The Journal of Systems and Software 214 (2024) 112062

Contents lists available at ScienceDirect

The Journal of Systems & Software

journal homepage: www.elsevier.com/locate/jss

In practice

Evaluating software security maturity using OWASP SAMM: Different approaches and stakeholders perceptions☆

Davide Fucci [a,*], Emil Alégroth [a], Michael Felderer [a,b,c], Christoffer Johannesson [d]

[a] Blekinge Institute of Technology, Karlskrona, Sweden
[b] German Aerospace Center (DLR), Cologne, Germany
[c] University of Cologne, Cologne, Germany
[d] Ericsson, Karlskrona, Sweden

ARTICLE INFO

Dataset link: https://doi.org/10.5281/zenodo.7730021

Keywords:
OWASP SAMM
Industry-academia collaboration
Software security

ABSTRACT

Background: Recent years have seen a surge in cyber-attacks, which can be prevented or mitigated using software security activities. OWASP SAMM is a maturity model providing a versatile way for companies to assess their security posture and plan for improvements.
Objective: We perform an initial SAMM assessment in collaboration with a company in the financial domain. Our objective is to assess a holistic inventory of the company security-related activities, focusing on how different roles perform the assessment and how they perceive the instrument used in the process.
Methodology: We perform a case study to collect data using SAMM in a lightweight and novel manner through assessment using an online survey with 17 participants and a focus group with seven participants.
Results: We show that different roles perceive maturity differently and that the two assessments deviate only for specific practices making the lightweight approach a viable and efficient solution in industrial practice. Our results indicate that the questions included in the SAMM assessment tool are answered easily and confidently across most roles.
Discussion: Our results suggest that companies can productively use a lightweight SAMM assessment. We provide nine lessons learned for guiding industrial practitioners in the evaluation of their current security posture as well as for academics wanting to utilize SAMM as a research tool in industrial settings.

Editor's note: Open Science material was validated by the Journal of Systems and Software Open Science Board.

1. Introduction

Companies, from small to large enterprises, are increasingly becoming the targets of cyberattacks.[1] The increasing threats from malevolent actors are pushing organizations to implement secure software development life cycles (SSDLC) to support software design and development as well as detection and removal of vulnerabilities (Ramirez et al., 2020).

Practitioners in the security community have proposed several initiatives to quantify security practices across different parts of an organization (Ramirez et al., 2020). One such initiative is the Software Assurance Maturity Model (SAMM) proposed by OWASP[2]—an open community dedicated to enabling application security through tools, guides, and methodologies. SAMM provides a tool (i.e., a questionnaire) for assessing an organization's current software security posture in five main areas, from governance to operations. Based on the outcome of the assessment, the organization can set an improvement target, create a roadmap, and track its progress through subsequent assessment iterations.

In 2010, according to an industry survey (Geer, 2010), approximately 30% of software companies used a secure development lifecycle. Since then, practitioners worldwide use SAMM[3] to improve their security posture. Researchers use SAMM to structure their analysis and interventions (e.g., for web application security Teodoro and Serrao, 2011) and to categorize existing literature (e.g., in secondary studies Wen, 2017). However, to the best of our knowledge, no previous

☆ Editor: Marcos Kalinowski.
* Corresponding author.
E-mail addresses: davide.fucci@bth.se (D. Fucci), emil.alegroth@bth.se (E. Alégroth), micheal.felderer@bth.se (M. Felderer), christoffer.johannesson@ericsson.com (C. Johannesson).
[1] https://www.redscan.com/news/nist-nvd-analysis-2021-record-vulnerabilities/
[2] https://owaspsamm.org
[3] https://owaspsamm.org/practitioners/

https://doi.org/10.1016/j.jss.2024.112062
Received 14 March 2023; Received in revised form 17 November 2023; Accepted 12 April 2024
Available online 22 April 2024
0164-1212/© 2024 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

work has investigated how the SAMM assessment is carried out in practice and how it is perceived by the practitioners who will eventually base their improvement activities on its results.

We performed an industrial case study focusing on the *initial SAMM assessment* at COMPANY. In a survey with 17 employees in different roles, we collected data using a modified version of the SAMM questionnaire to measure the understandability of the questions and the respondents' confidence in their answers. We compared the answers to the questionnaire with qualitative data collected in a focus group with seven participants from COMPANY.

The results show that different roles perceive the organization's maturity differently. However, the results also indicate that individuals have difficulty answering questions outside their expertise. We note that considering the respondents' perceptions of their answers has some impact in specific areas of the SAMM assessment. Regardless, we show that survey and focus group results overlap across several security practices.

This paper makes the following contributions.

- The design of the first study to investigate a real-world lightweight SAMM assessment.
- A novel approach and lessons learned to support organizations applying OWASP SAMM with multiple stakeholders.
- An improved version of the OWASP SAMM assessment questionnaire.

The paper is organized as follows. In Section 2, we introduce SAMM and other models that embed security in the software development lifecycle. In Section 3, we present our case study in the context of COMPANY, and show the results in Section 4. Section 5 discusses the results and presents lessons learned. Finally, we conclude the paper and provide future work in Section 6.

2. Related work

This section presents an overview of related work regarding security models in software development from practitioners and academic literature and an overview of OWASP SAMM.

2.1. Security models in software development

Several maturity models are focusing on evaluating the security posture of specific aspects of systems—e.g., managing data centers (Lima et al., 2017) and designing service-oriented architectures (Kassou and Kjiri, 2012). In this section, we focus on models that evaluate the maturity of software security activities.

Together with SAMM, there are several other maturity models for software security. Among these, the *Build Security In Maturity Model* (BSIMM) groups 121 security activities into five domains, each with three practices. Unlike SAMM, BSIMM is descriptive as it records the security activities of more than 200 companies since 2008.

Weir et al. (2021, 2022) present a 12-year longitudinal analysis of the data obtained from BSIMM. They first observed an increase in the use of software security groups, whereby security activities become diluted throughout the organizations due to the adoption of DevSecOps practices. Regarding the number of activities, they show a trend of steady increase from 2016. This trend of activities signals a *shift-left* of security practices as techniques are applied earlier in the lifecycle and become a prerogative of the developers rather than security experts. The authors clustered activities based on their co-occurrences in the last five years of their evaluation. More than 60% of the companies use activities that are heavily skewed towards the *Governance* domain, showing that despite the emphasis on shift-left, security is still a top-down (managerial) concern. The top 10 activities with the most significant increase in 2016–2020 belong to the *Governance* and *Intelligence* domains, with three supporting only managerial roles, two supporting only developers, and five supporting both. Conversely, of the top-10 decreasing activities in the same period, three support managerial roles, five support developers, and two support both. Based on these results, the authors proposed a minimal set of activities to focus on early on, such as the establishment of a security expert role within the development team and of a cross-organizational team focused on other aspects, such as security training and planning.

Tashi and Ghernaouti-Helie (2009) present a holistic conceptual model for assessing information security inspired by the existing ISO 2700 family of security standards. In particular, the authors map engineering security standards (technical and operational concerns) to non-engineering ones (legal, human, and organizational aspects). The resulting model allows users to evaluate information security according to dimensions such as effectiveness and efficiency.

Ramirez et al. (2020) report a survey of secure software development standards and models. They compare and classify eight maturity models, including SAMM, according to their coverage of the SDLC phases and their approach to risk management. Compared to standards such as PA-DSS[4] and TOGAF,[5] SAMM does not offer certification, nor does it specify detailed coding of threats as in OWASP AVSV[6] or SAFECode.[7] Nevertheless, SAMM is more focused on risk management and training, presenting a complete spectrum of activities in these areas. Moreover, it is one of the three standards introducing mitigation activities for known risks (together with TOGAF and NIST[8]).

Similarly, De Win et al. (2009) compares OWASP CLASP—the precursor of SAMM which focused on different security views—to the Microsoft SDL,[9] and McGraw's Touchpoints (McGraw, 2004). The authors identified common activities, linked them to six phases of development, and created a hierarchy of frameworks. Of the three, Microsoft SDL is the more focused, providing concrete activities, whereas CLASP has the most comprehensive scope and is the heaviest to apply.

