SRA.

Our participants mentioned the need for clear steps in the process and practical guidelines for a successful SRA, but the structure of knowledge representation was not explicitly mentioned. Thus, the presented qualitative evidence partially supports $H1_a$.

Presenting-Sharing Information The most relevant aspect considered by the participants is to have a well-defined terminology as this enhances the interoperability among experts and stakeholders. This was nicely summarized by one of our participants in the following statement:

> *"Where I see the shortfalls, is the place where people […] don't understand each other, […] Some common training would tell them that this is what you are talking about altogether"* (FG1; St. 3)

In addition to a well-defined terminology, documentation templates were also perceived as beneficial by our participants. Similarly, a well-defined terminology was considered to improve the comprehensibility of the outcomes. One of our participants reported on the matter as follows:

> *"There is no sense if you have a super method, if the results cannot be exchanged [...]. You have to share the situation"* (FG2; St. 1).

Presenting and sharing information is a challenge in SRA due to a lack of a well defined terminology, interoperability of outcomes, and comprehensibility of risk assessment outcomes. The latter challenge was explored in the some studies (Labunets et al. 2017b, a), but, the existing literature (summarised in Section 10) does not measure the effect of terminology specificity on the analysis (perceived and actual) outcomes. Therefore, there is a need to study the effect of **domain-specific terminology** definitions in the catalog on the efficacy of the SRA.

Our interviewees perceived well defined terminology as a key success factor (as seen from Figure 2). Therefore, the qualitative evidence gathered supports our hypothesis $H1b$.

Validating Information Compliance is a particularly important aspect for professionals, and typically not a concern for novice analysts or trainees (e.g., students). But, the ease with which information is validated **may also depend on the content** (and its amount) of the knowledge base.

Our analysis suggests that for validating information the key success factors are comparability of results and compliance to standards. Consistent use of knowledge representations with comprehensive content helps in obtaining fine-grained evidence which is often required for compliance analysis. Therefore, the results presented in Figure 2 support $H1c$.

5 Revised Theory of Knowledge Reuse in Security Risk Analysis

During the discussions with experts, more senior experts typically referred to catalogues and tool-supported checklists as way to validate one's own intuition, as follows:

> *Participant A: "In my opinion, the catalog should be used after the brainstorming, to verify, that all controls have been identified that are present in the system. I see it as an instrument to be used ex-post, more than for the definition of controls." …Participant B: "I agree on the catalog. It looks more useful when used in retrospect, to check what you thought."*

At the same time, a recent systematic study (Jafari and Rasoolzadegan 2020) has found that the structure of security patterns is designed to help teach security knowledge from experts to novices. Novice analysts are therefore significant consumers of community knowledge in which they actually seek new knowledge and such use seems significantly different from the use advocated by the experts who participated to the interviews.

Hence, we revised our theory by considering two type of SRA knowledge re-users: novice in SRA and experts in SRA. Figure 3 shows the revised theory of knowledge reuse in security risk analysis.

In our revised theory, most SRA senior experts mainly use catalogs to either present (share) information, by using the same terminology or to validate information. On the other hand, novice analysts may rely more on the catalogs to actually find information. Therefore, for novice users a clear catalog structure with a comprehensive amount of content without irrelevant information may affect both the actual and perceived efficacy of catalogs.

A another key observation from the discussion was that *a novice in security risk analysis is* not *necessarily a new employee, s/he might be a domain expert with several years of experience who has not been previously engaged in security risk assessment*. For example in SESAR one might find several safety experts who never performed a security risk assessment. Indeed, two participants explicitly stated that

Participant C (10yrs IT security experience): "In SESAR, it has to be usable by non-security experts (with support from (removed))"[…] participant D (35yrs experience): "The methodology is clear to those who need to manage systems, operations or regulate operations."

Such distinction is therefore important to revise our experimental hypothesis $H2a, H2b, H3$ that were formulated to investigate the research questions. The interviews have thus identified classes of possible subjects of interest

1. novices in both domain *and* security risk analysis
2. novices in domain *but* experts in security risk analysis (on a different domain)
3. experts in domain *but* novices in security risk analysis
4. experts in both domain *and* security risk analysis

We did not consider class two in our analysis as from our interviews was apparent that knowledge of domain is important to be able to generate applicable recommendations.

Participant E (FG1): "It is the users of the methods that possess the information. The users needs help from the method to organize the information."

So essentially members in class two would need to work in team with domain experts and thus would, as a team, join class four.

From novices perspectives, they use catalogs mostly to find information and adopt the appropriate language to present results. First, if a catalog does not have clear and logical *structure* it can affect novices' perceived efficacy and increase the effort needed to find the necessary information. Second, novice analysts can struggle with catalogs that are too big due to feeling of being overwhelmed by options. Hence, the *amount of information* presented in a catalog can affect both the actual and perceived efficacy of a security assessment. To investigate the structure and amount of information we propose the following alternative hypothesis:

*$H2a$: Novices using **specific** catalogs will have **better actual and perceived efficacy** than the novices who used catalogs with broad and extensive structure and large amount of content.*

Experts rely on their own knowledge as a source of information. However, their knowledge might be incomplete in a particular direction (e.g. safety expert but not security expert). They may benefit from the use of community knowledge in the same way that absolute novices do, to navigate the correctness and coverage of their intuition and to check the terminology used to present results to the particular customers. We conjecture that using catalogs with terminology that they are already familiar with can increase completeness of the results and have a positive effect on their actual and perceived efficacy.

Supplying a catalogue to experts in both security and domain would definitely make their work easier. As a participant of the focus group noted, they could quickly check that nothing was forgotten. However, they would not face any particular intellectual demand on their task. Thus, an intellectual demanding condition would for them to rely on their own knowledge only.

To investigate this research question we propose the following alternative hypothesis:

*$H2b$: **Domain and security experts** using catalogs will have **better actual and perceived efficacy** than those who did not use catalogues.*

As emerged from the interviews, security expertise (including knowledge of state of the art security methods, techniques and how to apply them, knowledge of the security attack landscape, security mitigations, effort in implementing countermeasures and monitoring for attacks, etc.) is valuable but scarce and time to perform the analysis is even scarcer.

At this point, to answer RQ3 we want to compare experts under the strain of intellectual demands.

From the perspective of domain experts who lack security expertise we have the alternative of supplying a general catalogue and a domain specific catalogue to provide them with different level of intellectual demands.

However, *H*2*b* already identifies that catalogues should make the work more effective. At this point the result that participants C and D would like is that security knowledge bases could be (to some extent) used by non security experts to do a "good enough" job when the actual security experts are not available.

*H*3*: Domain experts (with no security knowledge) using catalogs will have* ***actual efficacy similar*** *to security experts who conducted security analysis* ***without using catalogs***.

Evidence of (or against) *H*3 does not diminish the need of security experts. Rather than aiming to replace security experts, it only stipulate the presence of satisficing solutions.

Having said that, security knowledge base requires security experts to exist in the first pace, as it has to be updated, and communicated/transferred. The gap in security workforce still exists, the question is what can be done *despite* this shortage, and is it worth it.

6 Experimental Study

We designed an experiment and conducted it with experts and student participants to evaluate the revised theory of knowledge reuse in SRA (see Fig. 3).

6.1 Experiment Design

The experiment was executed in four phases.

Demographics Questionnaire First, subjects were given a questionnaire to fill-in about their previous knowledge and experience (e.g., with risk assessment). We used this questionnaire to collect demographic information about our participants.

