USENIX THE ADVANCED COMPUTING SYSTEMS ASSOCIATION

Engaging Company Developers in Security Research Studies: A Comprehensive Literature Review and Quantitative Survey

Raphael Serafini, Stefan Albert Horstmann, and
Alena Naiakshina, *Ruhr University Bochum*

https://www.usenix.org/conference/usenixsecurity24/presentation/serafini

This paper is included in the Proceedings of the
33rd USENIX Security Symposium.

August 14–16, 2024 • Philadelphia, PA, USA

978-1-939133-44-1

Open access to the Proceedings of the
33rd USENIX Security Symposium
is sponsored by USENIX.

Engaging Company Developers in Security Research Studies: A Comprehensive Literature Review and Quantitative Survey

Raphael Serafini
Ruhr University Bochum

Stefan Albert Horstmann
Ruhr University Bochum

Alena Naiakshina
Ruhr University Bochum

Abstract

Previous research demonstrated that company developers excel compared to freelancers and computer science students, with the corporate environment significantly influencing security and privacy behavior. Still, the challenge of recruiting a substantial number of company developers persists, primarily due to a lack of knowledge on how to motivate their participation in empirical research studies. To bridge this gap, we performed a literature review and identified a conspicuous absence of information regarding compensation and study length in the domain of security developer studies. To support researchers struggling with the recruitment of company developers, we conducted an extensive quantitative survey with 340 professionals. Our study revealed that 62.5% of developers prioritize security tasks over software engineering tasks, and 96.5% are willing to participate in security studies. Developers consistently ranked security higher than other barriers and motivators. However, repeat participants perceived security tasks as more challenging than first-time participants despite having 40% more general experience and 50% more security-related experience. Further, we discuss Qualtrics as a potential recruitment channel for engaging company developers, acknowledging various challenges. Based on our findings, we provide recommendations for recruiting a high number of company developers.

1 Introduction

Including professional software developers from the industry in empirical research is imperative for studying factors influencing usable security and privacy within corporate settings. These factors encompass organizational processes, security culture, and communication between security and privacy experts and software developers [7, 8, 25, 39]. However, recruiting professionals for security studies poses a consistent challenge for researchers. This challenge is amplified by several factors, including the time constraints, the high cost of recruitment, the geographical dispersion of potential participants, and the relatively small size of this target population [2–4, 34, 37, 38, 62, 76]. These challenges are particularly pronounced in the context of quantitative studies or studies involving practical tasks such as evaluating software development tools [1, 21, 37, 47, 69].

To address recruitment issues, researchers explored the characteristics of participants, focusing on programming and security expertise, in various online recruitment platforms and samples [33, 69]. One strategy is to enlist computer science (CS) students as proxies or source appropriate participants from crowdsourcing or freelance platforms, which can help to investigate new development approaches [59] or study design parameters like security prompting [44]. However, CS students and freelancers showed a different security behavior in security developer studies compared to professionals employed by organizations [18, 44, 76]. Further issues, such as a lack of programming skills of crowdsourcing or freelance participants, required the usage of screening questions [15, 16, 33, 69].

To address these research challenges, we focused on the hard-to-reach population of professional software developers employed at least part-time by organizations and indicating software development as the main part of their job, henceforth referred to as *company developers* in this work, and explored the underlying reasons for existing recruitment challenges. Researchers lack established guidelines, consensus, and recommendations concerning appropriate compensation levels. Participation in some security studies with professional software developers is uncompensated [3, 12], while others provide monetary compensation [35, 44]. Expectations for monetary compensation may also vary depending on study type, task complexity, or study length. Additionally, study participants often perceive study length as a participation barrier [60]. Despite these challenges, recommendations for determining appropriate study lengths are lacking.

To provide foundational guidelines for participant recruitment, we conducted an extensive literature review of security developer studies, revealing that many studies omit crucial methodological details related to compensation, study length, or participant recruitment. Consequently, researchers lack foundational knowledge to establish study parameters attracting company developers for security studies. However, keeping participants engaged in security studies is an essential researcher concern. To address this research gap, we investigated company developers' willingness to participate in empirical research and examined Qualtrics as a potential recruitment channel. We conducted a survey with 340 professional software developers employed in companies and explored the following research questions:

- **RQ 1:** What factors influence the security study participation of company developers?
- **RQ 2:** Are there differences in attitudes toward study participation between first-time and repeat participants?
- **RQ 3:** Compared to existing recruitment platforms concerning participants' programming and security experience, skills, and knowledge, is Qualtrics suitable for recruiting company developers?

Our research indicated 62.5% of developers favor security over software engineering tasks, with 96.5% willing to engage in security studies. Additionally, developers prioritized security over other barriers and motivators. We found that repeat participants perceived security as more challenging than first-time participants, although they had 40% more general and 50% more security experience than first-timers. Further, Qualtrics might be used as a recruitment channel for engaging company developers by acknowledging several challenges. Based on our findings, we provide study length and compensation guidelines for recruiting a high number of company developers.

2 Related Work

In the following, we discuss existing work on the recruitment of participants for software engineering and IT-security studies and factors affecting their participation.

Recruitment Guidelines: The current body of research on conducting software engineering studies provides researchers with guidance and case studies on conducting specific study tasks such as surveys [42], interviews [27], or experiments [66, 73]. Recruitment advice found in the literature tends to focus on recruiting a reliable and representative sample [55] and providing testimonials on specific issues and pitfalls during participant recruitment [20]. Although guidelines exist on how to recruit professionals for software engineering studies, they primarily focus on methodology [10], recruiting the right participants [56], or establishing cooperation between researchers and companies [58]. While these recommendations are useful for ensuring reliable data collection from recruited participants, they offer limited insights into how to effectively and efficiently motivate professional software developers employed in companies in significant numbers.

Recruitment Platforms: For participant recruitment, researchers can leverage personal and industrial contacts [25, 33, 44], rely on convenience samples such as CS students [9, 37, 47, 68, 69], or use crowdsourcing and freelancer platforms, such as Appen, Clickworker, MTurk, Prolific, Upwork, and Freelancer [16, 18, 24, 26, 33, 46]. However, these recruitment methods come with their own set of challenges, such as unreliable data [18, 51, 65], weak built-in tools [57], or contract limitations [24]. Thus, Kaur et al. [33] conducted a survey study with 706 participants and compared six recruitment platforms for software engineering and security studies. Their analysis was grounded in a literature survey of 59 developer studies in the field of usable security and privacy and revealed significant differences in programming and security experiences across participants from crowdsourcing platforms, Google Play, and universities. They suggested using MTurk for conducting quantitative studies that can tolerate a certain degree of fraudulent participation and use screening questions to detect and reject unsuitable participants. For smaller studies that require less noisy data, the authors suggested using Upwork for participant recruitment, a freelancer platform often used in previous research (e.g., [13, 25, 36, 41]). We based our survey on the work of Kaur et al. [33] to provide a comparable analysis of recruitment platforms and investigated the experiences, skills, and knowledge of software developers recruited with Qualtrics. Tahaei et al. [69] also compared different recruitment platforms, including four crowdsourcing platforms and a CS student mailing list. They found that MTurk participants exhibited the highest level of self-reported programming experience. However, they were also the least successful in solving programming skill screening questions developed by Danilova et al. [16]. No previous work presented recruitment insights into the sample of company developers or Qualtrics as a recruitment platform.

Motivation: Based on previous research on study design, persuasion, and the researchers' experience in conducting surveys, Smith et al. [63] analyzed survey invitation emails for factors associated with the recruitment of professional software developers. They found that most of the invitations did not include many persuasive factors. However, emails that addressed recipients directly contained similarity cues (e.g., belonging to the same company), compliments, scarcity cues (e.g., time limits or a maximum number of participants), rewards, and humor had higher response rates. Similarly, Brandt et al. [11] recommended highlighting the benefits for participants beyond monetary compensation (e.g., learning something new), personalizing communication, streamlining appointments (e.g., using a calendar service), and avoiding technical setups by providing an online environment. Additionally, they suggested appealing to different types of motivation, including fun, drive to produce knowledge, social connection, and self-improvement, to improve study participation rates further [6]. The most related work on the motivational factors of participant recruitment was conducted by Serafini et al. [60], who investigated factors influencing the motivation of professional software developers from companies to participate in security studies. They conducted 30 semi-structured interviews on participants' perceptions of study factors related to study design, recruitment methods, and data collection. The results showed that study length, topic, monetary compensation, and trust were influential factors for participation in developer studies. However, participants were concerned about high effort and low performance in security tasks. Compared to [60], we conducted a quantitative analysis and focused on establishing a foundational guideline for participant recruitment.

3 Systematic Literature Review

To provide foundational guidelines based on recruitment strategies and study design used in previous security developer studies, we conducted a comprehensive literature review and selected papers from the most popular top-tier conferences in the field of Security and Privacy (S&P), Human-computer Interaction (HCI) and Software Engineering (SE) research: (1) Conference on Computer and Communications Security (CCS), (2) Conference on Human Factors in Computing Systems (CHI), (3) IEEE Symposium on Security and Privacy (IEEE S&P), (4) Network and Distributed System Security (NDSS), (5) USENIX Security Symposium (USENIX), (6) International Conference on Software Engineering (ICSE), and (7) Symposium on Usable Privacy and Security (SOUPS).