To understand how the best practices recommended in models such as SAMM are performed in practice, Assal and Chiasson (2018) interviewed 15 teams (including developers, testers, and QA) from 15 companies in North America. The authors matched the practices elicited during the interviews with a list of 12 practices distilled from SAMM, BSIMM, and Microsoft SDL. Most of the teams that participated in the study perform their activities post-development—e.g. penetration testing. Additionally, the participants do not consistently integrate security in the lifecycle but rather at specific stages. Results also show, for design and implementation, that developers deliberately violate best practices or circumvent limits imposed by their framework to achieve their functional goals. Similarly, it is common to exclude security from functional testing plans and code review sessions. In contrast to Assal and Chiasson, in this study we perform a complete maturity assessment of the practices reported in OWASP SAMM.

Such et al. (2016) investigated the practical use of 25 assurance techniques extracted from the ISO 27001 standard. They conducted 14 interviews and then surveyed 115 practitioners, showing that architectural review, vulnerability scans, and penetration tests are the most cost-effective among the reported techniques.

Using a questionnaire based on BSIMM, Jaatun et al. (2015) surveyed 20 Norwegian companies operating in the public sector. They showed a significant variation in their maturity, with the most mature organization implementing 87 of the 112 BSIMM activities, while the least mature one implementing nine (avg = 44). Follow-up interviews with the companies' stakeholders revealed that comprehensive security strategies and means to set and evaluate security achievements are commonly missing.

[4] https://www.pcisecuritystandards.org
[5] https://www.opengroup.org/togaf
[6] https://owasp.org/www-project-application-security-verification-standard/
[7] https://safecode.org
[8] https://csrc.nist.gov/publications/detail/sp/800-37/rev-2/final
[9] https://www.microsoft.com/en-us/securityengineering/sdl/

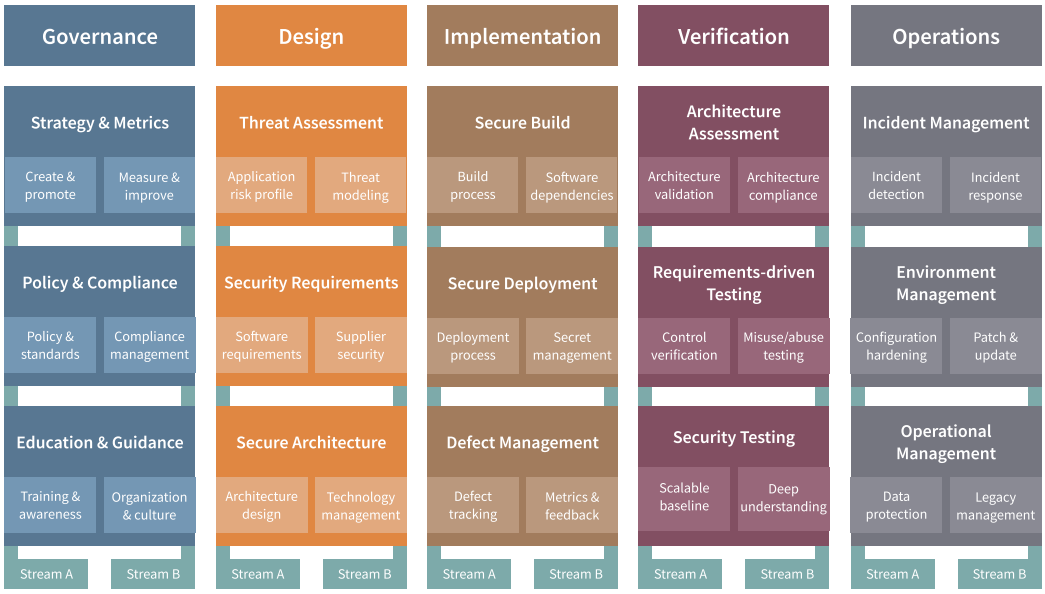


Fig. 1. OWASP SAMM structure.

2.2. The OWASP Software Assurance Maturity Model (SAMM)

OWASP SAMM is a prescriptive maturity model developed in 2008. It divides the SDLC into five business areas, each with three security practices that include activities helping the organization build its security assurance (see Fig. 1). The security practices have two streams with three maturity levels with increasingly more advanced objectives. Each stream has its objectives and covers different aspects of a practice. For example, the *Secure Build* practice is part of the *Implementation* business area, and it consists of six activities of increasing maturity divided into two streams—i.e., *Build Process* and *Software Dependencies*.[10]

The SAMM tool (i.e., questionnaire) consists of 90 questions (5 business areas *times* 3 practices *times* 2 streams *times* 3 maturity levels) which are answered on 4-point scales between 0 and 1, where zero indicates that the activity is *not present* and the remaining values (i.e., 0.25, 0.5, and 1) indicate increasingly broader scope in the adoption of the activity. Each practice receives a rating between 0 and 3 corresponds to the sum of the average answers in the two streams for each maturity level (see Formula (1)).

$$\text{Rating} = \sum_{\text{maturity}=1}^{3} \frac{(\text{StreamA} + \text{StreamB})_{maturity}}{2} \tag{1}$$

where $maturity \in [1, 2, 3]$ and $StreamA|B \in [0, 0.25, 0.50, 1]$

Therefore, a practice with a rating of 3 indicates that all activities in that practice, from the least to the most mature ones, are performed in the broadest scope possible. For example, for the *Build process* stream of the *Secure build* practice, one question asks:

> `Is your full build process formally described?`

Possible answers are:

> `No` (0 points)
> `Yes, for some applications` (0.25 points)
> `Yes, for at least half of the applications` (0.50 points)
> `Yes, for almost or all of the applications` (1 points)

[10] https://owaspsamm.org/model/implementation/secure-build/

After the assessment, SAMM allows the user to create a roadmap to improve on specific practices or areas over time. SAMM suggests dividing the roadmap into four phases, each lasting between three and six months. For this study, we created a roadmap based on the first SAMM assessment at COMPANY; however, its implementation is out of scope.

3. Research methodology implemented at COMPANY

We designed and carried out an exploratory case study following the guidelines of Runeson et al. (2012). As data collection methods, we chose a survey and focus group. Fig. 2 presents a timeline of the different phases.

3.1. Goal and research questions

COMPANY wants to find areas of improvement in its cybersecurity posture. To that end, we (researchers and industry practitioners) need an overview of COMPANY cybersecurity posture across teams and lifecycle phases.

Hence, our first goal is to **create a holistic inventory of security activities.** After discussing with COMPANY, we selected OWASP SAMM as a tool for collecting and assessing security-related activities. To address this goal, we answer the following research question.

RQ1: How do stakeholders in different roles evaluate the maturity of their organization using SAMM?

The answer to RQ1 fills the need of COMPANY and provides valuable lessons for other companies. In particular, we show how to consider the assessments of *multiple* roles when determining the maturity level of a company.

Our second goal is to **evaluate the SAMM assessment process and its questionnaire.** To address this goal, we answer the following research questions.

RQ2: How do different approaches for SAMM assessment impact its results?

RQ3: How do stakeholders in different roles perceive the SAMM assessment instrument?

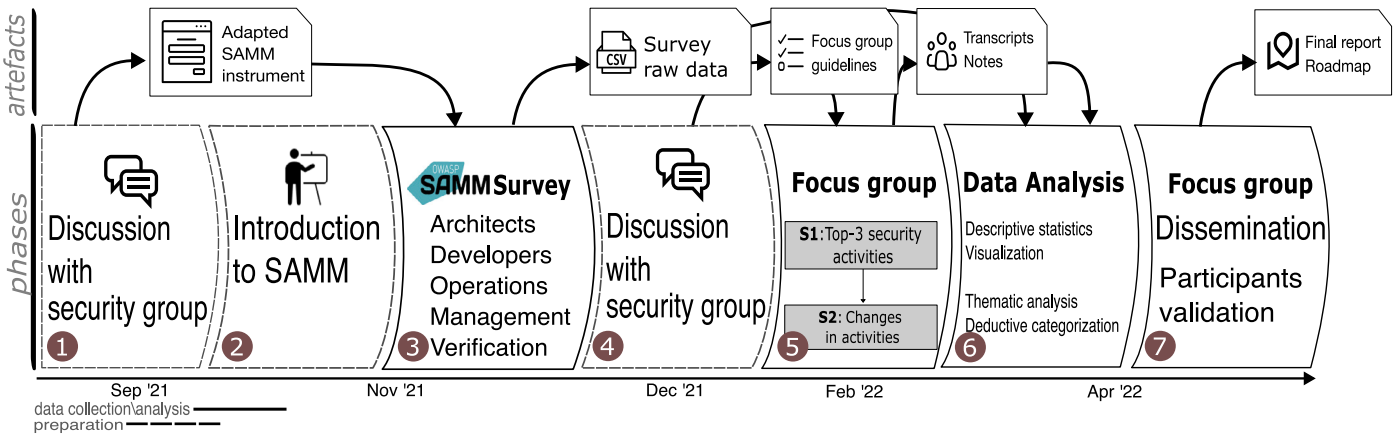


Fig. 2. Overview of the research workflow with Company.