Training A scenario description was administered to subjects by either an individual reading or by an introductory presentation. Then, a pre-training phase follows in which the

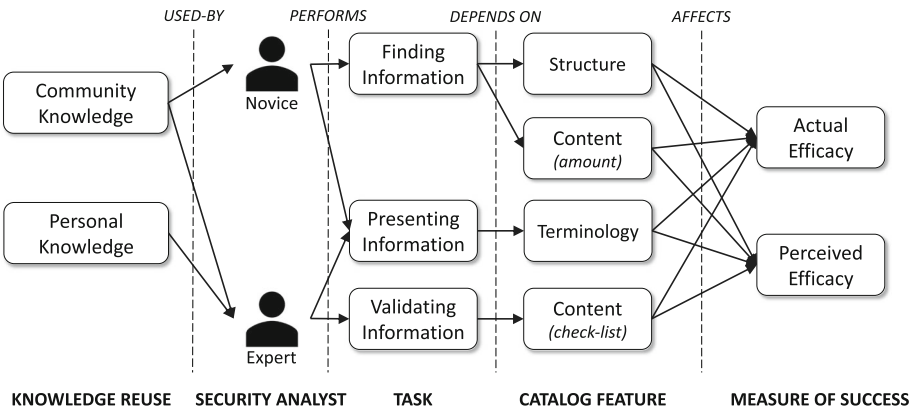


Fig. 3 Revised theory after interview study

expert in the method introduces the methodology to be used for the SRA through a step-by-step tutorial.

Analysis Sessions On the day of the experiment the participants were handed the relevant material from the training, the documentation of the scenario and the catalog (according to the treatment). The procedure of the experiment was repeated by the researchers. Next the participants could start applying the method to the scenario. During the analysis sessions the experimenters acted as observers and intervened only in case of procedural questions to avoid influencing the outcomes of the study.

Post-Experiment Qualitative Feedback A post-experiment questionnaire was handed out to the subjects to gather their perception of the method (and the catalog). Then they are organized into focus groups to discuss the drawbacks and benefits of the method and the catalogs they have used. A list of questions was used to guide the discussion, which was audio recorded for further analysis. The main positive and negative aspects reported in the focus groups were then recorded on post-it notes by the subjects.

6.2 Measures

In this work we measured *actual efficacy* as the *quality* of results produced by the subjects (in contrast to the related work (Opdahl and Sindre 2009; Scandariato et al. 2014; Karpati et al. 2014; Stålhane and Sindre 2014; Tuma and Scandariato 2018)). Using the number of threats and security controls as a performance metric would be meaningless in our experimental setup because there are many threats and security controls available in the catalogs, and subjects could include any of them in the analysis. Yet, they may be irrelevant. were sometimes made by industry assessors when evaluating the threats reported by students in previous studies

To assess *actual efficacy* we rely on expert-based assessment of the artefacts produced by the subject (as reported in (Opdahl and Sindre 2009; Massacci and Paci 2012) and Labunets et al. (2013)). These expert assessors were *not* the authors of the paper nor their colleagues. They were independent industry experts contracted for the task. To mitigate potential bias, we prepared assessment guidelines and shared them with the hired experts. In a nutshell, we asked each expert to rate the overall results on a 1–5 scale as follows:

- (1) *Bad*, when it was not clear which are the final threats or security controls for the scenario;
- (2) *Poor*, when threats/security controls were not specific for the scenario;
- (3) *Fair*, when *some* of them were related to the scenario;
- (4) *Good*, threats/security controls were specific for the scenario; and
- (5) *Excellent*, when the threats were significant for the scenario and the security controls propose an effective solution for the scenario.

To assess *perceived efficacy* we used both a quantitative and qualitative measure. We first asked the subjects to fill in a post-task questionnaire. The questionnaire contains 10–20 questions about different constructs specific to perceived usefulness (PU) and perceived ease of use (PEOU) variables (Davis (1989)). This approach has also been applied to measure perceived efficacy of security and safety methods in numerous studies (Opdahl and Sindre 2009; Stålhane and Sindre 2014; Wuyts et al. 2014; Mouaffo et al. 2014; Karpati et al. 2015) and validated in our previous experiments (Labunets et al. 2013, 2014b). Questions were formulated as opposite statements with answers on a 5-point Likert scale.

Data Analysis We used both difference and equivalence statistical tests. As our data are ordinal and comes from independent samples, we select the Mann-Whitney test. For the equivalence test we use TOST, which was initially proposed by Schuirmann (1981) and

Table 2 Descriptive statistics of the sample

Function	Min	Max	Mean	Median	σ
Mean of samples	1.48	4.05	2.92	2.96	0.68
σ of samples	0.35	1.28	0.86	0.835	0.25

is widely used in pharmacological and food sciences to answer the question whether two treatments are equivalent within a particular range δ (Food and Drug Administration 2001; Meyners 2012). We defined $\delta = 0.7$ empirically, corresponding to the pooled variance σ_p across 36 samples reported in the literature on knowledge reuse (ManSci, ISR, MISQ in a four year interval) on variables ranging over a 5-item Likert scale for demographic statistics as to account for natural variability of the data. From the identified papers we extracted descriptive statistics of ordinal variables for 36 samples. Table 2 reports descriptive statistics of variables means and standard deviations across collected samples from the articles in the literature from which we calculated the pooled variance.[3] The detailed papers are reported in Tab. 1 in Appendix.

6.3 SRA Method and Scenario

Usually, an SRA process needs to identify three main components: (1) assets that should be protected, (2) threats that can harm identified assets, and (3) security controls that can mitigate identified threats. To provide a common baseline to all participants, the experimenters have conducted an asset analysis beforehand and provided its results as part of the case documentation to all the teams. The task of the participants was to identify the security threats (2) and find appropriate security controls (3) (with or without catalogs, depending the treatment group).

Remotely Operated Tower (ROT) The application scenario was chosen among one of the ATM new operational scenarios that have already been assessed by SESAR with the SecRAM methodology: the Remotely Operated Tower (ROT). The Remote and Virtual Tower is a new operational concept proposed by SESAR.[4] The main change with respect to current operations is that control tower operators will no longer be located at the aerodrome. They will move to a Remotely Operated Tower Center. Each tower module will be remotely connected to (at least) one airport and consist of one or several Controller Working Positions. The operator will be able to do all air traffic management tasks (e.g. authorize landing, departure, etc.) from this position. The idea is that operator will be able to control remotely more than one airport. Visual surveillance by the air traffic controller will be replaced by a virtual reproduction of the Out of The Window view, by using visual information capture and/or other sensors such as cameras with a 360° view and overlaid with information from additional sources such as surface movement radar and surveillance radar. The first implementation of the ROT was in Sweden by LFV and Saab.[5]

[3] To calculate that pooled variance σ_p we used $\sigma_p = \sqrt{\sum_{i=1}^{k} \sigma_i^2 (N_i - 1) / \sum_{i=1}^{k} (N_i - 1)}$, where N_i is the size and σ_i is the variance of sample i. Using this formula on the collected dataset of 36 papers we chose $\delta = 0.7$ for the test.

[4] SESAR Project P12.04.07: Single Remote Tower, Technical Specification, Remotely Operated Tower Multiple Controlled Airports with Integrated Working Position.

[5] "LFV first in the world to have an operating license for remote towers" (http://news.cision.com/lfv/r/lfv-first-in-the-world-to-have-an-operating-licence-for-remote-towers,c9672916).

As is apparent from the description, the ROT concept is a complex cyber-physical information system encompassing both cyber-security issues (e.g., data confidentiality, integrity and availability of sensor data) as well as physical security issues, like on-site protection of the remotely located cameras and sensors.

Security Risk Assessment Method (SecRAM) is a method used in the Air Traffic Management (ATM) domain to conduct security analysis of operation concepts. It is highly aligned for compliance with steps resembling security standards (e.g., NIST 800-30). Crucially for this experiment, the method is supported by the use of (various) catalogs of threats and controls. Finally, previous results of using SecRAM for analyzing the same scenario were available to the experimenters for inspection and the authors of the method were available to train the participants. Depending on their treatment group, participants were handed (or not) specific catalog.

Catalogs The main characteristics of the two catalogue types are summarized in Table 3.