3.1 Paper Selection Process

The comprehensive literature review was conducted by two researchers R1 and R2. R1 reviewed titles, abstracts, and methodology sections of all papers published at the venues mentioned above and screened them according to the following inclusion criteria: (1) published in the years 2017-2022, (2) presented a study involving participants with programming skills, such as CS students, professional software developers employed in companies, freelancers, and crowdsourcing participants. We focused on recent studies and included CS students, freelancers, and crowdsourcing participants with programming skills, as they were often recruited as a convenience sample. R2 reviewed the resulting list of papers to ensure that all papers met the criteria. Both researchers extracted the following information from the resulting list of papers: year of publication, sample, recruitment channel, study type (online, lab, field), study task (survey, interview, practical), number of participants, compensation, study length, and IT-security focus. Any disagreements were resolved by reading the full paper.

In total, we identified 169 papers. We excluded one paper due to limited methodology (provided only information on the study task), another paper due to complex study design (included three different data collection methods)and four papers due to having different payment levels for the same task. The final list included 163 papers (see Replication Package [61]).

3.2 Systematic Literature Review Results

We identified five types of participants in the 163 papers; Software developers (100 paper), CS students (36 paper), security and IT experts (39 paper), admins (8 paper), and "other" (e.g., hackers and testers) (7 paper). In the following, we discuss the results of our systematic literature review analysis (see Replication Package [61]).

3.2.1 Recruitment Channels

The recruitment channels varied within the 163 papers, with social and regional contacts (75/163) being the most commonly used channels, often combined with unsolicited emailing (55/163). Other frequently used channels were social media (42/163), and university recruitment (36/163), especially for practical tasks that required a substantial number of participants while keeping costs and effort low. The least frequently used channels were snowball sampling (29/163), online forums/blogs (29/163), networking platforms (21/163) (e.g., LinkedIn), freelancer and crowdsourcing platforms (16/163) (e.g., Upwork, Clickworker), and security events (8/163). Additionally, some studies (4/163) used channels such as contacting an NGO to recruit participants. Twelve papers did not report the recruitment channels they used. While freelancer, crowdsourcing platforms and online forums/blogs were among the least used recruitment channels, they were often preferred for quantitative studies with many participants. On average, studies used2.06 (md: 2, σ: 1.2) recruitment channels. Universities, unsolicited email, and social and regional contacts were the most common recruitment channels, where no channel combination was reported.

3.2.2 Study Task and Type

Most papers (125/163) described an online study, usually a survey. While interview (72/163), practical (75/163), field (17/163), and laboratory (37/163) studies were less common than surveys (77/163), they often included additional surveys (e.g., entry or exit survey on demographics). For surveys, the highest number of participants were achieved (μ: 247.53, md: 102, σ: 416.81). Field (μ: 27.62, md: 24, σ: 19.12) and interview (μ: 18.33, md: 17, σ: 9.45) studies recruited the fewest number of participants. We identified 71 papers that focused on IT-security. However, we did not find significant differences in study design and recruitment compared to software engineering studies.

Table 1: Study Parameters.

Study Type	Online
Study Task	Survey
Study Language	English
Study Length	mean: 32.77 min (md: 28.5, σ: 18.85)
Recruitment Channel	Qualtrics Research Panels
Recruitment Duration	2.5 months
Participants	Company Developers (n = 340)
Compensation	$62.15 per participant (or Reward Points)

3.2.3 Study Length

Out of the 163 papers, 101 explicitly reported the study length, with an average of 77.75 (md: 52.5, σ: 93.76) minutes. Survey studies typically lasted about 16.58 (md: 14.4, σ: 17.52) minutes, interviews on average 50.93 (md: 45, σ: 25.65) minutes, and practical tasks had the longest length with 126.35 (md: 60, σ: 137.93) minutes. For practical tasks spanning multiple days or weeks, this information was often missing. 46 survey studies also lacked information on the study length.

3.2.4 Compensation

Out of the 163 papers, 77 did not mention any reward, 21 provided no compensation, 48 mentioned monetary compensation (μ: $60.09, md: $26.04, σ: 87.90), and 17 compensated participants with other incentives such as raffle participation, hardware products, charitable donations, or study credits. Of the 48 papers mentioning monetary compensation, 30 also reported the study length, which yielded a mean hourly wage of $41.07 (md: $31.91, σ: 54.77).

4 Methodology

To develop foundational guidelines for participant recruitment, we designed a survey based on our systematic literature analysis and related work. A Replication Package (RP) can be found in [61].

4.1 Study Parameters

Our literature analysis highlighted the importance of presenting detailed study information transparently and consistently. Thus, Table 1 provides an overview of our study parameters. We conducted a 30-minute online survey lasting 2.5 months (79 days) involving 340 company developers recruited through the Qualtrics service. The survey was conducted in English. The study incurred costs totaling €17,588.75 ($62.15 per participant). The Qualtrics service indicated compensation varied based on the research panel participants, including monetary compensation or points redeemable for services or products, such as video game currency. Qualtrics did not provide us with information on the detailed participant compensation plan.

4.2 Survey Design

Based on the systematic literature analysis, we designed questions concerning recruitment channels, study tasks and length, and the type and amount of compensation commonly used in recent research studies. We asked participants for previous experience with CS studies and explored their preferences and attitudes toward study topics, particularly in the area of IT-security and software engineering. We derived organizational and security-related statements Q3.1-Q3.20 for question Q3 from [8,45,46] (see Section D6 in RP) as well as time-related statements Q6.1-Q6.4 and trust-related statements Q6.22-Q6.25 for question Q6 from [16] and [5,19,23,75] accordingly (see Section D7 in RP). Based on [60], participants were asked about their willingness to participate in different types of studies and the factors that might influence their participation decisions. We further identified studies that lasted multiple days (e.g., [45, 49, 50, 64]) and thus considered longitudinal studies in our survey questions. Since code reviews were explored in the literature as an alternative approach to implementation tasks [17], we examined both in our survey. Finally, we collected developer and organizational backgrounds, coding experience, and participants' demographics. For this, we adopted coding and security experience questions from Kaur et al. [33], and included the SSD-SES scale from [74]. We used the exact wording and order of questions to provide consistency and allow comparison with previous research. The final survey consists of ten sections with 66 questions (see RP [61]).

4.3 Pilot Study

To test our questions' comprehensiveness and determine the survey length, we conducted a pilot study with eight participants involving two CS students, two professional software developers, and four researchers. Based on the pilot study results, we added an option allowing participants to indicate their expected monetary compensation with a familiar currency instead of using USD. Further, due to a mean completion time of 35.66 minutes in the pilot study, we reduced the number of items in our exploratory factor analysis (EFA) sets from 30 each to 26, 26, and 20 for the motivator, barrier, and security sets, respectively (see Section 4.6).

4.4 Recruitment

We used the Qualtrics research panel service to recruit participants for several reasons. First, we recruited a hard-to-reach population of company developers. As company developers, we consider professional software developers employed by organizations, at least part-time, who indicated software development as the main part of their job. We asked for 300 company developers evenly distributed across four major regions: Europe, Asia, Africa, and North America. Qualtrics could not guarantee the recruitment of a high number of participants from other regions, such as Australia. The service reported primarily recruiting participants from traditional, actively managed, double-opt-in research panels; sometimes, they referred to social media.

Second, unlike most crowdsourcing and freelance platforms, Qualtrics managed the sampling and pre-screening of participants according to our specified inclusion criteria. To compare first-time with repeat participants' results, we specified that at least 40% of participants should be first-time and at least 40% repeat participants. We screened out participants if they indicated to work less than part-time in a company or reported software development is not part of their job. Previous research showed that participants claiming to be software developers recruited via Qualtrics failed a simple programming task [18]. Therefore, we additionally used a screening question recommended for the recruitment of participants with programming skills [15] (see Section D3 in RP [61]). With 1,556 of the 2,380 participants screened out, we experienced a high failing rate. In Section 5.3.2, we discuss potential concerns associated with participants of the Qualtrics pool.

Third, we were promised the recruitment time of 8 to 10 days, ensuring a predictable timeline for data collection. However, the recruitment process lasted 2.5 months (79 days). The delay was due to many participants being rejected by our built-in exclusion and quality checks. In addition, we reviewed the data set to check for inconsistencies in our participant's responses. This resulted in Qualtrics struggling to fulfill our quota in time (see Section 4.5). To compensate for the high time discrepancies, Qualtrics recruited 40 additional participants without increasing our compensation cost, resulting in a final recruitment sample of 340 participants.

4.5 Participants

Of the 3,240 participants who started the survey, 672 did not meet our inclusion criteria and 188 were from regions already covered. Of the remaining 2,380 participants, 1,556 failed the programming screening questions, and 247 failed our attention check or provided low-quality answers. Due to inconsistencies in their responses or survey completion time below 10 minutes, we excluded 237 additional participants, resulting in a final sample of 340.

An overview of our 340 participants' demographics can be found in Table 2. The average age of participants was 37 years. 265 (77.94%) were male and 73 were female. Most participants (92.65%) held academic degrees. They had 8.8 years of professional software development experience on average, working 37.9 hours weekly. They also had, on average, 12.8 total years of software development experience and 5.85 years in software security. Most participants worked in companies with 100 to 499 employees, with an average team size of 22. The majority of our participants worked in a security-related field for security-oriented companies. In addition, participants were more likely to be at least partially involved in security tasks than not involved at all. Almost all (98.82%) participants indicated they were at least somewhat comfortable completing our survey in English, with 295 (86.76%) being very comfortable.