The answers to RQ2 and RQ3 provide an assessment of SAMM itself. With RQ2, we focus on the methods to gather data to answer the SAMM questionnaire. Obtaining the data necessary to perform the assessment is time-consuming as an individual needs to interview stakeholders in scope for the SAMM investigation (e.g., a single team, teams working on the same project, or the entire company). Therefore, we propose a novel, alternative, and *lightweight* approach in which we let individual stakeholders of different roles covering the five business areas answer the SAMM questionnaire independently and create an overall assessment based on their aggregated answers. With RQ3, we assess the easiness of answering the questionnaire, the stakeholders' confidence in their answers, and how these constructs influence the SAMM maturity score.

3.2. Case description

COMPANY develops a platform designed to enable providers, such as banks, credit unions, and mobile network operators to offer financial services accessible by customers using different devices. The financial services provided by the platform include payments, money transfers, microfinance, insurance, and mobile banking, supporting a range of payment methods. Therefore, the platform is designed to be highly scalable, flexible, and secure, with advanced fraud detection and prevention mechanisms built in. Moreover, as part of their offer, the products securely store financial data which can be turned into analytics for the respective financial service providers to gain insights into customers' behaviors and preferences.

The security team at COMPANY consists of five individuals, including a security officer, security experts, and pentesters. The experts collaborate with the development teams, for example, by participating in risk analysis sessions, code reviews, and giving support when addressing security bugs. Each development team also include a security master—usually a developer with additional training in security, who supports the rest of the team and the security experts.

3.3. Design and instrument

In this study phase, we designed the data collection procedures for the case study. In an initial discussion, the industry practitioners clarified company-specific terminology, organizational structure, tasks, and roles concerning security. Together with the COMPANY security team lead (i.e., the fourth author of this paper), we adapted the SAMM questionnaire[11] to COMPANY-specific terminology and added notes to clarify how the question should be interpreted in the context of the company. The SAMM interview questionnaire contains 90 questions (see Section 2.2). For each question in the questionnaire, we also attached the following two questions (i.e., *meta-questions*), which are answered on a 4-point scale.

- How easy was it for you to answer this question?
- How confident are you in the answer you provided to this question?

The questions capture the *easiness* and the *confidence* constructs, respectively. Moreover, we included two additional fields, (*survey starts* and *survey ends*), in the survey instrument to estimate the effort of filling in the questionnaire.

3.4. Data collection

We collected data in two steps; first, a survey using the updated SAMM questionnaire (step 3 in Fig. 2), and, second, a focus group with different COMPANY stakeholders (step 5 in Fig. 2).

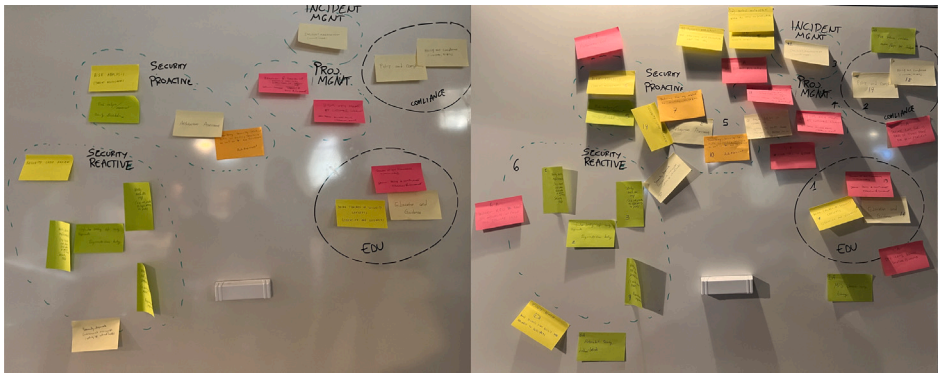
The security team lead (i.e., the fourth author), purposefully sampled representative survey participants (see Table 1) from COMPANY employees working in different roles. The average tenure of the participants in their role at COMPANY was 12 years. We determined the roles based on the recommendations of the SAMM assessment and the SAMM questionnaire—i.e., architects, developers, managers, operations, and verification. Thus, a purposeful sample aimed at covering all areas included in the assessment tool. Before distributing the survey, we held an online meeting with the participants to introduce the SAMM assessment process and the questionnaire (step 2 in Fig. 2). We distributed the survey via internal e-mail, giving the participants two weeks to reply.

The objective of this first step was to acquire the participants' perceptions of the level they performed security assurance at COMPANY in their own and other areas (i.e., SAMM business functions).

We discussed the survey results with the security group (step 4 in Fig. 2), which gave us insights to organize the focus group—i.e., we deemed it interesting and necessary to continue to include all role perspectives and accordingly invited representatives for each role. We collected data in a focus group with seven participants (three participants overlapping with the survey), representative of all roles we previously identified.

In the focus group, we elicited the state-of-practice of how COMPANY performs software security assurance. We captured nuances of how the practices were carried out in different areas and gained deeper insight into the organization's application of each practice. During the focus group, the first two authors moderated the discussion, which we structured in two sessions. First, we asked the participants to reflect on the top-3 security activities they perform, focusing on eliciting practices, processes, and tools. We used sticky notes, color-coded according

[11] https://owaspsamm.org/assessment



(a) Focus group 1st session: Top-3 security activities. (b) Focus group 2nd session: Changes in activities.

Fig. 3. Topics (clusters) discussed during focus groups.

Table 1
Survey and focus group participants.

ID	Role	Survey	Focus group	Time survey (min)
1	Management	✓	✓	42
2	Management	✓	✗	45
3	Management	✓	✗	19
4	Management	✗	✓	–
5	Developer	✓	✗	65
6	Developer	✓	✗	45
7	Developer	✓	✗	83
8	Developer	✓	✗	38
9	Developer	✓	✓	67
10	Developer	✓	✗	75
11	Developer	✓	✗	49
12	Developer	✓	✗	19
13	Architect	✓	✗	75
14	Architect	✗	✓	–
15	Operation	✓	✗	49
16	Operation	✓	✗	108
17	Operation	✗	✓	–
18	Operation	✗	✓	–
19	Verification	✓	✗	40
20	Verification	✓	✓	40
21	Verification	✓	✗	49

to the role, to register the answers and clustered them on a whiteboard visible to everybody. We clustered the sticky notes based on the SAMM areas Fig. 3a shows the whiteboard after the first focus group session.

We discussed, for each cluster, the answers given on the sticky notes to (i) validate our understandings (i.e., member checking) and (ii) let the participants present their rationale for their answers. We numbered the sticky notes to ease the traceability between the topic and the statements provided during the discussion. In the discussion, one researcher acted as moderator while another took notes.

In the second session, following the same structure, we asked the participants to add, remove, or replace any of the practices, processes, and tools presented on the sticky notes provided in the first session. Fig. 3b shows the whiteboard after the second focus group session. The purpose of this second session was to (i) confirm the results of the first session (e.g., that no one added practices that had already been mentioned) and (ii) identify if perceived areas of improvement aligned with areas of improvement shown by the SAMM assessment. The focus group lasted approximately three hours.

3.5. Data analysis

We analyzed the data in two steps according to the data sources—i.e., survey and focus group.