We used two catalogs to support the analysis with SecRAM in this experiment. To test our research hypotheses regarding the effects of a domain specific catalog, we selected the ATM domain catalog (DOM CAT) developed by the European Organization for the Safety of Air Navigation. This catalog matches well the needs of security analysts in assessing the risks for the ATM specific scenarios. As a representation of a general catalog (GEN CAT), we selected the BSI IT-Grundschutz catalogs developed by the German Federal Office for Information Security.

DOM CAT have clear and simple structure (32 threats divided into three topics with links to security controls), reasonable size (155 pages), support users with ATM-specific terminology, and propose effective controls. Table 8 provides an illustrative example of a threat catalogue entry describing the social engineering threat.

In comparison GEN CAT has a large corpus (621 threats and 1444 security controls divided in six topics with links between threats and controls in a separate section), large size ($\approx$ 2500 pages), supporting users with common security terminology, and covering a wide range of IT security problems and solutions. Figure 6 shows an example of a threat catalogue item describing the social engineering threat.

Table 3 Catalogues' main characteristics

Topic	DOM CAT	GEN CAT
Structure	Simple	Complex
Amount of content	Reasonable size	Very Large
Terminology	Domain-specific	Common Security
Threats	32 (Physical, Information and Procedural)	621 (Basic threats, Force Majeure, Organizational Shortcomings, Human Error, Technical Failure and Deliberate Acts)
Security controls	51 (Pre- and Post-controls)	1444 (Infrastructure, Organization, Personnel, Hardware and software, Communication and Contingency planning)
Link between threats and security controls	Yes (two-way); as a part of threats or security controls description	Yes (two-way); in a separate section

Task Difficulty In general, we aimed to make our experimental materials comparable to realistic cases and, therefore, worked in collaboration with the scenario owners and experts. Moreover, the participants of both experiments (professionals and students) did not have any prior knowledge of the SecRAM security method. Therefore, the task of performing the analysis using SecRAM was of comparable difficulty for all our participants.

6.4 Participants Recruitment and Demographics

Students First we conducted the experiment with 18 MSc students from different European universities in February 2014. Our participants were recruited through the European Institute of Innovation & Technology (EIT) network by offering a possibility to participate in the Winter School on Secure Design. The event was open for the master or last year's bachelor students with the background in security, privacy or related computer science areas. The subjects worked in groups of two. A significant share of the subjects (44%) reported a working experience of at least 3 years, some subjects (22%) reported $\leq$2 years of previous working experience, and the rest reported no previous working experience. Some student subjects (28%) reported previous involvement in security/privacy initiatives. They had limited expertise in safety and security regulations, while in security technologies they reported a general knowledge. Finally, student subjects also had no prior knowledge of the ATM domain.

Subjects were divided into two treatment groups: the first conducted an SRA with the support of a domain-specific catalog (DOM CAT-S), the second group with the support of a general (GEN CAT-S) one. Nine teams were randomly assigned to the catalogs: five teams applied the SESAR SecRAM method to the ROT scenario using EUROCONTROL ATM catalogs (DOM CAT-S), while the other four teams used BSI IT-Grundschutz (GEN CAT-S).

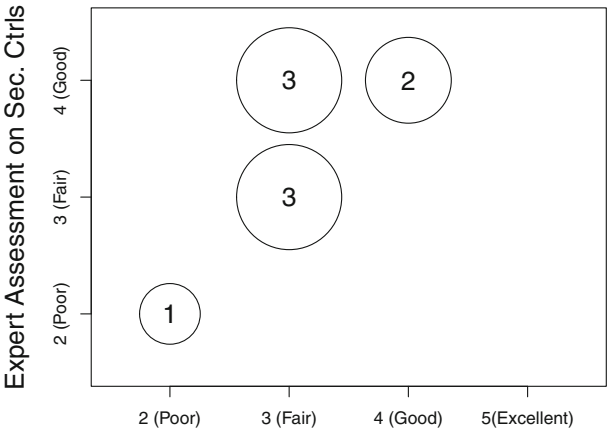
Industrial Experts Second we conducted the experiment with 15 industrial experts working at several Italian ATM companies. We recruited participants through SESAR network by inviting them to participate in a free training on SecRAM method by qualified experts. Most of the subjects (73.4%) reported that they had at least 5 years of working experience, some subjects (26.7%) reported 2–5 years of working experience. In addition, the majority of subjects (60%) reported that they had security/privacy knowledge; the rest reported no knowledge. Three out of 16 subjects reported from 3 months up to 2 years experience in SRA.

Expert subjects were divided into *three* treatment groups: the first two were analogous to the student treatments (DOM CAT, GEN CAT) and a third treatment which worked without a catalog (NO CAT). The nine subjects with security knowledge where split into a group without catalogues (NO CAT+SEC, five participants)) and the reminder four with a catalogue. The ten subjects assigned to a catalogue were then evenly split between the DOM CAT and GEN CAT catalogues. Again, we then asked them to apply individually the same method, namely SESAR SecRAM.

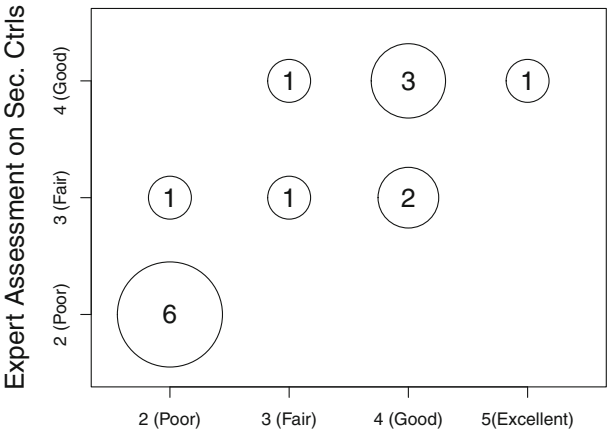
6.5 Results: Answering RQ2 and RQ3

Expert Assessments of Actual Efficacy

Figure 4a illustrates the average of experts' evaluation for threats (reported on the x-axis) and security controls (on the y-axis) for the student groups. Only one group out of nine performed poorly. Most groups (2/3) performed fairly good (or good) and there was no group that excelled in finding the threats.



(a) Student participant groups.



(b) Expert participant groups.

Fig. 4 Expert assessments of quality of threats and security controls (none of the security control results were assessed as excellent)

Figure 4b illustrates the average of experts' evaluation for threats (reported on the x-axis) and security controls (on the y-axis) for the expert participants who worked on the task alone. Six subjects out of 15 performed poorly. The rest mostly performed good (or fairly good) and one excelled in the reported threats. A second investigation of the expert assessments revealed that the expectations of the quality of results submitted by experts were higher (compared to student groups). Still, in cases of slight disagreements of the overall mark (e.g., "bad" and "fair") the assessor comments actually revealed an agreement on the quality. Hence, for our quantitative evaluation we used the average of the vote of the experts.

Table 4 summarises the quantitative results of the statistical analysis conducted to answer $RQ2$. Our analysis supports the $H2a$ hypothesis with respect to the perceived usefulness (PU), but not for actual efficacy and perceived ease of use (AE, PEOU). There is **no difference**

Table 4 Summary results ($H2a$: Novices)

	DOM CAT				GEN CAT				Statistical Tests		Eff. size
Threats											
	N	μ	med	σ	N	μ	med	σ	$TOST$	MW	d
AE	5	3.33	3.33	0.63	4	3.08	3.00	0.17	0.06	0.24	0.51
PU	10	3.64	3.57	0.46	8	3.07	3.07	0.42	0.37	0.02	1.30
PEOU	10	3.50	3.70	0.78	8	3.58	3.60	0.63	0.04	1.00	-0.10
Sec. Ctrls											
	N	μ	med	σ	μ	N	med	σ	$TOST$	MW	d
AE	5	3.27	3.33	0.64	4	3.58	3.67	0.17	0.09	0.50	-0.64
PU	10	3.71	3.71	0.41	8	3.21	3.29	0.34	0.22	0.01	1.31
PEOU	10	3.56	3.60	0.67	8	3.60	3.60	0.53	0.03	0.94	-0.07

in actual efficacy (AE) and perceived ease of use (PEOU) of the two catalogs for novice subjects without domain expertise. The results of TOST returned $p = 0.04$ for PEOU of threats and $p = 0.03$ for PEOU of controls. For actual efficacy (AE), we only have a TOST for equivalence at 10% confidence level.