4.6 Analysis

Quantitative Analysis: For statistical testing, we considered results with $p < .05$ significant. We used the independent t-test [67] to compare first-time and repeat participants for continuous data (e.g., EFA factor scores) and the Mann-Whitney U test [40] to examine differences in Likert type responses. We report results where the effect size exceeds the required effect size (Cohen's $d \geq 0.36$ with $\beta = 0.95$) of our power analysis. Unless otherwise noted, our results include responses from all 340 participants. Since we performed the Mann-Whitney U test on individual Likert items and the independent t-test on the mean of the overall factors of the EFA, we applied a correction factor of two. A significant right tail led to high standard deviations for questions about study length and compensation expectations. Therefore, we used the MAD (Median Absolute Deviation) method, as recommended in the literature [29], to exclude outliers exceeding 3.5 times the MAD. Using this approach, we removed at most 58 participants to calculate the mean, median, and standard deviation results of monetary compensation and study length expectations. We removed at most 21 participants while analyzing a subset of our participants. However, we did not remove outliers for statistical testing when comparing first-time and repeat participants due to a loss of statistical power. Outliers were also considered for statistical representation in Table 5. A detailed statistical analysis summary can be found in the RP [61].

Qualitative Analysis: Open-ended questions were analyzed by researchers R1 and R2 using inductive coding [70]. After two researchers independently coded answers, they merged their sets of codes. Due to the short and straightforward answers, all the discrepancies could be resolved by discussion. Although the main purpose of qualitative research is to explore a phenomenon in depth, we note how many participants stated specific themes to indicate their frequency and distribution.

Exploratory Factor Analysis: To explore the structure of participants' attitudes and concerns about study participation, we conducted an exploratory factor analysis [22]. Our study met several sample size guidelines for the EFA. A sample of at least 300 participants is considered satisfactory according to the literature [14]. A standard guideline is to have a minimum of 10 cases per item for analysis with $N \geq 300$. Following the recommendation, we retained variables with absolute factor loadings greater than 0.4 [22]. We used Kaiser-Meyer-Olkin (KMO) to ensure sampling adequacy [31, 32] and Bartlet's

Table 2: Demographics of the 340 participants.

Gender	Male: 265 (77.9%), Female: 73 (21.5%)	Non-binary: 1, NA: 1
Age	min = 19, max = 65, σ = 8.42	md = 37, μ = 36.56
Language	English = 130, German = 47, Zulu = 37	French = 25, Hindi 24, Other = 77
Region of Employment	Europe = 105, Asia = 87	North America = 82, Africa = 66
Education	B.A = 154, M.A = 126, A.A = 18	JD/MD = 17, Other = 25
English Comfortability	Very = 295, Somewhat = 41	Neither comfortable nor uncomfortable = 4
Previous Participation	Yes = 153	No = 187
Software Development experience in years	min = 1, max = 44, σ = 7.67	md = 13, μ = 12.76
Software Development work experience [in years]	min = 1, max = 40, σ = 6.70	md = 8, μ = 8.86
Security experience [in years]	min = 0, max = 42, σ = 4.43	md = 5, μ = 5.85
Company size	100-499: 100, 500-999: 78, 1000-4999: 59	20-99: 39, 10000+: 28, 5000-9999: 18, 10-19: 7, 1-9: 6, NA: 5
Number of co-worker	min = 0, max = 95, σ = 21.22	md = 15, μ = 21.75, solo = 18
Work hours	min = 20, max = 70, σ = 6.13	md = 40, μ = 38.61
Field security focus	Yes: 222, No: 23	Partially: 89, NA: 5
Company security focus	Yes: 285, No: 12	Partially: 39, NA: 4
Task security focus	Yes: 212, No: 29	Partially: 95, NA: 4

test for sphericity to measure the correlation between our items. The KMO measure assesses the proportion of variance among variables that might be common variance. It indicates if the dataset is suitable for the Factor Analysis by showing if the variables share enough common variance to provide meaningful factors. Bartlett's test evaluates whether the variables are unrelated and hence unsuitable for structure detection, ensuring the data is not too dispersed to reveal any underlying patterns. For all three sets, Bartlett's test was statistically significant. Therefore, we used the oblimin rotation, a commonly used oblique rotation method for correlated variables [22]. The oblimin rotation allows factors to be correlated, reflecting the complex nature of attitudes and concerns by acknowledging that human behaviors and opinions are often interrelated rather than independent. As suggested by Field [22], we determined the maximum number of factors using Horn's parallel analysis, and considered only meaningful factors for which all variables fit the same theme. Horn's parallel analysis compares the actual data's eigenvalues against those from random data to accurately decide the number of factors to retain by preventing the retention of factors that do not significantly contribute to the understanding of the data structure.

4.7 Limitations

This study has several limitations that must be considered when interpreting the results. First, our literature review might not include all relevant work featuring developer studies. Second, other factors may influence the motivation and intention of company developers to participate in empirical research studies that are not considered in this study. Third, desirability bias might have affected participants' statements on their study preferences. Fourth, Qualtrics did not provide detailed compensation or response rate for participant recruitment. Finally, while we wanted to ensure participants genuinely represented company developers, the high number of drop-outs due to our strict screening procedure might have introduced selection bias. Thus, the results might not fully encompass company developers' diverse perspectives and experiences.

4.8 Ethics

The institutional review board (IRB) of our university approved our project. Participants were provided with a consent form outlining the scope of the study, the data use, and retention policies. We also complied with the General Data Protection Regulation (GDPR). Participants were informed about the practices used to process and store their data and that they could withdraw their data during or after the study without any consequences. We assured participants that we would only evaluate and publish anonymized data and quotes. Participants had to give their consent before completing the survey. They were also asked to download the consent form for their use.

5 Results

In this section, we present our participants' attitudes and expectations concerning different study factors, discuss first-time and repeat participation, and compare Qualtrics to existing recruitment platforms. Data comparison based on region can be found in the RP. The average survey completion time was 32.77 minutes (md: 28.5, σ: 18.85).

5.1 Exploring Factors Influencing Security Study Participation (RQ1)

5.1.1 Recruitment Channel

We asked participants through which recruitment channels they would accept a study invitation. Table 3 provides an overview of our participants' recruitment preferences. Most channels received a high acceptance rate except unsolicited emailing (μ: 2.67), flyers and posters (μ: 3.17). Targeted emailing (μ: 4.01), CS related mailing lists (μ: 4.33), and employer recommendations (μ: 4.48) were the most preferred channels. Interestingly, participants employed in Europe showed the lowest likelihood of responding to most recruitment channels, while participants in Asia showed higher acceptance rates. Par-

Table 3: Average likelihood to accept study invitation via a recruitment channel.
Likert-Items with responses: 1: extremely likely, 5: extremely unlikely

	Referral	Unsolicited Emailing	Targeted Emailing	Social Media	Networking Platforms	Employer
Repeat participants	4.48	2.48	4.18	4.23	4.04	4.57
First-time participants	4.34	2.82	3.87	3.46	4.15	4.41
All	4.40	2.67	4.01	3.81	4.10	4.48
	Workshop	Conference	Flyer and Poster	Online Forum	Study Mailing List	CS Mailing List
Repeat participants	3.41	3.97	3.04	4.17	4.44	4.48
First-time participants	4.16	4.16	3.28	3.79	3.95	4.20
All	3.82	4.07	3.17	3.96	4.17	4.33

Table 4: Proportion of participants willing to participate in studies with different types, tasks and topics.

	Security	Lab	Field	Interview	Coding	Review
Repeat participants	0.987	0.797	0.935	0.850	0.961	0.843
First-time participants	0.947	0.813	0.765	0.872	0.957	0.936
All	0.964	0.841	0.805	0.861	0.958	0.894

ticipants from security-focused companies (μ: 4.53), having a security focus in their current field of activity (μ: 4.61), or working on security-related tasks (μ: 4.59) had higher acceptance rates to recruitment through their employer compared to those without a security focus, with acceptance rates of (μ: 4.18), (μ: 4.24), (μ: 4.30), respectively. Asking about additional recruitment channels, 23 participants mentioned crowdsourcing and survey websites like Indeed [30], Naukri [48], or Mypoints [43]. Nine participants were willing to be contacted by phone, while seven preferred study invitations through social media or networks (e.g., YouTube, WhatsApp, Telegram, Instagram). Five participants explicitly mentioned LinkedIn as an appropriate channel for study recruitment. Less common responses included the participants' workplace, university, friends and family, or other third parties (e.g., "CS Global Network," "professional bodies").

5.1.2 Study Task and Type

Table 4 shows the proportion of participants willing to participate in different types of developer studies. While participants were interested in all study types, most (96.4%) indicated a willingness to participate in security studies. When given the option to choose between security and software engineering studies or indicate no preference, 62.5% specifically favored security studies. Participants in security-focused companies (67.0%), field of activity (75.2%), or who worked on security-related tasks (71.2%) preferred security over software engineering studies, compared to 54.5%, 56.8%, and 47.6% of their non-security counterparts, respectively. Almost all participants (95.8%) were willing to participate in implementation tasks. Field studies were the least preferred (80.5%).