In the first step, we visualized the data from the survey using a spider diagram and bar charts and summarized it using descriptive statistics. We further grouped the results according to the role and weighted them according to confidence and easiness.

The second step of the analysis was qualitative. The two researchers, who were present at the focus group, applied deductive thematic analysis to the statements (n=83) collected during the focus group. We assigned each statement to a predefined theme—i.e., the 30 streams of SAMM security practices. Each statement could be associated with one or more themes. The two researchers involved in this analysis compared results to resolve disagreements and finally relied on a third researcher (the third author of this paper) for a final review. Once we obtained a stable characterization, the same two authors coded each statement as either supporting or impairing the fulfillment of each SAMM security practice. Moreover, the third author used the statements and their characterization to fill in the SAMM questionnaire. This allowed us to quantify the qualitative results and compare them to the first part of the analysis (i.e., survey results).

We validated the results in an online dissemination focus group (1 h) with all study participants. We presented the quantitative and qualitative results and asked the participants to comment and ask questions. Besides minor questions and clarifications, no objections were given to the results. We distributed the materials used in the seminar to the participants for further review and feedback. However, no additional feedback was given, implying that the participants agreed with the results of our analysis.

We provide a replication package[12] containing the SAMM questionnaires collected using the survey and focus group and the script to generate the results reported in this paper.

3.6. Validity threats

In this section, we discuss the potential threats to the validity of this work, grouped according to the categories proposed by Runeson et al. (2012).

Internal validity. This study does not seek to examine strong causal relationships; accordingly, internal validity is less of a concern. Comparisons (e.g., between stakeholders' roles and practices) are grounded in quantitative survey data, descriptive statistics, and visualization. We acknowledge that other causes may have impacted the results (e.g., the maturation of some participants between the survey and focus group). We also recognize that qualitative evidence could have been collected using, for instance, individual interviews. However, using the available

[12] https://doi.org/10.5281/zenodo.7730021

data sources, we were able to provide results to answer the research questions. In other words, additional data would provide more depth to the results but not necessarily new ones.

External validity. This threat is concerned with the generalizability of the results. This study involved 21 participants in different roles from a single company. The results can be used as a benchmark for companies of similar size, security demands, and domain as COMPANY. Therefore, we consider this study to have a high level of transferability but not analytical generalizability.

Construct validity. This threat concerns the operationalization of the constructs. Although for this study we consider and analyze an important construct which can have an effect on the results (i.e., participant's role), we acknowledge that there may be others (e.g., experience, gender) that can play a role in the assessment. We have used standard survey instruments (i.e., SAMM), which were further tailored by a stakeholder knowledgeable in the area. In our assessment, we included two additional questions regarding *easiness* and *confidence* in answering each of the 90 questions in SAMM. However, in total, the participants did not answer 34% of the *easiness* and 40% of the *confidence* questions. One explanation is that these participants could not understand such questions, posing a threat to the validity of the instrument. Finally, we performed member-checking of the data collected during the focus group.

Reliability. This threat concerns the extent to which the data and analysis depend upon the specific researchers. We address this by reporting extensive information regarding data collection, using a standard tool for the survey, and providing a replication package. We recognize that exact replication of the study results is unlikely, but the methodology is described in sufficient detail to replicate the research procedure.

4. Results

In this section, we report the results and answer the research questions.

4.1. RQ1: SAMM maturity evaluation by different roles

To answer RQ1, we plotted in a spider graph (see Fig. 4) the aggregated SAMM questionnaire answers according to the participants' roles. We show that different roles perceive varying levels of maturity in different practices. The only practice for which all roles seem to agree is *Secure Deployment*, assessed at a maturity level of approximately 1. Architects perceive a lower maturity than personnel in Verification in the *Security Requirements* practice and, to a lesser extent, in the *Secure Architecture*, which should fall under their responsibility. Moreover, architects perceive themselves as more mature than personnel in Operation in the *Operational Management* and *Environment Management* practices. Verification perceives higher maturity in practices that fall under their concerns, especially *Requirements* and *Security Testing*, compared to other roles. Similarly, developers perceive higher maturity in their areas of concern (e.g., *Secure Build* and *Secure Deployment*) but perceive lower maturity in neighboring areas, such as Architecture and Verification. Operations personnel perceive the lowest maturity across most practices, except those falling under their concern—i.e., *Incident*, *Environment*, and *Operational Management*. Nevertheless, other roles (architects and verification) perceive higher maturity in these areas. Finally, two practices concerning managers (i.e., *Strategy&Metrics* and *Policy&Compliance* show some of the lowest level of maturity, although they are perceived as the most mature by managers. The other practice in the *Governance* area, *Education&Guidance*, is perceived as more mature by architects and developers than managers. Conversely, verification and operations personnel perceive it as less mature.

RQ1: Different roles perceive maturity differently

All roles perceive higher-than-average maturity for most of the practices associated with their role. In particular, developers, verification, and management personnel perceived a high level of maturity in their areas of concern, whereas architects and operations personnel showed low maturity in their areas—i.e., *Design* and *Operation*.

4.2. RQ2: Different executions of SAMM assessment

On average, the participants took 49 min (std.dev = 25 min) to answer the SAMM questionnaire, including the two meta-questions, during the survey. The participants undertook this offline process at their own pace, which can be easily scaled up to the entire organization. On the other hand, the focus group—necessary to elicit the information needed to fill in the SAMM questionnaire—took approximately three hours and the availability of participants limits it.

In Fig. 5, we show how the survey questionnaire results overlapped with the results obtained from the focus group. Although there are minimal discrepancies for several practices (e.g., *Security Requirements* and *Security Testing* show exactly the same assessments), we observe two areas where the assessments vary substantially. In the *Design* area, there is one maturity point difference in the *Threat Assessment* and *Secure Architecture* practices. Practices in the *Operations* areas, specifically for *Operation Management* and *Environment Management*, show a similar variation.

RQ2: Differences between assessments are limited to specific practices

There are modest differences in the results yielded by a lightweight execution of SAMM (i.e., survey) compared to a more burdensome one (i.e., focus group). In particular, the latter underestimates the maturity level for the former, except for *Policy&Compliance*, *Architecture Assessment*, and *Incident Management* practices. Larger gaps are limited to the *Design* and *Operations* areas.

4.3. RQ3: Roles perception of SAMM questionnaire

We measured the perception of the SAMM instruments through two constructs, *easiness* and *confidence*.

Fig. 6 shows how the ease of answering questions related to different practices is distributed. Approximately 40% of the participants did not indicate an easiness level. Most questions are perceived as easy-to-answer with the exception of those in the *Strategy&Metrics* practice which were deemed difficult by managers, despite their involvement in the area. A similar observation holds for the *Policy&Compliance* practice. Architects found difficulties answering questions regarding *Security Requirements*, which falls under their area of concern, and *Requirements Testing*. In general, developers—and to a less extent managers—found it easier to answer questions across practices compared to other roles.

Fig. 7 shows a substantial agreement between the overall evaluation and the one consisting of easy-to-answer questions. *Operation* is the practice showing a higher maturity (although of less than half-point) when considering *easy* answers to the questionnaire. Considering the *difficult*-to-answer questions, the differences are more remarked. For example, participants evaluated the *Architecture Assessment* as zero—i.e., the lowest level of maturity—whereas the *Threat Assessment* practice shows half-point less maturity compared to the overall assessment. On the other hand, the *Defect Management* practice is evaluated one maturity point more than the overall evaluation, a result attributable to architects (i.e., the role reporting more mature score for this practice) finding it difficult to answer this part of the SAMM questionnaire.