However, we have found a difference in the perceived usefulness (PU). Namely, novice subjects without domain expertise perceive that domain-specific catalogs **are more useful** than general ones (MW test with $p = 0.02$ for threats and $p = 0.01$ for security controls with a large effect size).

> In other words, (absolute) novices are not able to exploit domain specific knowledge reuse to a satisfactory extent. They also find catalogues equally cumbersome to use. However, they *perceived* the domain specific catalogue as more effective. We only partially support $H2a$.

Table 5 summarises the quantitative results of the statistical analysis conducted to answer the second part of $RQ2$.

Table 5 Summary results ($H2b$: Security Experts)

	CAT				NO CAT				Statistical Tests		Eff. size
	N	μ	med	σ	N	μ	med	σ	$TOST$	MW	d
Threats											
AE	4	4.12	4.00	0.63	5	2.80	2.50	0.45	0.95	0.02	2.49
PU	4	3.64	3.64	0.44	5	3.77	4.00	0.46	0.09	0.75	-0.28
PEOU	4	3.50	3.60	0.53	5	3.64	3.80	0.52	0.10	0.69	-0.27
Sec. Ctrls											
AE	4	3.75	3.75	0.29	5	2.80	3.00	0.57	0.72	0.03	2.02
PU	4	3.68	3.50	0.41	5	3.77	3.71	0.41	0.10	0.69	-0.23
PEOU	4	3.30	3.40	0.66	5	3.64	3.80	0.52	0.19	0.61	-0.58

Our analysis supports the $H2b$ hypothesis with respect to the actual efficacy (AE), but not for perceived efficacy (PU, PEOU). Security experts who used catalogs have **better actual efficacy** than security experts who performed analysis without catalogs (MW test returned $p = 0.02$ for threats and $p = 0.03$ for controls). For PU of threats and controls and for PEOU of threats both groups reported comparable results. For PEOU of controls the results neither are equivalent nor different.

Table 6 summarises the quantitative results of the statistical analysis conducted to answer $RQ3$. The results are inconclusive when comparing actual efficacy of a security analysis conducted by non-security experts with catalogs and security experts without a catalog. Neither equivalence nor difference tests showed statistically significant results. Possibly, using a catalog may improve the actual efficacy of domain experts slightly, but not enough compared to the efficacy of security experts without a catalog. We note that for security controls security expert without catalogs reported better PEOU than non-security experts with catalogs (however this is not a significant result $p = 0.07$). We would need a bigger experimental sample to answer $RQ3$.

7 Refined Theory of Knowledge Reuse in Security Risk Analysis

Table 7 presents a summary of results. For every research question the table lists the investigated hypotheses, result of testing the hypothesis (✓ fully supported, (✓) partially supported, ✗ not supported), and the type of evidence supporting the result.

We base our theory refinement on the results presented in Section 6 and short post-experiment post-it sessions. Figure 5 shows the refined theory of knowledge reuse in security risk analysis. Compared to Fig. 3 the refinement differs in three effects. First we have not found evidence of one effect (crossed out arrow from terminology to PE in Fig. 5). Second, we found mixed evidence for the effect of content (as check-list) on AE and PE (dashed arrows in Fig. 5). We provide a more in depth discussion for each catalog feature.

Catalog Structure Thanks to its basic layout, clear tables (see threat entry example in Table 8) and its relatively short length (155 pages), the domain-specific catalog is generally perceived by the subjects as easier to browse and to read.

Table 6 Summary results ($H3$: domain vs security experts)

	Dom. Expert CAT				Sec. Expert. NO CAT				Statistical Tests p-value		Effect Size
	N	μ	med	σ	N	μ	med	σ	$TOST$	MW	d
Threats											
AE	6	2.50	2.50	0.71	5	2.80	2.50	0.45	0.15	0.50	-0.50
PU	6	3.33	3.50	0.66	5	3.77	4.00	0.46	0.25	0.23	-0.75
PEOU	6	3.20	3.30	0.59	5	3.64	3.80	0.52	0.31	0.36	-0.78
Sec. Ctrls											
AE	6	2.50	2.50	0.45	5	2.80	3.00	0.57	0.12	0.40	-0.59
PU	6	3.31	3.43	0.61	5	3.77	3.71	0.41	0.26	0.21	-0.87
PEOU	6	3.00	2.90	0.55	5	3.64	3.80	0.52	0.52	0.07	-1.19

Table 7 Summary of answers to $RQ1$, $RQ2$, and $RQ3$

RQ	Hypothesis, Result, and Evidence
$RQ1$	$H1_a$: *Knowledge representation is a key success factor.*
	(✓) supported by qualitative evidence (Section 4)
	$H1_b$: *Domain specific terminology is a key success factor.*
	✓ supported by qualitative evidence (Section 4)
	$H1_c$: *Comprehensive amount of information is a key success factor.*
	✓ supported by qualitative evidence (Section 4)
$RQ2$	$H2a$: *Novices* + DOM CAT > *Novices* + GEN CAT.
	(✓) for PU; supported by quantitative evidence (Section 6)
	$H2b$: *Sec experts* + CAT > *Sec experts* + NO CAT.
	(✓) for AE; supported by quantitative evidence (Section 6)
$RQ3$	$H3$: *Dom experts* + CAT = *Sec experts* + NO CAT.
	✗ not enough evidence found (Section 6)

> *"I read only the titles [namely the reference to the "Generic Threat" and the "Attack Threat"], they were quite explanatory, therefore a very short consultation of the catalog allowed me to produce enough content."* (DOM CAT subject)

In contrast the general catalog consists of a long list of items, which was perceived as "not user-friendly at a first read" (GEN CAT subject) and "difficult to navigate and master due to its length and structure" (GEN CAT subject).

Second, the link between threats and security controls was appreciated by our participants (even more so, novices):

> *"[Having] identified the threat, finding the controls was really a mechanical task."* (DOM CAT subject)

In contrast, the general catalog does not provide this support (see example in Fig. 6) and therefore the findings are affected:

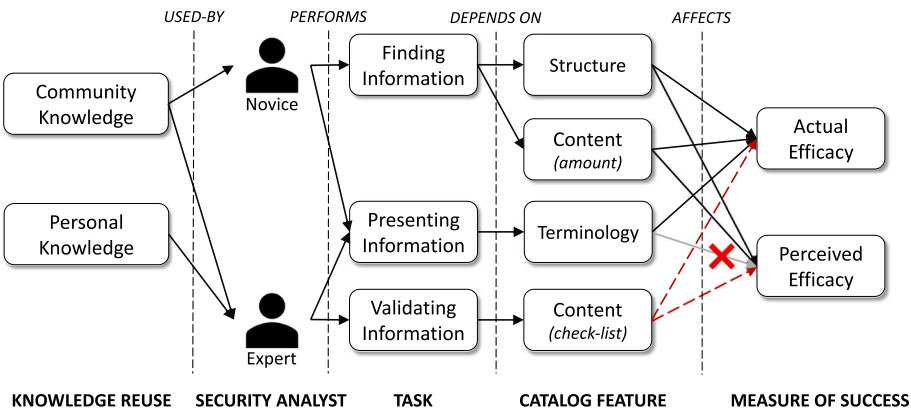


Fig. 5 Refined theory after the experiments

Table 8 A mock example of a threat catalog entry similar to the DOM-CAT

Generic Threat &ID	Specific Threat	C	I	A	Potential Impact	Increased Threat Warning Signs	Examples of previous use against [stakeholdes]	Threat Capabilities	Preventive Controls	Post- event Controls
Information INFF6	Hackers/ Social Engineering	x	x	x	[List of potential impacts on primary assets, e.g.	[List of cues signalling the rise of a corresponding threat, e.g.: Hackers affect data.]	[Examples of real cases illustrating the threat within the ATM domain. For illustrative purposes, we added an example from the Internet that was not a part of the original catalog]	Hackers - Social Engineering skills	PE-X; PE-Y; PE-Z	PO-A; PO-B; PO-C
					- Compromise of data or information.		In April 2020, the hackers behind the Maze Ransomware breached and successfully encrypted the information systems of VP San Antonio Aerospace. This company provides maintenance, repair, and overhaul services.			
					- Denial of services.]		The hackers gained access through remote desktop connection using a compromised administrator account and attacked the organisation's domain controllers, intranet servers, and file servers on two domains[a].			