5.1.3 Study Length

We asked participants how much time they would be willing to spend on different study tasks. Table 5 shows our participants' preferred study length. On average, participants were willing to spend 25.61 minutes on surveys (md: 20, σ: 15.77), 31.56 minutes on interviews (md: 30, σ: 20.68), 56.89 minutes on programming (md: 60, σ: 41.86) and 56.61 minutes on code review studies (md: 60, σ: 41.38). Participants from Africa were willing to spend more time on programming or code review studies, while participants from Asia were willing to spend more time on all study tasks except code. Most (81.18%) participants were willing to participate in studies lasting several days, weeks, or months. However, 93.48% of participants expected better hourly compensation than for one-day studies.

5.1.4 Compensation

We asked for participants' expected monetary compensation rates for different study types and tasks using their usual currency. The expected compensation rates for a 15-minute study are presented in Table 6.[1] Participants expected significantly less compensation for online (\$6.50-\$16.32), interview (\$8.76-\$15.64), and survey (\$6.50-\$12.34) studies for 15-minute studies than other study types and tasks. By contrast, implementation (\$13.14-\$20.12), code review (\$16.32-\$24.04), field (\$12.34-\$24.04) and laboratory (\$11.29-\$23.5) studies had the highest compensation rates expectations. Regarding regional differences, Africa, and North America generally had, on average, similar expectations. Europe excepted slightly lower compensation in most cases. By contrast, participants in Asia expected significantly lower compensation for all combinations, always expecting less than a third compared to other regions.

Table 7 shows the expected compensation rates for a 60-minute study. Compensation rates did not increase linearly with study length, with expectations almost always only two times higher for 60-minute studies compared to 15-minute studies. Participants' expected compensation rates were of-

[1] Participants' expected compensation rates were converted to USD by using exchange rates from January 25, 2023

Table 5: Average study length by previous participation in minutes.

	Survey	Interview	Implementation	Code Review
Repeat participants	17.55	21.60***	41.83***	45.51
md / σ	15.0 / 8.05	20.0 / 18.39	30.0 / 36.61	45.0 / 39.97
First-time participants	30.38	37.88***	64.72***	60.49
md / σ	25.0 / 18.29	30.0 / 20.35	60.0 / 42.97	60.0 / 40.81
All	25.61	31.56	56.89	56.61
md / σ	20.0 / 15.77	30 / 20.68	60 / 41.86	60 / 41.38

Table 6: Average expected monetary compensationin USD for a 15-minute study.

	Online Survey	Lab Survey	Field Survey	Online Interview	Lab Interview	Field Interview
Repeat participant	9.62	14.31	14.46	11.47	16.49	16.65
md / σ	10.0 / 5.94	15.0 / 9.09	15.0 / 8.84	10.8 / 6.59	16.2 / 9.96	16.2 / 10.19
First-time participant	2.67	6.39	8.04	4.58	8.62	10.48
md / σ	2.0 / 2.33	5.4 / 5.33	6.0 / 7.26	3.6 / 3.86	6.0 / 7.52	8.1 / 9.58
All	6.5	11.29	12.34	8.76	14.12	15.64
md / σ	5.0 / 6.0	10.0 / 9.6	10.8 / 10.21	6.42 / 7.28	10.8 / 11.76	12.0 / 13.54

	Online Code	Lab Code	Field Code	Online Review	Lab Review	Field Review
Repeat participant	17.66	24.5	24.28	24.06	32.28	31.39
md / σ	16.2 / 11.52	21.6 / 17.74	20.26 / 17.4	20.0 / 18.01	24.65 / 26.54	25.0 / 26.06
First-time participant	7.0	11.41	13.31	7.95	13.08	15.08
md / σ	5.0 / 6.79	7.2 / 10.88	10.8 / 12.01	5.4 / 7.96	8.64 / 12.43	10.8 / 13.72
All	13.14	19.02	20.12	16.32	23.5	24.04
md / σ	10.0 / 11.5	15.0 / 16.28	16.2 / 16.77	10.8 / 15.2	19.99 / 21.8	18.18 / 21.53

Table 7: Average expected monetary compensation in USD for a 60-minute study.

	Online Survey	Lab Survey	Field Survey	Online Interview	Lab Interview	Field Interview
Repeat participant	18.92	27.25	26.7	21.3	29.72	29.28
md / σ	20.0 / 10.53	22.95 / 17.57	24.15 / 16.05	20.8 / 11.15	25.0 / 18.23	26.97 / 17.09
First-time participant	10.83	22.12	26.58	17.01	27.46	32.67
md / σ	8.52 / 9.09	16.2 / 19.54	20.0 / 24.25	12.35 / 14.65	20.8 / 24.68	24.0 / 31.12
All	15.12	24.95	26.88	19.3	28.76	30.98
md / σ	12.0 / 11.32	21.6 / 19.15	21.6 / 21.12	18.0 / 13.41	24.3 / 21.81	25.0 / 24.81

	Online Code	Lab Code	Field Code	Online Review	Lab Review	Field Review
Repeat participant	32.8	42.95	41.77	45.19	58.64	56.82
md / σ	25.0 / 23.31	30.0 / 32.52	30.0 / 31.45	30.0 / 36.77	40.0 / 50.31	40.0 / 47.31
First-time participant	22.94	35.29	39.46	26.84	39.32	44.48
md / σ	15.12 / 20.82	24.5 / 32.91	29.85 / 38.07	18.98 / 25.98	29.85 / 37.49	30.0 / 44.13
All	27.47	38.51	39.93	35.48	48.15	49.58
md / σ	21.6 / 22.27	30.0 / 32.03	30.0 / 33.29	21.6 / 32.13	32.4 / 44.09	32.4 / 44.34

ten based on previous study participation (41.2%), job salary (28.5%), or this study's compensation (25.6%). Regarding compensation alternatives, 282 participants selected Amazon vouchers, 145 chose hardware products, 60 might decide to reject the study compensation, and 56 indicated a preference for non-anonymous charitable donations. Participants suggested no alternative forms for non-monetary compensation but payment channels (e.g., PayPal) or cryptocurrency.

5.1.5 Barriers, Motivators and Security Attitudes

In this section, we present the results of our EFA on three sets of questions: attitudes toward participation motivators, barriers, and IT security. All overall and individual KMO scores were above 0.8 except for variable 20 in the security set, which had a score of 0.5401. Since this variable still exceeds the minimum threshold of 0.5, we retained this variable, indicating that our three sets were suitable for the EFA. We report the individual factor loadings, the mean response of all, repeat and first-time participants, and the distribution of the Likert responses. Tables providing an overview of the resulting factors and their loadings and showing the distribution of the Likert responses for all, first-time and repeat participants can be found in the RP [61].

Motivators: We asked participants 26 questions about their attitudes toward study factors that could be perceived as motivators for study participation. We set the number of factors to five. We dropped five items iteratively due to items loading on factors that did not fit thematically until we obtained a sensible EFA output. The number of factors did not change. All factors had higher agreement rates than in the barrier set. Most participants agreed with the statements at least somewhat (μ: 3.93-4.29) suggesting participants placed more value on motivating than deterring factors. Altruism (μ: 4.14) and transparency (μ: 4.29) were the highest scoring factors, indicating that participants were interested in sharing their knowledge and experiences with the community. Still, they wanted to know how their data would be collected, stored, and processed.

Barriers: We asked participants 26 questions about their attitudes toward study factors that could be perceived as barriers to study participation. We set the number of factors to six. For most factors in this set, responses were equally distributed across the scale with a slight inclination towards agreement as evidenced by most means being close to or above three (μ: 2.81-3.65). Thus, only some participants perceived these factors as significant barriers to participation. Commitment (μ: 3.65), privacy concerns (μ: 3.32) and time-constraints (μ: 3.30) were the highest scoring factors. In contrast to the findings by [60], study participation uncertainty (μ: 2.81) was the lowest scoring factor in our analysis.

IT-Security: We asked participants 20 questions about their attitudes toward IT-security studies. We set the number of factors to five. There was a very high agreement for all factors (μ: 3.92-4.54), indicating that participants feel responsible for security and see a benefit for their company and themselves in improving their security knowledge and skills. Factors regarding security culture (μ: 4.54), responsibility (μ: 4.49) and risk awareness (μ: 4.43) were scored highest. While the agreement for task difficulty (μ: 3.92) was the lowest in this set, it was still very high compared to factors in the other sets.

RQ 1 – Summary

Almost all participants were willing to participate in security studies. Targeted emails and employer recommendations were the most accepted recruitment channels, whereas unsolicited emails were the least preferred. Participants were willing to invest about 30 minutes in surveys and interviews and 60 minutes in implementation and code review studies. The expected compensation was between \$15.12 and \$49.58, on average; however, it varied on the participants' country, type, and task of the study. Altruism and transparency were the highest-scoring motivators for participants. The highest-scoring barriers were commitment, privacy concerns, and time constraints.

5.2 Attitudes of First-Time vs. Repeat Participants in Research (RQ2)

e asked participants for CS-related study experience. 187 (55%) participated for the first time in a research study, while 153 (45%) indicated to have study experience, with an average number of participation of 8.81 (md: 8, σ: 8.91).

5.2.1 Recruitment Channel

First-time and repeat participants showed minor differences in recruitment channel acceptance rates. In particular, repeat participants were more receptive (μ: 4.23) to social media compared to first-time participants (μ: 3.46) (see Table 3).