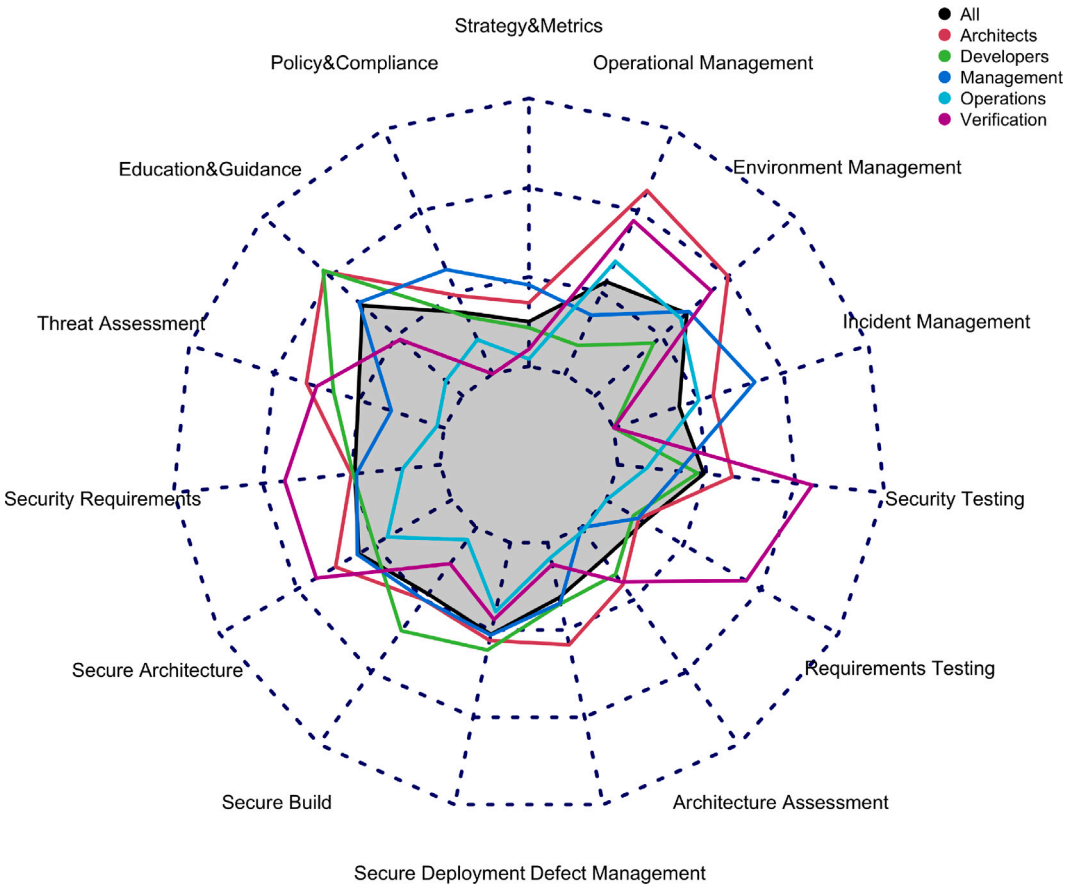


Fig. 4. Aggregated answers to the SAMM questionnaire by roles. The inner-most level corresponds to a maturity of 0 (i.e., the practice is not present) and the outer-most to a maturity of 3 (i.e., the practice is formalized and integrated).

We show how the confidence level in answering questions related to the SAMM practices is distributed in Fig. 8. Across practices, for approximately one-third of the answers, the participants decided not to fill in the additional question regarding confidence. The results are mostly distributed between *confident* and *unconfident* across the SAMM areas. Conversely, we only observe extreme (i.e., *very unconfident* or *very confident*) results for a few practices. The practices *Strategy&Metrics*, *Policy&Compliance*, *Threat Assessment*, and *Operational Management* did not receive *very confident* answers despite the appropriate roles (i.e., managers, architects, and operations personnel) participating in the assessment. *Incident Management* and *Environment Management* received the highest confidence rating; however, these results are, mostly based on the perceptions of developers and managers rather than operations personnel—i.e., the role which is the more involved with the activities within these practices. In general, architects and managers provided more confident answers, even for practices outside their concerns, whereas verification personnel did not report confidence for any questions across practices.

As shown in Fig. 9, confident answers result in a more mature assessment across practices (only for *Requirements Testing* the assessment is the same). Such divergence is generally limited to approximately half-point differences, with the largest difference observed in the *Incident Management*. Nevertheless, the general "shape" of the maturity distributions across practices is maintained. Conversely, the shape of the distribution of unconfident answers is contained within the overall one. Roles not in the *Governance* area (e.g., architects, operations) tend to overestimate its maturity despite providing unconfident answers. On the other hand, roles such as managers and operations underestimate the maturity of *Threat Assessment* and *Secure Build* areas when providing unconfident answers.

RQ3: Most roles can easily and confidently answer the questionnaire

Architects and managers are the roles that found the questionnaire more difficult to answer with respect to other roles. However, they also reported more confident answers. In particular, practices in the *Governance* business area appear problematic—i.e., they show answers with low confidence and high difficulty. Answers rated with low confidence and high difficulty change the distribution of maturity more remarkably compared to high-confidence and high-easiness ones.

5. Discussion

Our experience running the SAMM assessment as a survey and as a focus group shows that maturity is, in some areas of the SSDLC, perceived differently by different stakeholders based on their roles. Performing a SAMM assessment showed COMPANY insights into their way of working about security and brought up misalignments that can hinder prioritization and implementation of improvement actions. For example, according to personnel concerned with verification, the *Requirement Testing* practice has a maturity level of 2, indicating that reviews are performed to discover application-specific risks against the security requirements.[13] However, managers perceived this practice as opportunistic at best (i.e., maturity ≤ 1). Based on this perception, it is plausible that managers can be misguided into realizing improvements in this practice, neglecting other practices, such as *Strategy&Metrics*, which are in fact more needed according to the evaluation results.

[13] https://owaspsamm.org/model/verification/requirements-driven-testing/

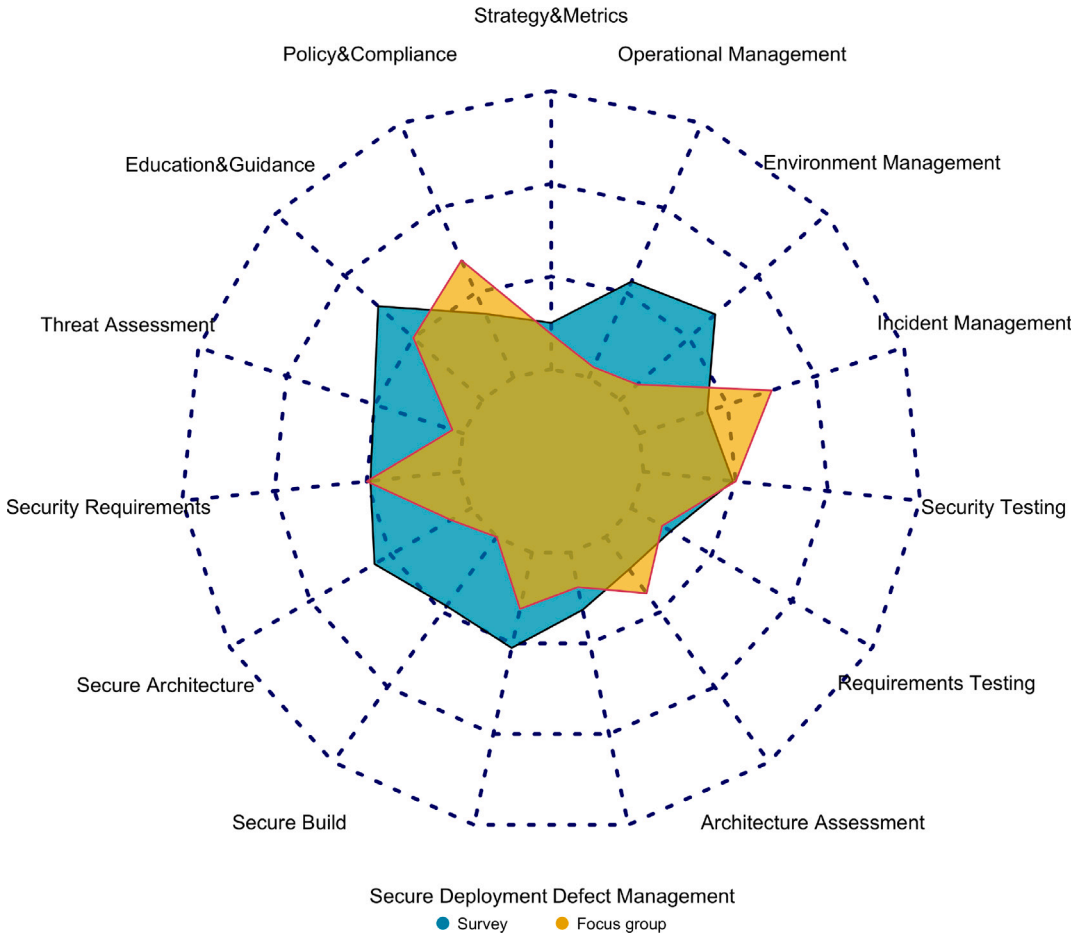


Fig. 5. Overlap between the overall SAMM questionnaire survey results and SAMM questionnaire results compiled using statements from the focus group.