Note: C - Confidentiality, I - Integrity, A - Availability.
Source of example: https://www.bleepingcomputer.com/news/security/us-aerospace-services-provider-breached-by-maze-ransomware/

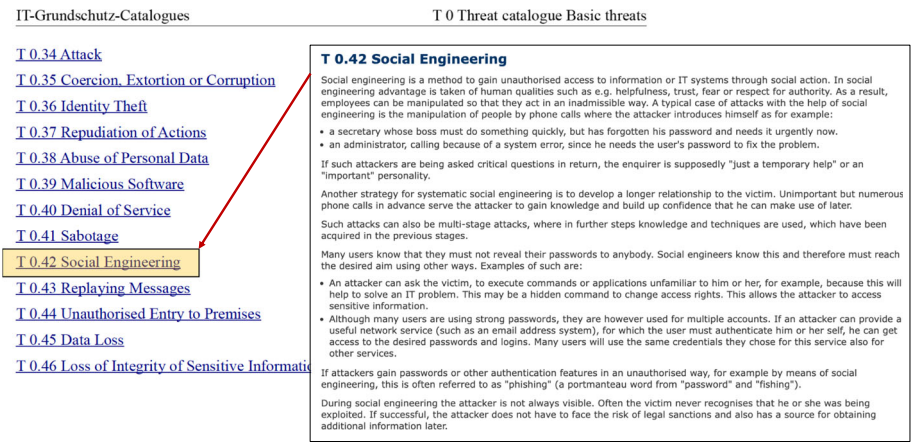


Fig. 6 Threat description entry from the GEN-CAT

> *"The identification of security controls was more difficult because you had to map them with the threats previously identified but there was no direct link in the catalog. It was mainly due to a problem of usability of the catalog."* (GEN CAT subject)

Catalog Content (Amount of Information) Security novices may feel overwhelmed by the amount of information. This is particularly the case for the general catalog, judged by one participant as:

> *"Difficult to consult for non-technical people." (GEN CAT subject)*

An interesting statement in this regard comes from an expert who was not assigned to any catalog but who had the chance to glance at the nearby group using a general catalog:

> *"I saw people near to me; they were not able to find stuff in the catalog, they were lost in the pages and eventually they always came with the same two or three items." (NOCAT subject)*

Catalog Content (Checklist) Regarding the ability of catalogs to cover a variety of threats and controls, the opinions expressed by the subjects was quite varied: security experts claimed that the suggestions in both catalogs were very generic, rather than specific, precise and well-defined threats and controls. The same result arose from the domain-specific catalog:

> *"[The catalog provided a] list of non-specific threats impacting the specific concept under investigation." (GEN CAT subject with security expertise)*

For security experts the catalog is perceived as a 'checklist'. In this way, the catalog is supposed to validate the efficiency and the coverage of the identified threats and security controls. In contrast, security novices were in general more satisfied by use of (any) catalog.

> *"I found the catalog useful, but I noticed that many threats were repeated." (DOM CAT subject)*

Catalog Terminology One feature of the catalog perceived as essential by every subject, irrespective of the type of catalog employed, is the fact that a catalog by itself provides a common terminology for all users. As suggested by our subjects:

"The catalog could be seen as a useful tool, able to formalize the controls that have been formulated in an informal way, and to lead them back into a common nomenclature." (DOM CAT subject)

The demand for a standard language caused by the need of sharing, discussing and presenting results by all stakeholders is an essential feature of the risk assessment process. Unsurprisingly, this aspect is mostly perceived as valuable by subjects who were not assigned to any catalog.

8 Suggestions for SRA Knowledge Reuse

We provide insights into the effective use of security knowledge catalogs worth further empirical validation.

Our study showed that despite a higher perceived usefulness (PU), novices using a more structured catalog were not more effective. Novices are expected to struggle with catalogs that are too big due to feeling overwhelmed by the number of security threats and controls. Hence, the amount of information presented in a catalog was expected to affect both the actual efficacy and perceived usefulness of a security assessment. However, we have not found a significant difference nor equivalence between the treatment groups (DOM CAT-S) and (GEN CAT-S) (Table 4 TOST returned 0.06, 0.37 for threats and 0.09, 0.22 for mitigations). It is encouraging to find that novices were (rather) underestimated and that they were able to conduct security risk assessment with both domain-specific and general catalogs.

Suggestion for Trainees The evidence regarding the equivalence of AE and PEOU of domain specific vs general catalogs suggests that novice analysts could (in principle) choose the catalog based on their preference and it would not effect the quality of the analysis. However, it is possible that novice analysts are at first less productive (not measured here) when working with a domain-specific catalog. For instance, beginner analysts might benefit from using a general catalog to first train the mechanics of the analysis technique (in our work SecRAM) and avoid getting overloaded with too much new terminology. Assuming that trainees pursue their career within the domain, it is crucial for them to also learn domain specific terminology for presenting and validating the analysis results (which are the other key SRA tasks). Second, trainees (with some experience with the technique) might benefit from using the domain-specific catalog to improve their understanding of domain-specific terminology (as discussed in Section 4), including security threats and controls. In addition, the evidence of higher perception of efficacy with domain-specific catalogs by novices might have a second order effect on the analysts. It may also raise the confidence levels of trainees and their motivation.

Suggestion for Experts The evidence presented in Table 6 suggests that experts using a catalog can still improve the quality of analysis outcome (AE). We have observed that experts use catalogs of threat and controls with a completely different goal in mind. Novice analyst will typically consult the catalog (as an oracle) to find the threats and controls, with the ultimate goal of reaching the correct outcomes (or minimizing the mistakes). In contrast, experts may already know some security threats and controls are applicable for a given domain component. Instead, they will browse the catalog (as a check-list) to make sure they have not missed out on an important threat, with the ultimate goal of analysis completeness.

Therefore, a domain-specific catalog with a more efficient querying mechanism to avoid time-consuming manual browsing could be beneficial for experts. For instance, an automated

support to query all applicable security mitigations for a category of security threats would be useful to quickly select the appropriate mitigations. Domain experts could use such tools to provide an initial analysis which could be submitted to security experts for a quality check. We cannot claim that domain experts could leverage such tools without consulting security experts because we have not found conclusive results for $H3$ (which essentially tested whether security experts without catalogs have the same "performance" as domain expert *with* catalogs). More experiments would be needed.

Similarly, domain experts (with experience in security risk analysis) could benefit from using the domain-specific catalog to avoid terminology misunderstandings (Tuma et al. 2021) and to verify analysis completeness.