5.2.2 Study Task and Type

Table 4 displays company developers' willingness to participate in research studies with different tasks and topics considering first-time and repeat participants. While first-time and repeat participants expressed their willingness to participate in studies with a focus on security, field studies were the least preferred study type (80.5%) for repeat participants (76.5%).

5.2.3 Study Length

We compared the results of first-time and repeat participants using the independent t-test. First-time participants were willing to spend more time on all tasks than repeat participants, with statistically significant differences ($p < 0.0005$) for interview (Cohen's $d = 0.42$) and implementation (Cohen's $d = 0.37$) tasks (see Table 5).

5.2.4 Compensation

Repeat participants expected higher compensation for most study types and task combinations for a 15-minute study than first-time participants, especially for code review and online studies (see Table 6). We used the t-test again to compare the results of first-time and repeat participants. However, no statistically significant results exceeded the effect size of our power analysis. For a 60-minute study, repeat participants aligned with first-time participants' expectations for all study combinations except for online studies (see Table 7). For online studies, they based their expectations more on the associated task than the type, resulting in higher expectations for online studies than first-time participants. Repeat participants mainly referenced past study compensation (55.5%), while first-time participants referred to their job salary (39%).

5.2.5 Barriers, Motivators and Security Attitudes

Motivators: Repeat participants placed more importance on motivational factors, as they had higher means for all factors and almost all items than first-time participants (see RP, Table 29). There were also statistical differences between first-time and repeat participants for personalization, self-interest, and personal development, indicating that repeat participants expected to be compensated for participation.

Barriers: For the privacy concerns and confidentiality factors, repeat participants scored statistically higher than first-time participants (see RP, Table 30). Repeat participants might be more critical in the context of privacy due to past experiences, with a significant difference in past negative experiences with data privacy (item 22) supporting this. First-time participants scored statistically higher on the commitment factor, indicating they were more concerned about time and conditional commitment (e.g., useful contribution).

IT-Security: We observed statistically significant differences between first-time and repeat participants for task difficulty and risk awareness. Repeat participants perceived security as a more challenging task, while first-time participants were more worried about their company's risks (see RP, Table 28).

5.2.6 Understanding Participation Motivations

First-Time Participants: We asked first-time participants why they chose to participate in our study. 67 expressed a general interest in the study topic or software development in general. Twenty-six participants wanted to gain new knowledge, while 25 each participated to share their expert opinion or to receive the compensation offered. Twenty-one participants felt this study was connected to their job, and 11 wanted to help. Another 21 participants did not indicate a specific reason beyond receiving an invitation. Participants indicated further reasons for study participation, such as having fun (4), relieving boredom (2), seeking variety (1), or challenging oneself (1). Only 27 (14.4%) first-time participants indicated to be regularly invited to CS studies. We asked first-time participants why they had not participated in the past. The majority (114) indicated they had never received invitations for this type of study. Thirty-one did not provide any reason, while 18 cited a general lack of time as the primary barrier to participation. Other less common reasons included lack of interest (3), studies already at full capacity (2), insufficient compensation (1) or relevance (1), COVID-related restrictions (1), or unclear requirements (1).

Repeat Participants: By contrast, 138 (90%) repeat participants indicated to be regularly invited to CS studies. We asked them for the type of studies they had previously participated in. Tables providing a summary of participants' previous study involvement can be found in the RP [61]. Most studies were conducted online (128) or at the participant's workplace (106). Participants frequently engaged in implementation (120), survey (103), or code review tasks (90). Only 43 participants participated in an interview study. General software development (125) was the most common study topic, followed by security (79) and privacy and data (77). Organization (44) and third-party libraries (39) were the least common study topics. We asked repeat participants if they had any negative experiences participating in CS studies in the past. We received 81 responses. Half of our participants (42) reported no negative experiences. Thirteen participants referred to less professional researchers or settings, such as poor preparation or moderation during group activities. Three participants cited technical problems or a less adequate environment. Others reported problems such as high task demands (3), long study length (2), use of deception (2), or low compensation (1).

RQ 2 – Summary

First-time participants were willing to spend more time on studies than repeat participants. They were also more likely to expect compensation based on their job salary than repeat participants who used previous studies as a baseline. First-time participants prioritized user security and were more aligned with organizational security policies, while repeat participants valued job recognition and perceived greater implementation challenges when working on security tasks.

5.3 Qualtrics for Company Developer Participation in Security Studies (RQ3)

We compared the characteristics of our sample recruited through Qualtrics with those investigated by Kaur et al. [33]. Tables providing an overview of the descriptive statistics can be found in the RP [61].

5.3.1 Sample Comparison

Programming Experience: While comparable to those recruited from platforms like MTurk or Google Play, our Qualtrics sample exhibited higher proficiency in software testing and reverse engineering, with a composition of software developers (72.35%), engineers (27.35%), data science/machine learning specialists (24.41%), and DevOps professionals (20.88%). Our participants had elevated proficiency levels in development areas, particularly in frontend (μ: 4.14) and backend (μ: 4.20) development, and showed substantial experience in software testing (78.24%), networking (40.29%), vulnerability research (38.53%), and reverse engineering (30.88%). In contrast to Kaur et al., our sample was more gender-diverse (21.5% female developers) and reported higher proficiency in the top 15 programming languages.

Security Experience: Our participants were more inclined to work in larger companies with security champions (49.7%) and made security-critical decisions collaboratively (77.4%). They were at least twice as likely to have security certifications, attend security events, engage in CTF contests, disclose vulnerabilities, and submit bug reports compared to Kaur et al.'s samples. Further, they demonstrated extensive experience across security features, with higher usage reported in encryption/decryption (65.29%), authorization/authentication (59.41%), and API keys (58.53%), alongside more frequent application of security problem identification techniques. Finally, on the SSD-SES scale, our participants scored higher in vulnerability identification and mitigation (μ: 35.61), indicating stronger confidence in security-related tasks, with a moderate score in security communication (μ: 24.55), showcasing significant differences from all samples in Kaur et al.'s study.

5.3.2 Qualtrics as a Recruitment Platform for Company Developers

We tested Qualtrics as a promising recruitment platform for company developers. We opted for a comparable sample to [33]. For this, the Qualtrics service needed to conduct feasibility checks with individual panel providers in each country. With a recruitment duration of 2.5 month (79 days), Qualtrics required significantly more time than the promised 8 to 10 days for participant recruitment. However, our literature review suggested similar recruitment timeframes through alternative channels. Past work often recruited fewer participants (n = 45) within a similar time frame of 1.5 months with Upwork [28] or more participants (n = 333) over a longer period of 10 months with Prolific and Upwork [74]. Further, Tahaei et al. [69] recruited 613 participants across various platforms within 51 days.

Additionally, Kaur et al. [33] recommended using screening questions for recruitment on platforms such as MTurk. We also used screening questions recommended by Danilova et al. [15], resulting in 1556 participants being screened out of our study.

With 65.38%, the rate at which developers failed the screeners was notably higher than students' performance in previous studies. For example, in [69], freelancers recruited through Appen, Clickworker, MTurk, and Prolific failed screeners at rates of 100%, 36.8%, 91.6%, and 66.8%, respectively. However, the failure rate for students recruited through mailing lists was 10.8% [69]. This discrepancy raises concerns about the reliability and validity of participants recruited through the Qualtrics platform and suggests potential limitations in the screening process or participants' demographics within the pool. Thus, further investigation into the used screeners, screening criteria, and participant demographics is needed to assess and address potential quality issues associated with Qualtrics participants. The cost of using Qualtrics also needs to be considered, as the platform's per-participant charge of $62.15 for a 30-minute survey was significantly higher compared to alternative platforms like Upwork ($15-$120 per hour [71]), where participants could be compensated more flexibly based on the task complexity and duration.

RQ 3 – Summary

Compared to developer samples explored by Kaur et al. [33], our Qualtrics-recruited sample had higher proficiency in software testing and reverse engineering, focusing more on security-related tasks. Our participants reported greater proficiency in the top 15 programming languages, were more likely to work in teams involving security champions at larger companies, and demonstrated higher engagement in security-related activities and confidence in vulnerability identification and mitigation. Challenges with Qualtrics recruitment included longer-than-promised recruitment durations and a high screening rejection rate.

Table 8: Participant acceptance rates for different study lengths (in minutes).

Percentiles	Survey	Interview	Coding	Code Review
10%	60.00	60.00	120.00	120.00
25%	30.00	59.25	60.00	60.00
33%	30.00	30.00	60.00	60.00
50%	20.00	30.00	60.00	60.00
67%	15.00	20.00	30.00	30.00
75%	15.00	15.00	25.00	30.00
90%	10.00	10.00	15.00	15.00

6 Discussion and Recommendations

Improving the recruitment process of security studies with developers is critical for human-centered research's long-term success. This study serves as an initial step in guiding future research on recruitment strategies and establishing a foundational baseline for study design.

Security Studies: Our research indicated 62.5% of developers favor security over software engineering tasks, with 96.5% willing to engage in security studies. Additionally, developers prioritized security over other barriers and motivators. This preference is crucial for high numbers of participants and needs to be considered, e.g., in study design involving security deception. However, repeat participants perceived security as more challenging than first-time participants, although they had 40% more general and 50% more security experience than first-timers. Our results suggested these concerns might be mitigated by announcing security studies as a learning and development opportunity, aligning with corporate culture, and enhancing company reputation. Participants in security roles working for security companies predominantly preferred recruitment through their employers.