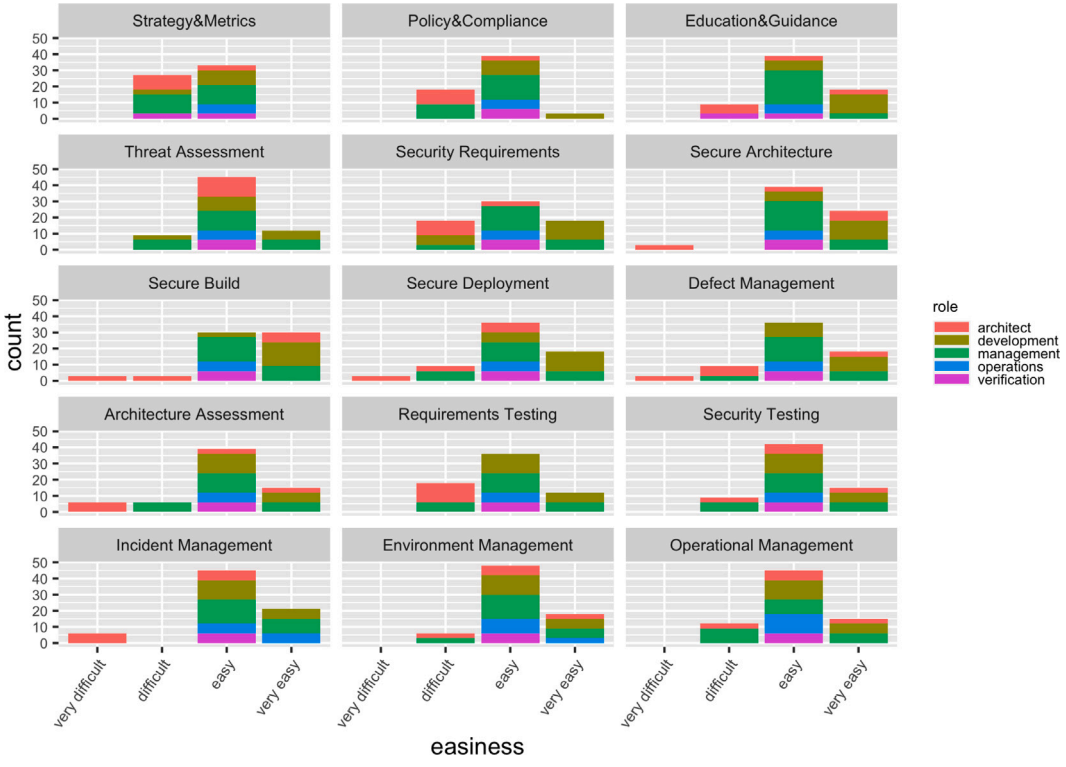


Fig. 6. *Easiness* for the answers about each SAMM practices by different roles.

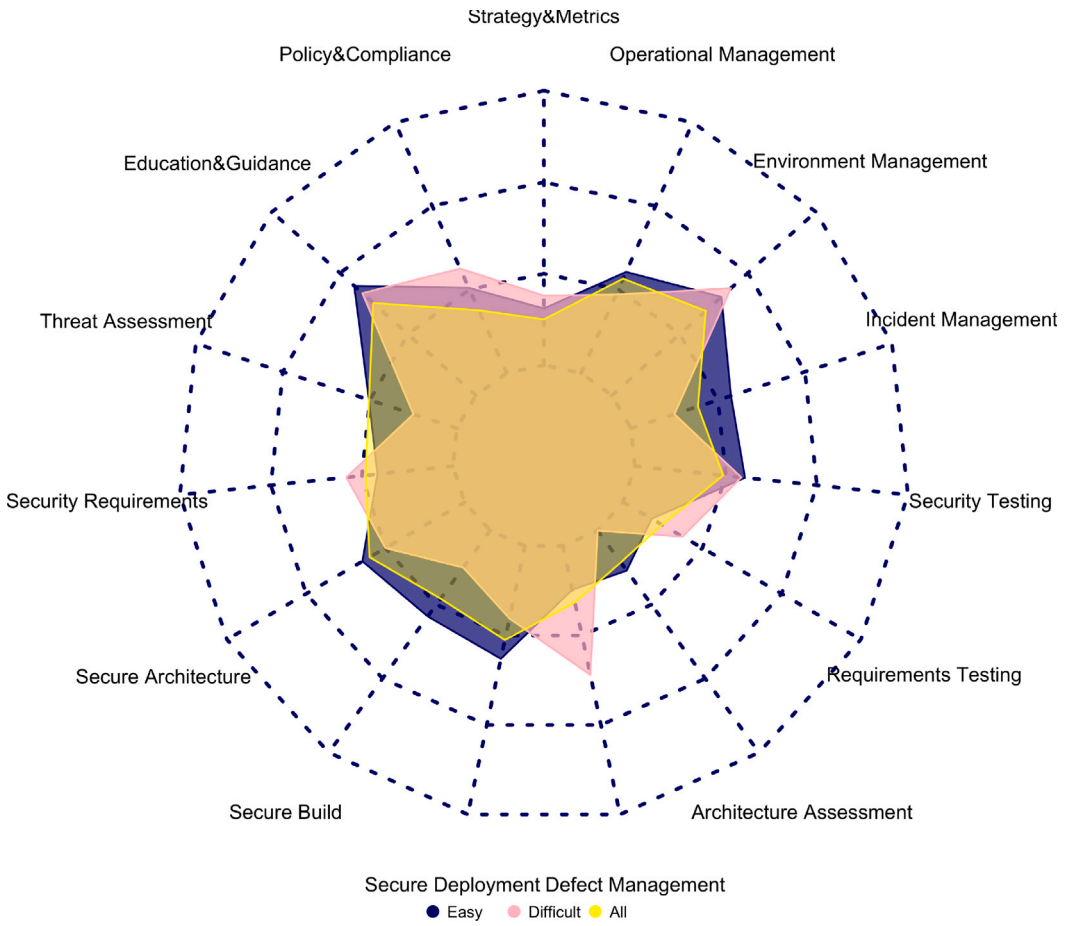


Fig. 7. Overlap between the overall SAMM questionnaire results and results considering *easy* and *difficult* answers. *Easy* are answers marked as *easy* or *very easy*, *Difficult* are answers marked as *difficult* or *very difficult*.