9 Threats to Validity

We consider the threat of the various catalog length (e.g., the general catalog is ≈2500 pages, but the domain-specific is ≈55 pages) impacting the perceived efficacy. We mitigated this threat to **internal validity** by making available domain-specific catalogs of relatively large size (155 pages) and by preparing an index of the general catalogs (≈55 pages) that contained the list of available threats and security controls for ease of reference. The subjects had also access to the full version of the general catalog in electronic form (≈2500 pages). Secondly, we relied on expert assessors (who could have different opinions) for measuring the quality of analysis outcomes. To mitigate this threat, we reviewed the assessments to verify their validity. We observed that the expert assessors were consistent in marking bad and moderately good outcomes. Although, the Kendall's W test demonstrated moderate agreement between three expert in assessing the quality of threats (Kendall's $W = 0.45$) and security controls (Kendall's $W = 0.47$). We could only observe a slight difference in two assessments (one related to threats of a subject and one related to security controls of another subject). However, we observed that even when experts slightly disagreed on the mark, they would actually agree in the comment on the deficiencies of the evaluated work.

The main threat to **conclusion validity** is related to the *sample size*. We have included 33 subjects. But, Meyer et al. Meyer and Seaman (2013) show that it is possible to achieve significant results also for small samples. In addition, more than half of our participants (18) were experts with between 5 and 15 years of experience in performing SRA. This is still a fairly large sample considering the lack of security experts in most organizations (Blažič 2021; CyberSeek 2019). To understand the possible effect of subjects' background on the results we collected information about subjects' through demographics and a background questionnaire at the beginning of the study. To mitigate possible previous knowledge about the object of the study the subjects were given a step-by-step tutorial on the SRA method and received a textual description of the application scenario. Another threat to conclusion validity could be the variety of security analyses considered to be of low quality by the experts in the second experiment (6 out of 15). We attribute this effect to the assessors tendency to evaluate student outcomes with lower expectations. In addition, we would consider equal assessments of quality of outcomes a bigger validity threat as it could mean that the exercise was either too easy or too hard.

The main threat to **generalizability of results** is that both the risk assessment method and scenario were chosen within the ATM domain. However, the chosen risk assessment method is compliant with the ISO 27005 standard and the NIST Standard which can be

applied to different domains not just to the ATM. Therefore, this threat is fairly limited in our study. Another threat to validity is the *realism of the experimental setting*. Our experiment significantly counters this threat in comparison to the literature (Karpati et al. 2015; Mouaffo et al. 2014; Sindre and Opdahl 2005; Stålhane and Sindre 2014) as we used a duration of two days rather than a couple of hours or less. This longer duration, suggested by Labunets et al. (2013), allowed us to use a complex enough application scenario and thus to generalize our results to real projects. In addition to the longer duration, we limited threats to conclusion validity because (a) subjects were trained by an expert in the method who usually trains professionals working in the ATM domain, and (b) subjects had two full days to apply the method to a new ATM operational concept.

10 Related Work

We contextualize our contributions with respect to the related literature investigating knowledge reuse from an empirical angle, particularly what concerns reuse of security-relevant knowledge.

Empirical Investigations of Security Knowledge Reuse Riaz et al. (2017) have proposed the use of templates for security requirements and have reported a varying degree of success (depending on the type of requirements). While recall of expected security goals has been low the participants have been able to identify more implied goals (Riaz et al. 2016). Another empirical study on situational awareness (Hibshi et al. 2016) to analyze the difference in patterns of decision based on implicit knowledge re-use (expert vs not-expert) has identified several interesting patterns but no major difference in situational awareness of attacks was reported on the basis of the background knowledge of the participants.

Yskout et al. (2015) have investigated the effect of using security patterns on the security of a software design when used by designer with limited security expertise. The study revealed that the designers who used security patterns did not produce a more secure software design. However, they did not investigate which capabilities of the security patterns had an effect on the security of the software design.

Knowledge-Based Security Threat Analysis Security threat and risk analyses consist of techniques and methods that are used for systematically analyzing the attacker's profile vis-a-vis assets of value to organizations. This study investigates the SecRAM technique, however several alternative approaches have been proposed (Tuma et al. 2018). Such techniques are often performed on models representing the software architecture of a system. The purpose of analyzing security threats at this stage is to ultimately identify security holes and plan for necessary security solutions. Therefore, we consider existing literature that makes use of knowledge base (threat catalogs, vulnerability data bases, etc.) (Abe et al. 2013; Almorsy et al. 2013; Berger et al. 2016; Shostack 2014; Deng et al. 2011) to perform such analysis as related work. We refer the interested reader to a systematic literature review (Tuma et al. 2018) for a more detailed list of knowledge-based threat analysis techniques.

Security Catalogs da Silva Santos (2016); Santos et al. (2017) present a catalog of common architectural weaknesses (Common Architectural Weakness Enumeration, CAWE). CAWE identifies and categorizes types of vulnerabilities originating in software architecture design and provides mitigations accordingly. da Silva Santos (2016) also analyze the vulnerabilities of four real systems to discover their cause and find that up to 35% vulnerabilities were actually rooted in the architectural design. Arce et al. (2014) compiled a list

of top 10 security design flaws and provide guidelines on how to avoid them. The illustrative examples showcasing the flaws are very useful for understanding the practical saecurity impacts.

Security Attack Databases Common Attack Pattern Enumerations and Classifications (CAPEC) (Barnum 2008) is a comprehensive repository of known attack patterns employed by adversaries to exploit known weaknesses in software systems. In addition to the description of an attack, CAPEC provides attack consequences and possible mitigations, as well as a list of vulnerabilities and weaknesses related to each attack. Finally, it is associated with Common Weakness Enumeration (CWE) MITRE (2020b) project, which is a community-developed list of common software security weaknesses. CWE aids developers and security practitioners since it serves as a common language for describing security weaknesses in architecture and implementation. It also provides different mitigation and prevention techniques that could be used to eliminate weaknesses.

11 Conclusion

Security catalogs are an essential part of the SRA process: "as the [security] field evolves and establishes best practices, knowledge management can play a central role in encapsulating and spreading the emerging discipline more efficiently" (Barnum and McGraw 2005).

The aim of catalogs of threats and security controls is to put best security practices into a uniform format that can be re-used. In this paper, we have presented an interview study with 20 experts and experiments with 33 participants (both from academia and industry) studying the impact of codified knowledge (catalogs) on the SRA process. We have investigated the effect of using domain-specific catalogs versus general catalogs in both qualitative and quantitative terms. We have compared them with the effects of using the same method by security experts but without catalogs. The results of our study also supported hyp. 3 of Gray and Meister (2004) that the impact on knowledge sourcing will be stronger under conditions of higher intellectual demands.

In summary, our study shows that with the use of the catalogs a satisfactory number of threats and controls can be identified. If security expertise is expensive to get, a domain-specific security catalog is your second-best bet.

In this work, we proposed a new evidence-based theory for security knowledge reuse and identified a set of features that contribute to a knowledge reuse process. We encourage researchers to explore further the effect of these features on the core SRA task and validate our findings through large-scale experiments with practitioners and novice security analysts and more specific case studies with companies.

Acknowledgements The authors would first like to express gratitude to the participants of this study who have generously contributed. We would also like to thank our past-co-authors Martina De Gramatica and Le Minh Sang Tran from UTrento, Bjorn Sohaug and Ketil Stoelen from SINTEF, Alessandra Tedeschi and Martina Ragosta from DeepBlue, John Hird and Rainer Koehele from Eurocontrol for helping us with the data collection and organizing the focus groups and the trainings. Without everybody's time and expertise this paper would not have been possible.

Part of the the work was done while K. Labunets and F.M. Paci were at the University of Trento. This research was partially supported by the SESAR project EMFASE and the European Union and the FP7 project 285223 (SECONOMICS).

The datasets generated and analysed during the study are not publicly available due confidentiality constraints with industrial partners but example data can be made available from the corresponding author on reasonable request.

Declarations

Conflicts of interest The authors declare that they have no conflict of interest.