Study Length: Our results suggested that survey studies typically met participant expectations concerning study length, while interviews and practical tasks often exceeded them. First-time participants prioritized time flexibility, suggesting that emphasizing flexibility could attract more participants. Table 8 provides an overview for percentiles (10,

Table 9: Participant acceptance rates based on compensation rates for a 15-minute study.

Percentiles	Online Survey	Lab Survey	Field Survey	Online Interview	Lab Interview	Field Interview
10%	0.66	1.20	1.44	1.08	1.44	2.00
25%	1.20	3.24	4.35	2.70	5.00	5.40
33%	2.16	5.00	5.79	4.32	6.00	8.32
50%	5.00	10.00	10.80	6.42	10.80	12.00
66%	10.00	14.29	15.00	10.80	16.20	18.90
75%	10.00	16.20	16.20	10.80	21.60	21.60
90%	15.00	24.90	25.00	20.00	29.70	32.40
Percentiles	Online Code	Lab Code	Field Code	Online Review	Lab Review	Field Review
10%	1.20	1.81	2.02	1.20	2.00	2.00
25%	3.60	5.40	6.38	4.37	6.00	7.36
33%	5.29	7.95	10.00	6.00	9.44	10.79
50%	10.00	15.00	16.20	10.80	19.99	18.18
66%	16.20	22.59	24.62	17.28	27.00	27.00
75%	20.00	30.00	30.00	24.82	35.77	35.10
90%	30.00	43.20	43.20	40.00	53.68	54.00

Table 10: Participant acceptance rates based on compensation rates for a 60-minute study.

Percentiles	Online Survey	Lab Survey	Field Survey	Online Interview	Lab Interview	Field Interview
10%	2.40	4.02	5.00	3.24	4.80	5.00
25%	5.40	9.66	10.80	8.52	11.34	12.00
33%	8.10	12.00	14.65	11.37	16.08	16.20
50%	12.00	21.60	21.60	18.00	24.30	25.00
66%	20.00	30.00	30.00	24.25	35.00	37.80
75%	21.60	37.80	37.80	27.00	43.20	43.20
90%	30.00	50.00	52.72	36.00	54.00	60.00
Percentiles	Online Code	Lab Code	Field Code	Online Review	Lab Review	Field Review
10%	4.80	5.00	5.64	4.88	5.51	6.00
25%	10.80	14.40	15.06	12.00	17.37	16.20
33%	13.59	19.21	20.00	16.20	20.00	20.31
50%	21.60	30.00	30.00	21.60	32.40	32.40
66%	30.00	43.20	47.52	41.68	50.67	54.00
75%	40.00	54.00	60.00	52.65	73.80	77.80
90%	60.00	86.40	97.20	80.60	108.00	114.48

25, 33, 50, 66, 75, 90) describing how many percentages of our participants would accept a study invitation with the indicated length showing that 90% of participants would take a 15-minute programming task. Detailed percentile tables by region are available in the RP [61]. Participants from Africa and Asia generally showed a higher willingness to engage in longer tasks.

Compensation: Our literature review revealed significant inconsistencies in compensation practices that could introduce a recruitment bias affecting study samples' diversity and representativeness. Interviews and practical tasks showed a wider compensation range, with six out of 13 online interviews paid at least 50% above participant expectations. While a higher payment might affect participants' performance, researchers missed the opportunity to increase their sample size by reallocating research funds. Still, the most extended interview study out of these 13 studies offered the lowest hourly compensation of $20 for 2 hours. Additionally, compensation for online practical tasks often met or exceeded expectations, in contrast to lab-based studies, which tended to compensate less than our participants would have expected. This discrepancy is even more pronounced considering the limited data from field studies, where the only interview that provided an hourly compensation rate also reported the lowest rate. Table 9 and Table 10 provide an overview of compensation expectation percentiles (10, 25, 33, 50, 66, 75, 90) for a 15-minute and 60-minute study. Most participants expected compensation at least equivalent to their country's minimum wage. Repeat participants used previous studies as a baseline for compensation. While participants from Asia expected lower compensation, other regions typically requested higher rates. Additionally, participants expected higher compensation for long-term studies.

Recruitment Channel: Compared to platforms in previous work, where participants often indicated freelancing as their main profession, Qualtrics stated to have access to developers employed in organizations through panels and social media. Since past research found different security behaviors between freelancers and developers employed by organizations [18, 44, 76], we tested Qualtrics as a promising recruitment platform for company developers. We encountered challenges with Qualtrics recruitment due to a higher recruitment duration than promised and a high screening rejection rate that might be significant factors for researchers. However, our target demographic comprises professional software developers employed in companies, presenting a challenge in accessibility compared to freelance developers [33]. Therefore, Qualtrics' ability to recruit company developers in significant numbers might be promising in enhancing the ecological validity of study findings. Further research is required to explore the ecological validity of study results for both sample groups.

7 Call to Action

7.1 Research Transparency

Our literature review showed that many studies omitted factors such as study length and compensation. With only 30 of the 163 papers in our literature review explicitly mentioning hourly compensation, a lack of transparency poses a significant challenge for researchers to base their study design and recruitment decisions on prior research. This might affect the comparability and reproducibility of security studies. Additionally, even if this information is included, locating this "blink-and-you'll-miss-it piece of data" is often challenging. Hence, we advocate for establishing clear guidelines regarding the consistent reporting of such information in research.

To improve the transparency of research, we propose a dedicated section *Study Parameters* (see Section 4.1) supported by a table (see Table 1), including the following study details: (1) study type, (2) study task, (3) study language, (4) study length, (5) recruitment channel, (6) recruitment duration, (7) participants, (8) compensation (type, amount, currency, ethical considerations). Based on the conducted study, this information might be extended. Adopting such a standard procedure might be a critical step to enhance the clarity and efficiency of research design and process of future human-centered security research. Providing this information could further benefit the creation of a study design benchmark relevant to the security community.

7.2 Ethical Study Compensation

Our literature review revealed significant inconsistencies in compensation practices raising ethical concerns. For example, three out of four online surveys compensated participants less than our participants expected. This is particularly concerning considering studies that involve participants from different regions but do not adjust compensation to reflect regional economic realities. Most studies offered compensation rates below the average hourly wage of software developers in higher-income countries such as the U.S. ($59.71 [72]) while exceeding the average hourly compensation (assuming 40hr/week) in lower-income countries such as India ($1.56 to $11.73 [53]) and South Africa ($4.46 to $21.51 [54]). Wages in European countries (e.g., Germany $19.39 to $38.25 [52]) were more in line with the compensation offered in most studies. To mitigate these issues and improve the recruitment of participants for future security developer studies, we advocate the adoption of consistent compensation based on participants' expectations, the local minimum wage, and time requirements reflecting the effort associated with different study tasks and types. By implementing a respectful compensation approach based on these principles, the security community might further foster high-value human-centered research standards respecting the participant pool's global diversity.

8 Conclusion

Recruiting a high number of participants for software engineering and security studies is challenging. Many participants indicated not receiving study invitations regularly, even though they would accept most recruitment channels and strategies to be recruited through. Past studies have often used few recruitment channels, relied heavily on personal and professional contacts, and thus often struggled to reach enough potential participants. Diversifying the recruitment channels used is an essential first step. We investigated the attitudes and expectations of first-time and repeat study participants and provided recommendations on monetary compensation and study length for future developer studies. Software developers were motivated to participate in developer studies if they benefited from participation. Since there are significant differences in expected compensation based on employment region, we advise a careful setup of study setting and compensation level. Future research might investigate the effect of different payment levels on study performance and offering non-anonymous charity donations as an alternative form of compensation. Finally, we call to action to foster a transparency-friendly research environment by consistently reporting essential recruitment details such as channels employed, compensation provided, and study length. This practice would establish a fundamental benchmark for future research endeavors. Future research might also need to investigate how the results will replicate over time by considering the effect of currently popular AI assistants on developer security studies and their implications on compensation and study length.

Acknowledgments

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2092 CASA - 390781972.

References

[1] Yasemin Acar, Michael Backes, Sascha Fahl, Simson Garfinkel, Doowon Kim, Michelle L. Mazurek, and Christian Stransky. Comparing the Usability of Cryptographic APIs. In *2017 IEEE Symposium on Security and Privacy (SP)*, pages 154–171, 2017.

[2] Yasemin Acar, Michael Backes, Sascha Fahl, Doowon Kim, Michelle L. Mazurek, and Christian Stransky. You Get Where You're Looking for: The Impact of Information Sources on Code Security. In *2016 IEEE Symposium on Security and Privacy (SP)*, pages 289–305, 2016.

[3] Yasemin Acar, Sascha Fahl, and Michelle L. Mazurek. You are Not Your Developer, Either: A Research Agenda for Usable Security and Privacy Research Beyond End Users. In *2016 IEEE Cybersecurity Development (SecDev)*, pages 3–8, 2016.

[4] Yasemin Acar, Christian Stransky, Dominik Wermke, Michelle L. Mazurek, and Sascha Fahl. Security Developer Studies with Github Users: Exploring a Convenience Sample. In *Proceedings of the Thirteenth USENIX Conference on Usable Privacy and Security*, SOUPS '17, page 81–95, USA, 2017. USENIX Association.