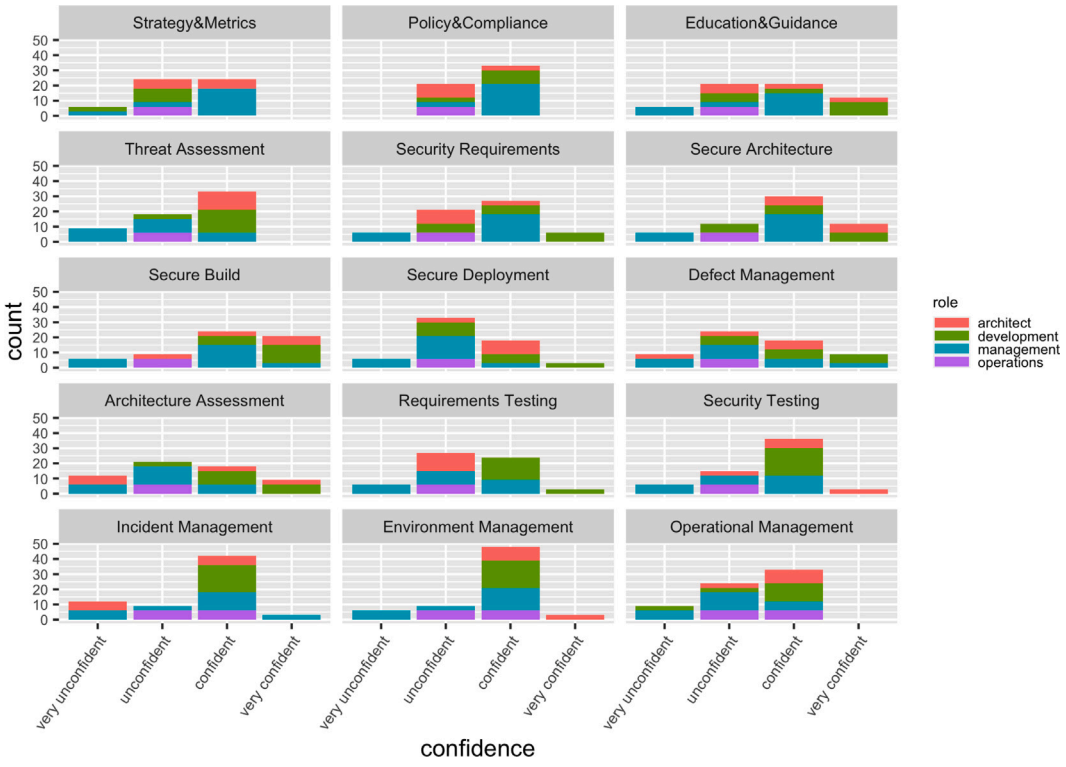


Fig. 8. *Confidence* for the answers about each SAMM practice by different roles.

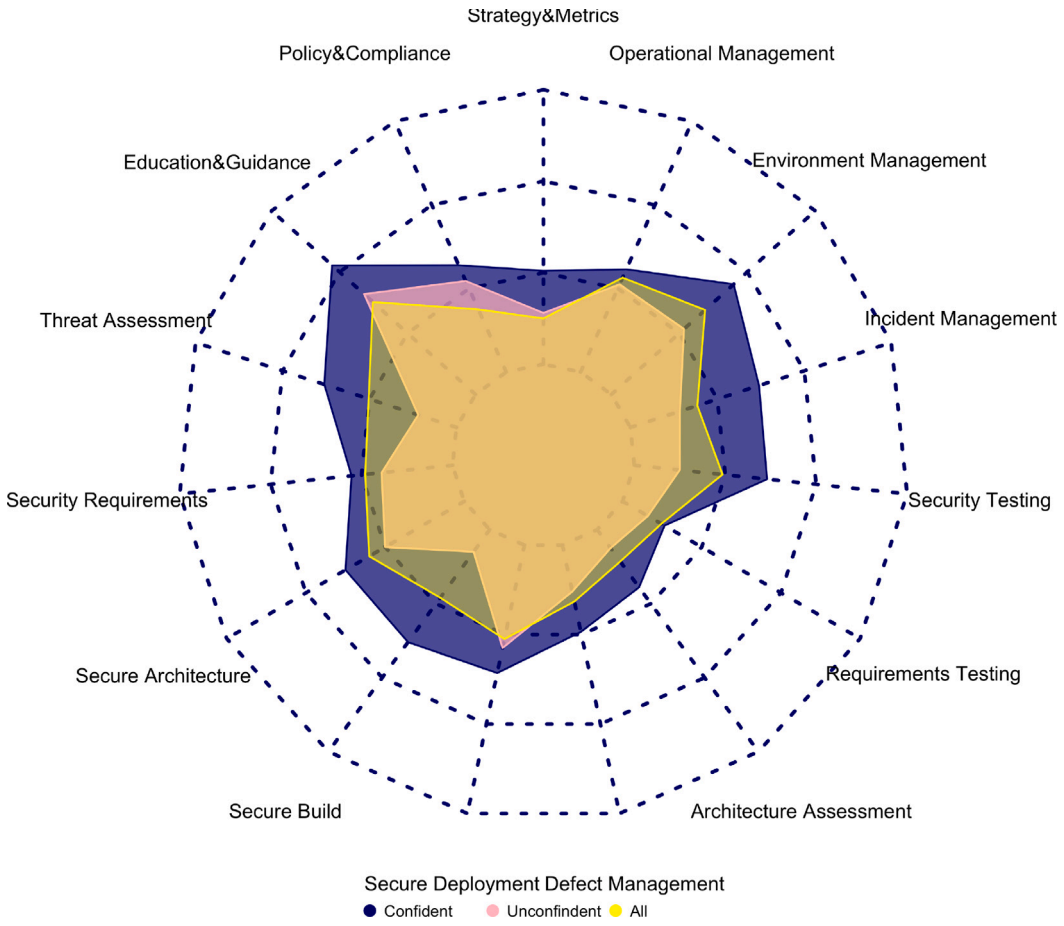


Fig. 9. Overlap between the overall SAMM questionnaire results and results considering *confident* and *unconfident* answers. *Confident* are answers marked as *confident* or *very confident*, *Unconfident* are answers marked as *unconfident* or *very unconfident*.

Lesson learned 1 — Industry

A SAMM assessment involving different stakeholders provides insights that can support decision-makers in establishing more accurate improvement plans.

At COMPANY, the verification team performs penetration testing or manual test, as well as sophisticated test automation for most of the applications as part of their reported *Security Testing* practices.[14] However, developers do not seem to be aware of this, which may result in duplicate work (e.g., running a static code analyzer at the development and verification phase), or friction in bug-fixing—e.g., a developer may question the existence of a bug reported by the verification team if they are not aware of the team's high maturity (Breu et al., 2010).

Lesson learned 2 — Industry

The different evaluations between roles can help pinpoint the source of confusion and friction in related areas—e.g., development and verification.

A specific instance of the above observation can reveal issues not strictly related to security assessment but still crucial for the company. In our case, COMPANY recently transitioned towards a DevOps culture and related practices they are striving to implement. Nevertheless, the SAMM assessment survey showed that the operations team underestimates COMPANY maturity compared to others, except in their area of responsibility. The results of the focus group showing a gap in the relevant practices confirmed this observation. This result shows a challenge in collaboration with the operation team, hindering DevOps.

Lesson learned 3 — Industry

Companies planning to follow Dev(Sec)Ops, can use the SAMM evaluation to check whether a silo exists between operations and other teams.

Although we showed some differences between different ways of executing a SAMM assessment, a lightweight approach closely matches the results of a more laborious one, except a few visible gaps. However, the lightweight approach allowed us to pinpoint areas rated with different maturity by different stakeholders. In the case of COMPANY, such areas are, for example, *Operational Management* and *Environment Management*. The gap in maturity in these practices shown as a result of the focus group, coupled with the low confidence scores for their related questions, grants COMPANY further scrutiny (e.g., with ad-hoc workshops or focus groups).

Lesson learned 4 — Industry

The SAMM assessment is comprehensive but costly. A lightweight approach can provide a valuable assessment, and point out areas where more involved investigations are necessary.

[14] https://owaspsamm.org/model/verification/security-testing/

The lightweight assessment relies on offline answers to the questionnaire. Therefore, there is no control over the respondent's perception—e.g., whether they understood the questions and had enough knowledge to answer them. We introduced two meta-questions to deal with this shortcoming and obtain answers which can be contextualized across two dimensions—i.e., confidence and easiness. Interestingly, at COMPANY the questions that were easy-to-answer and were answered with more confidence showed a higher level of maturity than the rest. Such answers can possibly yield a more precise assessment as they contain less uncertainties about the current state of affairs within the company.

Lesson learned 5 — Industry

Simple questions related to confidence and easiness can be used to weigh the SAMM assessment. Unconfident and difficult answers can be filtered out to provide a more reliable assessment upon which an improvement plan can be developed.

The previous observation does not make unconfident and difficult answers useless. Within COMPANY, they used these additional assessments to reflect on the SAMM instrument itself and on the culture around security in the company.

Lesson learned 6 — Industry

Low confidence scores can be used to select practices that are not well-known and customize activities aimed at spreading awareness about it to individual roles.

Lesson learned 7 — Industry&Academia

Low easiness scores can be used to identify issues in the terminology in the SAMM assessment instrument and fine-tune it to the needs of a specific company.