Information Systems Studies Using 5-item Likert Scale

To collect the sample of IS studies we used Google Scholar search service, as it allows to search in the text of the papers, and the following criteria:

- Publication year: between 2010 and 2016.
- Journals: "MIS Quarterly" (MISQ), "INFORMS Information Systems Research" (ISR), and "INFORMS Management Science" (ManSci).
- Search terms: ("5-point scale" OR "Likert scale") AND ("standard deviation" OR "stdev").

In total our search returned 22 papers that were published in MISQ, 20 papers in ISR, and 14 papers in ManSci journals. After checking the papers, we obtained a sample of 7 papers that reported descriptive statistics (incl. number of subjects, mean and standard deviation of each group) of variables on a 5-item Likert scale. Table 9 describes the final set of selected papers.

Table 9 Information systems studies using 5-item likert acale

Authors	Title	Journal	Year	Variables	#subj.
Ferratt et al.	Synergy and Its Limits in Managing Information Technology Professionals	ISR	2012	Job search behaviour	251
Gopal and Koka	The asymmetric benefits of relational flexibility: evidence from software development outsourcing	MISQ	2013	Requirements Uncertainty; Human Asset Specificity; Employee Turnover; Client MIS Experience; Quality	105
Gove and Parson	Is query reuse potentially harmful? Anchoring and adjustment in adapting existing database queries	ISR	2010	Reuse of queries results in higher confidence in query correctness	157
Maruping and Magni	Motivating Employees to Explore Collaboration Technology in Team Contexts	MISQ	2014	CT Exploration; Team Empowerment; Continued Intention to Explore; Continued Expectation to Explore; Perceived Usefulness; Facilitating Conditions; Personal Innovativeness with IT; Intention to Explore; Training; Task Interdependence	212
Montizaan et al.	The Impact of Negatively Reciprocal Inclinations on Worker Behavior: Evidence from a Retrenchment of Pension Rights	Man. Sci.	2016	Negative reciprocity; Positive reciprocity	5287
Phang et al.	What Motivates Contributors vs. Lurkers? An Investigation of Online Feedback Forums	ISR	2016	Civic skill	101
Sutanto et al.	Addressing the Personalization-Privacy Paradox: An Empirical Assessment From a Field Experiment on Smartphone Users	MISQ	2013	Excessive advertisements	60

References

Abe T, Hayashi S, Saeki M (2013) Modeling security threat patterns to derive negative scenarios. In: Proc. of the 20th Asia-Pacific Software Eng. Conf., vol. 1. IEEE, p 58–66

Agency CIS (2023) Cisa security bulletins. https://www.cisa.gov/uscert/ncas/bulletins

Almorsy M, Grundy J, Ibrahim AS (2013) Automated software architecture security risk analysis using formalized signatures. In: Proc. of the 35th Int. Conf. on Software Eng., p 662–671

Arce I, Clark-Fisher K, Daswani N, et al (2014) Avoiding the top 10 software security design flaws. IEEE Comput Soc Cent Secure Des (CSD), Tech Rep

Arora A, Belenzon S, Patacconi A (2015) Knowledge sharing in alliances and alliance portfolios. Available at SSRN 2719747

Barnum S (2008) Common attack pattern enumeration and classification (CAPEC) schema. Department of Homeland Security

Barnum S, McGraw G (2005) Knowledge for software security. IEEE Secur Priv 3(2):74–78

Berger BJ, Sohr K, Koschke R (2016) Automatically extracting threats from extended data flow diagrams. In: Proc. of the 8th Int. Symp. on Eng. Secure Software and Systems, pp. 56–71

Bharadwaj AS (2000) A resource-based perspective on information technology capability and firm performance: an empirical investigation. MIS Quart 24:169–196

Blažič BJ (2021) Cybersecurity skills in eu: New educational concept for closing the missing workforce gap. In: Cybersecurity Threats with New Perspectives

Boh WF (2008) Reuse of knowledge assets from repositories: A mixed methods study. Inform Manag 45(6):365–375

Boh WF (2014) Knowledge sharing in communities of practice: examining usefulness of knowledge from discussion forums versus repositories. Data Base Adv Inf Sy 45(2):8–31

BSI G (2017) Bsi standards 100-1, 100-2, 100-3, 100-4. https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Standards-und-Zertifizierung/IT-Grundschutz/it-grundschutz_node.html

Caralli R, Stevens J, Young L, et al (2007) Introducing octave allegro: Improving the information security risk assessment process. Tech. Rep. CMU/SEI-2007-TR-012, Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA, http://resources.sei.cmu.edu/library/asset-view.cfm?AssetID=8419

Center NCS (2021) 10 steps to cyber security. https://www.ncsc.gov.uk/collection/10-steps

Cruzes DS, Jaatun MG, Bernsmed K, et al (2018) Challenges and experiences with applying microsoft threat modeling in agile development projects. In: Proc. of the 25th Australasian Software Eng. Conf., IEEE, pp 111–120

CyberSeek (2019) Cybersecurity Supply/Demand Heat Map. https://www.cyberseek.org/heatmap.html

Davis FD (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quart 13:319–340

Deng M, Wuyts K, Scandariato R et al (2011) A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements. Req Eng 16(1):3–32

Dixon N (2002) The neglected receiver of knowledge sharing. Ivey Businees J 66:35–40

Food and Drug Administration (2001) Guidance for industry: Statistical approaches to establishing bioequivalence

Fredriksen R, Kristiansen M, Gran BA, et al (2002) The coras framework for a model-based risk management process. In: Proc. of the 21st Int. Conf. on Computer Safety, Reliability, and Security, Springer, pp 94–105

Gamma E, Helm R, Johnson R et al (1995) Design Patterns: Elements of Reusable Object-oriented Software. Addison Wesley, Boston

Garicano L, Wu Y (2012) Knowledge, communication, and organizational capabilities. Organ Sci 23(5):1382–1397

Gray PH, Meister DB (2004) Knowledge sourcing effectiveness. Manag Sci 50(6):821–834

Gritzalis D, Iseppi G, Mylonas A, et al (2018) Exiting the risk assessment maze: A meta-survey. ACM Comput Surv 51(1). https://doi.org/10.1145/3145905

Group SSI (2021) Building security in maturity model (bsimm12). https://www.bsimm.com

Guest G, MacQueen KM, Namey EE (2011) Applied Thematic Analysis. Sage, Thousand Oaks

Hibshi H, Breaux TD, Riaz M et al (2016) A grounded analysis of experts' decision-making during security assessments. J Cybersecurity 2(2):147–163

for Internet Security C (2023) Cis critical security controls. https://www.cisecurity.org/controls

Jafari AJ, Rasoolzadegan A (2020) Security patterns: A systematic mapping study. J Comput Lang 56:100938

Jedlitschka A, Juristo N, Rombach D (2014) Reporting experiments to satisfy professionals' information needs. Empir Soft Eng 19(6):1921–1955

Kankanhalli A, Lee OKD, Lim KH (2011) Knowledge reuse through electronic repositories: A study in the context of customer service support. Inform Manag 48(2):106–113