[5] Bushra A. Alahmadi, Louise Axon, and Ivan Martinovic. 99% False Positives: A Qualitative Study of SOC Analysts' Perspectives on Security Alarms. In *31st USENIX Security Symposium (USENIX Security 22)*, pages 2783–2800, Boston, MA, August 2022. USENIX Association.

[6] Judd Antin and Aaron Shaw. Social Desirability Bias and Self-Reports of Motivation: A Study of Amazon Mechanical Turk in the US and India. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, page 2925–2934, New York, NY, USA, 2012. Association for Computing Machinery.

[7] Hala Assal and Sonia Chiasson. Security in the Software Development Lifecycle. In *Proceedings of the Fourteenth USENIX Conference on Usable Privacy and Security*, SOUPS '18, page 281–296, USA, 2018. USENIX Association.

[8] Hala Assal and Sonia Chiasson. Think Secure from the Beginning: A Survey with Software Developers. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–13, New York, NY, USA, 2019. Association for Computing Machinery.

[9] Sebastian Baltes and Stephan Diehl. Worse Than Spam: Issues In Sampling Software Developers. In *Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, ESEM '16, New York, NY, USA, 2016. Association for Computing Machinery.

[10] Hans Christian Benestad, Erik Arisholm, and Dag Ingar Kondrup Sjøberg. How to Recruit Professionals As Subjects in Software Engineering Experiments. In Hustad E., Munkvold B.E., Rolland K., and Flak L.S., editors, *IRIS (Information Systems Research in Scandinavia), August 6-9, Kristiansand Norway*. Department of Information Systems, Agder University College, August 2005.

[11] Carolin Brandt and Andy Zaidman. Strategies and Challenges in Recruiting Interview Participants for a Qualitative Evaluation. In *International Workshop on Recruiting Participants for Empirical Software Engineering, co-located with the 44th International Conference on Software Engineering (RoPES - ICSE 2022)*. ROPES, May 2022.

[12] Souti Chattopadhyay, Nicholas Nelson, Yenifer Ramirez Gonzalez, Annel Amelia Leon, Rahul Pandita, and Anita Sarma. Latent Patterns in Activities: A Field Study of How Developers Manage Context. In *Proceedings of the 41st International Conference on Software Engineering*, ICSE '19, page 373–383. IEEE Press, 2019.

[13] Yan Chen, Steve Oney, and Walter S. Lasecki. Towards Providing On-Demand Expert Support for Software Developers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, page 3192–3203, New York, NY, USA, 2016. Association for Computing Machinery.

[14] Andrew L. Comrey and Howard B. Lee. *A First Course in Factor Analysis (2nd ed.)*. Psychology Press., 1992.

[15] Anastasia Danilova, Stefan Horstmann, Matthew Smith, and Alena Naiakshina. Testing Time Limits in Screener Questions for Online Surveys with Programmers. In *Proceedings of the 44th International Conference on Software Engineering*, ICSE '22, page 2080–2090, New York, NY, USA, 2022. Association for Computing Machinery.

[16] Anastasia Danilova, Alena Naiakshina, Stefan Horstmann, and Matthew Smith. Do You Really Code? Designing and Evaluating Screening Questions for

Online Surveys with Programmers. In *Proceedings of the 43rd International Conference on Software Engineering*, ICSE '21, page 537–548. IEEE Press, 2021.

[17] Anastasia Danilova, Alena Naiakshina, Anna Rasgauski, and Matthew Smith. Code Reviewing as Methodology for Online Security Studies with Developers - A Case Study with Freelancers on Password Storage. In *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*, pages 397–416. USENIX Association, August 2021.

[18] Anastasia Danilova, Alena Naiakshina, and Matthew Smith. One Size Does Not Fit All: A Grounded Theory and Online Survey Study of Developer Preferences for Security Warning Types. In *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering*, ICSE '20, page 136–148, New York, NY, USA, 2020. Association for Computing Machinery.

[19] Constanze Dietrich, Katharina Krombholz, Kevin Borgolte, and Tobias Fiebig. Investigating System Operators' Perspective on Security Misconfigurations. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, CCS '18, page 1272–1289, New York, NY, USA, 2018. Association for Computing Machinery.

[20] F. Ebert, A. Serebrenik, C. Treude, N. Novielli, and F. Castor. On Recruiting Experienced GitHub Contributors for Interviews and Surveys on Prolific. In *The 1st International Workshop on Recruiting Participants for Empirical Software Engineering (RoPES)*, 2022.

[21] Robert Feldt, Thomas Zimmermann, Gunnar R. Bergersen, Davide Falessi, Andreas Jedlitschka, Natalia Juristo, Jürgen Münch, Markku Oivo, Per Runeson, Martin Shepperd, Dag I. Sjøberg, and Burak Turhan. Four Commentaries on the Use of Students and Professionals in Empirical Software Engineering Experiments. *Empirical Softw. Engg.*, 23(6):3801–3820, dec 2018.

[22] Andy Field. *Discovering Statistics Using IBM SPSS Statistics*. Sage Publications Ltd., 4th edition, 2013.

[23] Alexander Gamero-Garrido, Stefan Savage, Kirill Levchenko, and Alex C. Snoeren. Quantifying the Pressure of Legal Risks on Third-Party Vulnerability Research. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, CCS '17, page 1501–1513, New York, NY, USA, 2017. Association for Computing Machinery.

[24] Marco Gutfleisch, Jan H. Klemmer, Yasemin Acar, Sascha Fahl, and Martina Angela Sasse. Recruiting Software Professionals for Research Studies: Lessons Learned with the Freelancer Platform Upwork. In *International Workshop on Recruiting Participants for Empirical Software Engineering, co-located with the 44th International Conference on Software Engineering (RoPES - ICSE 2022)*. ROPES, May 2022.

[25] Marco Gutfleisch, Jan H. Klemmer, Niklas Busch, Yasemin Acar, M. Angela Sasse, and Sascha Fahl. How Does Usable Security (Not) End Up in Software Products? Results From a Qualitative Interview Study. In *2022 IEEE Symposium on Security and Privacy (SP)*, pages 893–910, 2022.

[26] Joseph Hallett, Nikhil Patnaik, Benjamin Shreeve, and Awais Rashid. "Do this! Do that!, and Nothing will Happen" Do Specifications Lead to Securely Stored Passwords? In *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*, pages 486–498, 2021.

[27] S.E. Hove and B. Anda. Experiences from Conducting Semi-structured Interviews in Empirical Software Engineering Research. In *11th IEEE International Software Metrics Symposium (METRICS'05)*, pages 10 pp.–23, 2005.

[28] Nicolas Huaman, Alexander Krause, Dominik Wermke, Jan H. Klemmer, Christian Stransky, Yasemin Acar, and Sascha Fahl. If You Can't Get Them to the Lab: Evaluating a Virtual Study Environment with Security Information Workers. In *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*, pages 313–330, Boston, MA, August 2022. USENIX Association.

[29] Boris Iglewicz and David C Hoaglin. *Volume 16: how to detect and handle outliers*. Quality Press, 1993.

[30] Indeed. https://indeed.com/. Accessed: 2023-10-16.

[31] Henry F. Kaiser. A second generation little jiffy. *Psychometrika*, 35:401–415, 1970.

[32] Henry F. Kaiser and John Rice. Little Jiffy, Mark Iv. *Educational and Psychological Measurement*, 34(1):111–117, 1974.

[33] Harjot Kaur, Sabrina Amft, Daniel Votipka, Yasemin Acar, and Sascha Fahl. Where to Recruit for Security Development Studies: Comparing Six Software Developer Samples. In *31st USENIX Security Symposium (USENIX Security 22)*, pages 4041–4058, Boston, MA, August 2022. USENIX Association.

[34] Mannat Kaur, Michel van Eeten, Marijn Janssen, Kevin Borgolte, and Tobias Fiebig. Human Factors in Security Research: Lessons Learned from 2008-2018, 2021.

[35] Katja Kevic, Braden M. Walters, Timothy R. Shaffer, Bonita Sharif, David C. Shepherd, and Thomas Fritz. Tracing Software Developers' Eyes and Interactions for Change Tasks. In *Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering*, ESEC/FSE 2015, page 202–213, New York, NY, USA, 2015. Association for Computing Machinery.

[36] Alexander Krause, Jan H. Klemmer, Nicolas Huaman, Dominik Wermke, Yasemin Acar, and Sascha Fahl. Pushed by Accident: A Mixed-Methods Study on Strategies of Handling Secret Information in Source Code Repositories. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 2527–2544, Anaheim, CA, August 2023. USENIX Association.

[37] Katharina Krombholz, Wilfried Mayer, Martin Schmiedecker, and Edgar Weippl. "I Have No Idea What i'm Doing": On the Usability of Deploying HTTPS. In *Proceedings of the 26th USENIX Conference on Security Symposium*, SEC'17, page 1339–1356, USA, 2017. USENIX Association.

[38] Thomas D. LaToza, Gina Venolia, and Robert DeLine. Maintaining Mental Models: A Study of Developer Work Habits. In *Proceedings of the 28th International Conference on Software Engineering*, ICSE '06, page 492–501, New York, NY, USA, 2006. Association for Computing Machinery.

[39] Ze Shi Li, Colin Werner, Neil Ernst, and Daniela Damian. GDPR Compliance in the Context of Continuous Integration, 2020.