A SAMM assessment starts with a preparation phase which involves defining scope and expectations, identifying stakeholders, and spreading the word in the organization. Moreover, in our study, we needed to adapt the wording of the questionnaire to COMPANY-specific terminology.

Lesson learned 8 — Academia

Support from stakeholders in charge of security is fundamental for the preparation phase of the SAMM assessment.

Our experience running the assessment with different approaches gave us a good understanding of where knowledge in the different security areas lies in the organization. This is essential information when conducting further studies—i.e., when sampling participants. Moreover, the results of a general SAMM assessment can be the launch-pad for a more specific security assessment depending on the organization's domain. In the case of organizations in the financial domain, such as COMPANY, parts of the results can, for instance, be utilized to check compliance to some of the PCI-DSS requirements.[15]

Lesson learned 9 — Academia

The SAMM assessment can help define a research agenda in one or more areas.

[15] https://www.pcisecuritystandards.org/standards/

6. Conclusion and future work

OWASP SAMM is a maturity model that can help companies identify their strengths and weaknesses in different security areas and gain insights into improvement areas. In this work, we first performed a lightweight SAMM assessment (i.e., online survey) within a company in the financial sector involving 17 participants in five different roles. We then triangulated the survey result in a second, more heavyweight assessment based on a focus group with seven participants. We show that different roles perceive their maturity differently and that the different assessments show different results but are limited to specific practices. Moreover, we evaluated the perception of the questionnaire tool used to perform the assessment, showing that, for most roles, it was easy to answer and that most roles were confident in their answers.

In the future, we seek to replicate this work, focusing on the feasibility of scaling up the SAMM assessment to a larger part of the organization. One of our aims is to understand why survey and focus group assessments may differ. In future studies, we will fill-in the SAMM assessment *during* the focus group and steer the discussion to address any discrepancies with survey data. Finally, we are interested in mapping the SAMM results to other types of assessments specific to a domain relevant to our partner companies.

CRediT authorship contribution statement

Davide Fucci: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Emil Alégroth:** Conceptualization, Formal analysis, Investigation, Validation, Writing – original draft, Writing – review & editing. **Michael Felderer:** Conceptualization, Formal analysis, Validation, Writing – original draft, Writing – review & editing. **Christoffer Johannesson:** Funding acquisition, Project administration, Resources, Supervision, Validation, Writing – original draft.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Davide Fucci reports financial support was provided by KK-stiftelsen. Emil Alegroth reports financial support was provided by KK-stiftelsen. Michael Felderer reports financial support was provided by KK- stiftelsen.

Data availability

A replication package is available at https://doi.org/10.5281/zenodo.7730021.

Acknowledgment

This work was supported by the KKS foundation through the S.E.R.T. Research Profile project at Blekinge Institute of Technology.

References

Assal, H., Chiasson, S., 2018. Security in the software development lifecycle. In: Fourteenth Symposium on Usable Privacy and Security. SOUPS 2018, pp. 281–296.

Breu, S., Premraj, R., Sillito, J., Zimmermann, T., 2010. Information needs in bug reports: improving cooperation between developers and users. In: Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work. pp. 301–310.

De Win, B., Scandariato, R., Buyens, K., Grégoire, J., Joosen, W., 2009. On the secure software development process: CLASP, SDL and touchpoints compared. Inf. Softw. Technol. 51 (7), 1152–1171. http://dx.doi.org/10.1016/j.infsof.2008.01.010.

Geer, D., 2010. Are companies actually using secure development life cycles? Computer 43 (6), 12–16. http://dx.doi.org/10.1109/mc.2010.159.

Jaatun, M.G., Cruzes, D.S., Bernsmed, K., Tøndel, I.A., Røstad, L., 2015. Software security maturity in public organisations. In: Information Security: 18th International Conference, ISC 2015, Trondheim, Norway, September 9-11, 2015, Proceedings 18. Springer, pp. 120–138.

Kassou, M., Kjiri, L., 2012. SOASMM: A novel service oriented architecture security maturity model. In: 2012 International Conference on Multimedia Computing and Systems, vol. 1, pp. 912–918. http://dx.doi.org/10.1109/icmcs.2012.6320279.

Lima, M.V.M., Lima, R.M.F., Lins, F.A.A., 2017. A multi-perspective methodology for evaluating the security maturity of data centers. In: 2017 IEEE International Conference on Systems, Man, and Cybernetics. SMC, http://dx.doi.org/10.1109/smc.2017.8122775.

McGraw, G., 2004. Software security. IEEE Secur. Priv. 2 (2), 80–83.

Ramirez, A., Aiello, A., Lincke, S.J., 2020. A survey and comparison of secure software development standards. In: 2020 13th CMI Conference on Cybersecurity and Privacy (CMI) - Digital Transformation - Potentials and Challenges(51275). pp. 1–6. http://dx.doi.org/10.1109/cmi51275.2020.9322704.

Runeson, P., Höst, M., Rainer, A., Regnell, B., 2012. Case study research in software engineering. In: Guidelines and Examples. John Wiley and Sons, p. 237.

Such, J.M., Gouglidis, A., Knowles, W., Misra, G., Rashid, A., 2016. Information assurance techniques: Perceived cost effectiveness. Comput. Secur. 60, 117–133. http://dx.doi.org/10.1016/j.cose.2016.03.009, URL: https://www.sciencedirect.com/science/article/pii/S0167404816300311.

Tashi, I., Ghernaouti-Helie, S., 2009. A security management assurance model to holistically assess the information security posture. In: 2009 International Conference on Availability, Reliability and Security, vol. 1, pp. 756–761. http://dx.doi.org/10.1109/ares.2009.28.

Teodoro, N., Serrao, C., 2011. Web application security: Improving critical web-based applications quality through in-depth security analysis. In: International Conference on Information Society. i-Society 2011, pp. 457–462. http://dx.doi.org/10.1109/i-society18435.2011.5978496.

Weir, C., Migues, S., Ware, M., Williams, L., 2021. Infiltrating security into development: Exploring the world's largest software security study. In: Proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering. pp. 1326–1336. http://dx.doi.org/10.1145/3468264.3473926.

Weir, C., Migues, S., Williams, L., 2022. Exploring the shift in security responsibility. IEEE Secur. Priv. 20 (6), 8–17. http://dx.doi.org/10.1109/MSEC.2022.3150238.

Wen, S.-F., 2017. Software security in open source development: A systematic literature review. In: 2017 21st Conference of Open Innovations Association. FRUCT, pp. 364–373. http://dx.doi.org/10.23919/FRUCT.2017.8250205.

Davide Fucci is an Assistant Professor at Blekinge Institute of Technology (Sweden). His research interests lie in data-drive requirements engineering, test automation, security testing, and human aspects of software development. His research is performed in close collaboration with industry. Dr. Fucci has published over 60 articles in international journals and conferences and received several best paper awards. He has served on the program committees of over 20 academic conferences, on the editorial or review boards of several top-tier software engineering journals. He is a member of ACM, ACM SIGSOFT, IEEE and IEEE Computer Society.

Emil Alégroth is a docent at Blekinge Institute of Technology. His research has been focused on automated testing, in particular GUI test automation, with several impactful publications in the area. Emil's research has been primarily empirical in nature, conducted in co-production with industry with companies such as Saab, Ericsson and Spotify. He has also operated for several years as the CEO of a company developing test solutions for industry and co-founded companies in the domain of software testing.

Michael Felderer is the Director of the Institute for Software Technology at the German Aerospace Center (DLR), a full professor at the University of Cologne, Germany and an associate professor at the University of Innsbruck, Austria. He holds a Ph.D. and a habilitation degree in computer science. His research interests include software quality, security engineering, software architectures as well as software engineering and system engineering for AI, digital twins and quantum computing. Prof. Felderer is an internationally well-recognized researcher in software and systems engineering. He has published over 200 papers and received more than 10 best paper awards.

Christoffer Johannesson is a principal security master providing input in the design process regarding thread modeling, risk assessment, vulnerability analysis as well as contributing to the secure coding guidelines and training. In his work, Christoffer leads the security master community, helping to safeguard product security and organize security activities. He hold a Bachelors' degree in Cybersecurity.