Karahasanovic A, Kleberger P, Almgren M (2017) Adapting threat modeling methods for the automotive industry. In: Proc. of the 15th European Conf. on Embedded Security in Cars, p 1–10
Karpati P, Redda Y, Opdahl AL et al (2014) Comparing attack trees and misuse cases in an industrial setting. Inform Soft Tech 56(3):294–308
Karpati P, Opdahl AL, Sindre G (2015) Investigating security threats in architectural context: Experimental evaluations of misuse case maps. J Syst Soft 104:90–111
Knowles MS (1970) The modern practice of adult education; andragogy versus pedagogy
Labunets K, Massacci F, Paci F, et al (2013) An experimental comparison of two risk-based security methods. In: Proc. of the 7th ACM/IEEE Int. Symp. on Empirical Software Eng. and Measurement, p 163–172
Labunets K, Paci F, Massacci F, et al (2014a) A first empirical evaluation framework for security risk assessment methods in the atm domain. Proc. of the 4th SESAR Innovation Days
Labunets K, Paci F, Massacci F, et al (2014b) An experiment on comparing textual vs. visual industrial methods for security risk assessment. In: Proc. of the 4th IEEE Int. Workshop on Empirical Requirements Eng. at the 22nd IEEE Int. Requirements Eng. Conf., pp. 28–35
Labunets K, Massacci F, Paci F, et al (2017a) Model comprehension for security risk assessment: an empirical comparison of tabular vs. graphical representations. Empir Soft Eng 22(6):3017–3056
Labunets K, Massacci F, Tedeschi A (2017b) Graphical vs. tabular notations for risk models: on the role of textual labels and complexity. In: Proc. of the 12th ACM/IEEE Int. Symp. on Empirical Software Eng. and Measurement, IEEE, pp 267–276
Leach MJ (2006) Evidence-based practice: A framework for clinical practice and research design. Int J Nurs Pract 12(5):248–251
Lund MS, Solhaug B, Stølen K (2010) Model-driven risk analysis: the CORAS approach. Springer Science & Business Media
Markus LM (2001) Toward a theory of knowledge reuse: Types of knowledge reuse situations and factors in reuse success. J Manag Inform Syst 18(1):57–93
Massacci F, Paci F (2012) How to select a security requirements method? a comparative study with students and practitioners. In: Proc. of the 17th Nordic Conf. on Secure IT Systems, Karlskrona, Sweden, Springer, Karlskrona, pp 89–104
Meyer JP, Seaman MA (2013) A comparison of the exact Kruskal-Wallis distribution to asymptotic approximations for all sample sizes up to 105. J Exp Educ 81(2):139–156
Meyners M (2012) Equivalence tests-a review. Food quality and preference 26(2):231–245
MITRE (2022) Mitre att&ck enterprise matrix. https://attack.mitre.org/matrices/enterprise/
MITRE (2020a) CVE - Common Vulnerabilities and Exposures. https://cve.mitre.org
MITRE (2020b) CWE - Common Weakness Enumeration. https://cwe.mitre.org
Mouaffo A, Taibi D, Jamboti K (2014) Controlled experiments comparing fault-tree-based safety analysis techniques. In: Proc. of the 18th Int. Conf. on Evaluation and Assessment in Software Eng., ACM, p 46:1–46:10
Newell A (1982) The knowledge level. Artif Intell 18(1):87–127
O'Dell C, Grayson CJ (1998) If only we knew what we know: Identification and transfer of internal best practices. Calif Manag Rev 40(3):154–174
Opdahl AL, Sindre G (2009) Experimental comparison of attack trees and misuse cases for security threat identification. Inform Soft Tech 51(5):916–932
OWASP (2021) Owasp top 10. https://owasp.org/www-project-top-ten/
Pilat L, Kaindl H (2011) A knowledge management perspective of requirements engineering. In: Proc. of the 5th IEEE Int. Conf. on Research Challenges in Information Science, IEEE, p 1–12
Publicas MDA (2012) Magerit - methodology for information systems risk analysis and management. https://administracionelectronica.gob.es/pae_Home/pae_Documentacion/pae_Metodolog/pae_Magerit.html
Raman R, Bharadwaj A (2010) Knowledge and agency based performative deviations in practice transfer routines: The case of evidence-based medicine. Available at SSRN 1907412
Riaz M, Stallings J, Singh MP, et al (2016) Digs: A framework for discovering goals for security requirements engineering. In: Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement. Association for Computing Machinery, New York, NY, USA, ESEM '16. https://doi.org/10.1145/2961111.2962599
Riaz M, King J, Slankas J et al (2017) Identifying the implied: Findings from three differentiated replications on the use of security requirements templates. Empir Softw Eng 22(4):2127–2178
Rus I, Lindvall M (2002) Knowledge management in software engineering. IEEE Soft 19(3):26–38
SANS (2011) SANS Top 25 Software Errors. https://www.sans.org/top25-software-errors/
Santos JC, Tarrit K, Mirakhorli M (2017) A catalog of security architecture weaknesses. In: Proc. of the Int. Conf. on Software Architecture Workshops, p 220–223

Scandariato R, Wuyts K, Joosen W (2014) A descriptive study of microsoft's threat modeling technique. Req Eng 1–18
Scandariato R, Wuyts K, Joosen W (2015) A descriptive study of microsoft's threat modeling technique. Req Eng 20(2):163–180
Schuirmann D (1981) On hypothesis-testing to determine if the mean of a normal-distribution is contained in a known interval. Biometrics 37(3):617
Schultze U, Leidner DE (2002) Studying knowledge management in information systems research: discourses and theoretical assumptions. MIS Quart 26:213–242
Schultze U, Stabell C (2004) Knowing what you don't know? discourses and contradictions in knowledge management research. J Manag Stud 41(4):549–573
Schumacher M, Fernandez-Buglioni E, Hybertson D et al (2006) Security Patterns: Integrating Security and Systems Engineering. John Wiley & Sons, Chichester
la Sécurité Des Systèmes D'information (ANSSI) AND (2019) Ebios risk manager. https://www.ssi.gouv.fr/uploads/2019/11/anssi-guide-ebios_risk_manager-en-v1.0.pdf
Shostack A (2014) Threat modeling: Designing for security. John Wiley & Sons, Indianapolis
da Silva Santos JC (2016) Toward establishing a catalog of security architecture weaknesses. https://scholarworks.rit.edu/theses/9004
Sindre G, Opdahl AL (2005) Eliciting security requirements with misuse cases. Req Eng 10(1):34–44
Souag A, Mazo R, Salinesi C, et al (2015) Reusable knowledge in security requirements engineering: a systematic mapping study. Req Eng 1–33
of Standards NI, Technologies (2023) Cyber security framework v1.1. https://www.nist.gov/cyberframework
of Standards NI, Technology (2012) Nist special publication 800-30 - revision 1 - guide for conducting risk assessment. https://www.nist.gov/privacy-framework/nist-sp-800-30
of Standards NI, Technology (2020) Nist special publication 800-53 - revision 5 - security and privacy controls for information systems and organizations. https://csrc.nist.gov/publications/detail/sp/800-53/rev-5/final
Stålhane T, Sindre G (2014) An experimental comparison of system diagrams and textual use cases for the identification of safety hazards. Int J Inform Syst Model Design 5(1):1–24
Tuma K, Scandariato R (2018) Two architectural threat analysis techniques compared. In: Proc. of the 12th European Conf. on Software Architecture, Springer, pp 347–363
Tuma K, Widman M (2021) Seven pain points of threat analysis and risk assessment in the automotive domain. IEEE Secur Priv 19(5):78–82
Tuma K, Calikli G, Scandariato R (2018) Threat analysis of software systems: A systematic literature review. J Syst Softw 144:275–294
Tuma K, Sion L, Scandariato R, et al (2020) Automating the early detection of security design flaws. In: Proc. of the 23rd ACM/IEEE Int. Conf. on Model Driven Eng. Languages and Systems, p 332–342
Tuma K, Sandberg C, Thorsson U et al (2021) Finding security threats that matter: Two industrial case studies. J Syst Soft 179:111003
Wasko MM, Faraj S (2000) "It is what one does": why people participate and help others in electronic communities of practice. J Strat Inf Syst 9(2):155–173
Wuyts K, Scandariato R, Joosen W (2014) Empirical evaluation of a privacy-focused threat modeling methodology. J Syst Soft 96:122–138
Yin RK (2010) Qualitative Research from Start to Finish. Guilford Press, New York
Yskout K, Scandariato R, Joosen W (2015) Do security patterns really help designers? In: Proc. of the 37th Int. Conf. on Software Eng., IEEE, p 292–302
Zhang C, Budgen D (2012) What do we know about the effectiveness of software design patterns? IEEE Trans Soft Eng 38(5):1213–1231