[40] Henry B Mann and Donald R Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, pages 50–60, 1947.

[41] Jaron Mink, Harjot Kaur, Juliane Schmüser, Sascha Fahl, and Yasemin Acar. "Security is not my field, I'm a stats guy": A Qualitative Root Cause Analysis of Barriers to Adversarial Machine Learning Defenses in Industry. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 3763–3780, Anaheim, CA, August 2023. USENIX Association.

[42] Jefferson Seide Molléri, Kai Petersen, and Emilia Mendes. An Empirically Evaluated Checklist for Surveys in Software Engineering. *Information and Software Technology*, 119:106240, 2020.

[43] MyPoints. https://www.mypoints.com/. Accessed: 2023-10-16.

[44] Alena Naiakshina, Anastasia Danilova, Eva Gerlitz, and Matthew Smith. On Conducting Security Developer Studies with CS Students: Examining a Password-Storage Study with CS Students, Freelancers, and Company Developers. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–13, New York, NY, USA, 2020. Association for Computing Machinery.

[45] Alena Naiakshina, Anastasia Danilova, Eva Gerlitz, Emanuel von Zezschwitz, and Matthew Smith. "If You Want, I Can Store the Encrypted Password": A Password-Storage Field Study with Freelance Developers. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–12, New York, NY, USA, 2019. Association for Computing Machinery.

[46] Alena Naiakshina, Anastasia Danilova, Christian Tiefenau, Marco Herzog, Sergej Dechand, and Matthew Smith. Why Do Developers Get Password Storage Wrong? A Qualitative Usability Study. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, CCS '17, page 311–328, New York, NY, USA, 2017. Association for Computing Machinery.

[47] Alena Naiakshina, Anastasia Danilova, Christian Tiefenau, and Matthew Smith. Deception Task Design in Developer Password Studies: Exploring a Student Sample. In *Proceedings of the Fourteenth USENIX Conference on Usable Privacy and Security*, SOUPS '18, page 297–313, USA, 2018. USENIX Association.

[48] Naukri. https://www.naukri.com/. Accessed: 2023-10-16.

[49] Timothy Nosco, Jared Ziegler, Zechariah Clark, Davy Marrero, Todd Finkler, Andrew Barbarello, and W. Michael Petullo. The Industrial Age of Hacking. In *29th USENIX Security Symposium (USENIX Security 20)*, pages 1129–1146. USENIX Association, August 2020.

[50] Hernan Palombo, Armin Ziaie Tabari, Daniel Lende, Jay Ligatti, and Xinming Ou. An Ethnographic Understanding of Software (In)Security and a Co-Creation Model to Improve Secure Software Development. In *Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020)*, pages 205–220. USENIX Association, August 2020.

[51] Nikhil Patnaik, Joseph Hallett, Mohammad Tahaei, and Awais Rashid. If You Build It, Will They Come? Developer Recruitment for Security Studies. In *International Workshop on Recruiting Participants for Empirical Software Engineering, co-located with the 44th Inter-*

national Conference on Software Engineering (RoPES - ICSE 2022). ROPES, May 2022.

[52] PayScale. Software Developer Salary in Germany in 2024. https://www.payscale.com/research/DE/Job=Software_Developer/Salary, 2024. Accessed: 2024-02-29.

[53] PayScale. Software Developer Salary in India in 2024. https://www.payscale.com/research/IN/Job=Software_Developer/Salary, 2024. Accessed: 2024-02-29.

[54] PayScale. Software Developer Salary in South Africa in 2024. https://www.payscale.com/research/ZA/Job=Software_Developer/Salary, 2024. Accessed: 2024-02-29.

[55] A. Rainer and C. Wohlin. Recruiting participants and sampling items of interest in field studies of software engineering. In *The 1st International Workshop on Recruiting Participants for Empirical Software Engineering (RoPES)*, 2022.

[56] Austen Rainer and Claes Wohlin. Recruiting credible participants for field studies in software engineering research. *Information and Software Technology*, 151:107002, 2022.

[57] Brittany Reid, Markus Wagner, Marcelo d'Amorim, and Christoph Treude. Software Engineering User Study Recruitment on Prolific: An Experience Report. In *International Workshop on Recruiting Participants for Empirical Software Engineering, co-located with the 44th International Conference on Software Engineering (RoPES - ICSE 2022)*. ROPES, May 2022.

[58] Norsaremah Salleh, Rashina Hoda, Moon Ting Su, Tanjila Kanij, and John Grundy. Recruitment, Engagement and Feedback in Empirical Software Engineering Studies in Industrial Contexts. *Information and Software Technology*, 98:161–172, 2018.

[59] Iflaah Salman, Ayse Tosun Misirli, and Natalia Juristo. Are Students Representatives of Professionals in Software Engineering Experiments? In *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*, volume 1, pages 666–676, 2015.

[60] Raphael Serafini, Marco Gutfleisch, Stefan Albert Horstmann, and Alena Naiakshina. On the Recruitment of Company Developers for Security Studies: Results from a Qualitative Interview Study. In *Nineteenth Symposium on Usable Privacy and Security (SOUPS 2023)*, pages 321–340, Anaheim, CA, August 2023. USENIX Association.

[61] Raphael Serafini, Stefan Albert Horstmann, and Alena Naiakshina. – Replication Package – Engaging Company Developers in Security Research Studies: A Comprehensive Literature Review and Quantitative Survey. https://figshare.com/articles/dataset/Engaging_Company_Developers_in_Security_Research_Studies_A_Comprehensive_Literature_Review_and_Quantitative_Survey_Replication_Package_pdf/25298338, 2024.

[62] Dag I. K. Sjøberg, Bente Anda, Erik Arisholm, Tore Dybå, Magne Jürgensen, Amela Karahasanovic, Espen F. Koren, and Marek Vokác. Conducting Realistic Experiments in Software Engineering. In *Proceedings of the 2002 International Symposium on Empirical Software Engineering*, ISESE '02, page 17, USA, 2002. IEEE Computer Society.

[63] Edward Smith, Robert Loftin, Emerson Murphy-Hill, Christian Bird, and Thomas Zimmermann. Improving Developer Participation Rates in Surveys. In *2013 6th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE)*, pages 89–92, 2013.

[64] Rock Stevens, Faris Bugra Kokulu, Adam Doupé, and Michelle L. Mazurek. Above and beyond: Organizational efforts to complement u.s. digital security compliance mandates. *Proceedings 2022 Network and Distributed System Security Symposium*, 2022.

[65] Kathryn T. Stolee and Sebastian Elbaum. Exploring the Use of Crowdsourcing to Support Empirical Studies in Software Engineering. In *Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*, ESEM '10, New York, NY, USA, 2010. Association for Computing Machinery.

[66] Christian Stransky, Yasemin Acar, Duc Cuong Nguyen, Dominik Wermke, Elissa M. Redmiles, Doowon Kim, Michael Backes, Simson Garfinkel, Michelle L. Mazurek, and Sascha Fahl. Lessons Learned from Using an Online Platform to Conduct Large-Scale, Online Controlled Security Experiments with Software Developers. In *Proceedings of the 10th USENIX Conference on Cyber Security Experimentation and Test*, CSET'17, page 6, USA, 2017. USENIX Association.

[67] Student. The probable error of a mean. *Biometrika*, 6(1):1–25, 1908.

[68] Mohammad Tahaei and Kami Vaniea. Lessons Learned From Recruiting Participants With Programming Skills for Empirical Privacy and Security Studies. In *International Workshop on Recruiting Participants for Empirical Software Engineering, co-located with the 44th Inter-*

national Conference on Software Engineering (RoPES - ICSE 2022). ROPES, May 2022.

[69] Mohammad Tahaei and Kami Vaniea. Recruiting Participants With Programming Skills: A Comparison of Four Crowdsourcing Platforms and a CS Student Mailing List. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery.

[70] David R. Thomas. A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation*, 27(2):237–246, 2006.

[71] How much does it cost to hire a software developer? https://www.upwork.com/hire/software-developers/cost/. Accessed: 2023-10-16.

[72] U.S. Bureau of Labor Statistics. Software Developers, Quality Assurance Analysts, and Testers. https://www.bls.gov/ooh/computer-and-information-technology/software-developers.htm, 2022. Accessed: 2024-02-29.

[73] Sira Vegas, Ôscar Dieste, and Natalia Juristo. Difficulties in Running Experiments in the Software Industry: Experiences from the Trenches. In *2015 IEEE/ACM 3rd International Workshop on Conducting Empirical Studies in Industry*, pages 3–9, 2015.

[74] Daniel Votipka, Desiree Abrokwa, and Michelle L. Mazurek. *Building and Validating a Scale for Secure Software Development Self-Efficacy*, page 1–20. Association for Computing Machinery, New York, NY, USA, 2020.

[75] David Gray Widder, Laura Dabbish, James D. Herbsleb, Alexandra Holloway, and Scott Davidoff. Trust in Collaborative Automation in High Stakes Software Engineering Work: A Case Study at NASA. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.

[76] Khaled Yakdan, Sergej Dechand, Elmar Gerhards-Padilla, and Matthew Smith. Helping Johnny to Analyze Malware: A Usability-Optimized Decompiler and Malware Analysis User Study. In *2016 IEEE Symposium on Security and Privacy (SP)*, pages 158–177, 2016.