Siiri Lassila

CONSIDERING PRIVACY: AN ANALYSIS OF PLAYERS' EXPERIENCES OF A SERIOUS PRIVACY GAME FOR SOFTWARE ENGINEERS

JYVÄSKYLÄN YLIOPISTO
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA
2023

ABSTRACT

Lassila, Siiri
Considering Privacy: An Analysis of Players' Experiences of a Serious Privacy Game for Software Engineers
Jyväskylä: University of Jyväskylä, 2023, 70 pp.
Information systems science, Master's Thesis
Supervisors: Mikkonen, Tommi; Sarrala, Tuisku

There exists a gap between privacy as a concept and how privacy is executed in software. Software engineers often consider privacy a secondary concern and they do not have the necessary knowledge to properly build software with privacy in mind. Privacy is not dead!-game developed by Tuisku Sarrala aims to teach engineers needed mental tools to develop their privacy thinking. This thesis investigated the experiences of the software engineering students that participated in the first trial of the game. The students either played the experimental game or a control version. The results from this trial uncovered areas of improvement and that the control group actually found the game more playable and useful than the experimental-game group. Further research is needed to get a holistic view of the game's value as an educational and practical tool.

Keywords: privacy, gamification, serious game, educational games

TIIVISTELMÄ

Lassila, Siiri
Considering Privacy: An Analysis of Players' Experiences of a Serious Privacy Game for Software Engineers
Jyväskylä: Jyväskylän yliopisto, 2023, 70 s.
Tietojärjestelmätiede, pro-gradu tutkielma
Ohjaajat: Mikkonen, Tommi; Sarrala, Tuisku

Yksityisyyden käsitteen ja sen välillä, miten yksityisyys toteutetaan ohjelmistoissa, on kuilu. Ohjelmistokehittäjät pitävät yksityisyyttä usein toissijaisena huolenaiheena, eikä heillä ole tarvittavaa tietoa rakentaakseen ohjelmistoja kunnolla yksityisyyden näkökulmasta. Tuisku Sarralan kehittämä Privacy is not dead!-pelin tavoitteena on opettaa insinööreille tarvittavia työkaluja yksityisyysajattelunsa kehittämiseen. Tässä pro-gradu tutkielmassa tutkittiin työkalun ensimmäiseen kokeilukierrokseen osallistuneiden ohjelmistosuunnittelijoiden kokemuksia. Osallistujat pelasivat joko koko peliä tai kontrolliversiota. Tämän tutkielman tulokset paljastivat parannuskohteita pelissä ja että lumeryhmä piti peliä pelattavampana ja hyödyllisempänä kuin toinen ryhmä. Lisätutkimusta tarvitaan, jotta saadaan kokonaisvaltainen käsitys pelin arvosta opetusvälineenä ja käytännöllisenä työkaluna.

Asiasanat: yksityisyys, pelillistäminen, opettavainen peli

GRAPHS

TABLES

FIGURES

CONTENT

1 INTRODUCTION

Privacy issues are important to consider for any software engineer for ethical, economic, legal, and reputational reasons. Failure to protect sensitive and private data could potentially affect the company's reputation and revenue as well as lead to legal consequences for breaching privacy laws and regulations. Although differing by region, data privacy laws affect all firms, and an especially strident policy is the General Data Protection Regulation (GDPR) in Europe that imposes heavy fines (up to 20 million euros or 4% of the global turnover of the preceding fiscal year, whichever is higher if the violation is severe, 10 million or 2% for mild violations) on companies that are in violation of its standards (GDPR, 2021a). A particularly grievous recent example of privacy failure would be the Facebook-Cambridge Analytica scandal for which Facebook was eventually fined 5 billion dollars by the Federal Trade Commission (FTC) which is the largest fine ever imposed on a company for a privacy violation (FTC, 2019). A company as huge as Facebook can perhaps afford to pay the fine and endure the ensuing reputational damage but for a smaller company such a loss could be crippling. Therefore, it is essential that when developing software engineers consider different privacy concerns, both for the company and for the users whose personal data is being collected. However, research has found that engineers often consider privacy to be a secondary concern (Senarath et al., 2019; Spiekermann & Cranor, 2009) and that when it comes to choosing between respecting privacy and fulfilling system functionality requirements, engineers favour the functionality (Senarath & Arachchilage, 2018). The engineers perceive there to be a lack of compatibility between engineering tools and privacy practices (Senarath et al., 2019) and this gap does actually exist (Kostova et al. 2020). All this leads to a gap between privacy as a concept and the tools used to implement it, where mental tools for privacy thinking should be.

The gamified method of a serious privacy game developed by Tuisku Sarrala (Sarrala, 2022) could be one possible way to improve the mental tools needed for privacy thinking by evoking engineers’ privacy thinking and helping them build a mental model of privacy. Gamification is the process of imbuing gameful elements into a non-game context in order to motivate users to perform certain tasks (Huotari & Hamari, 2012; Detering et al 2011; Hamari et al., 2014). Serious games are virtual games that have a useful purpose beyond entertainment, often, the purpose of learning and education (Girard et al., 2012). Past research has also shown that gamification has good potential for improving learning and knowledge retention. (Kapp, 2012; Girard et al., 2012; Hamari et al., 2014).

The experimental game developed by Sarrala is based on the idea of serious games and scenario development, which is a problem-solving tool for complex ideas, that is based on imagining different futures (Schoemaker, 1993). This thesis is a case study of the first trial experience of software engineering students playing the game developed by Sarrala, as well as a control group that plays a different version of the game.

The research questions this thesis attempts to answer are as follows:

RQ01 Do the players find the game effective?

RQ02 How was the players' experience of the game?

RQ03 What differences are there between the groups that played the experimental game vs the group that played the control game?

RQ04 Does the game affect how much interest the player has in the topic of privacy?

The thesis has one hypothesis that goes as follows:

HP01 the group that played the experimental game found the game more effective than the control group

This thesis studies if this gamified method of developing developer's privacy thinking would be a viable option for software engineers to use when developing software, so privacy concerns would be taken into account and addressed more holistically, or if further development of the game is needed. By evaluating the players' experiences and feedback on this tool, the results should uncover potential areas of improvement and further development.

Chapter Two provides the necessary background for the thesis. It first talks about the topic of privacy, how it is defined and how it is engineered, then about scenario development, which is an aspect of the privacy game, and gamification, what it is, and how it can benefit learning. Finally, Chapter Two introduces the serious game developed by Tuisku Sarrala. Chapter Three is on the research design; it lays out the research method, data collection, analysis, and research setting. Chapter Four lays out the results of the data analysis. Chapter Five discusses the answers to research questions derived from the results and discusses the implications for research and practice. Chapter Six is the conclusion; a summary of the study, a discussion on the limitations of this study, and suggestions for future research opportunities are included in this chapter.

2 CONCEPTUAL BACKGROUND

This chapter lays out the conceptual background for this thesis by describing and delving into the concepts and theories that the thesis is built upon.

2.1 Privacy design

Privacy has been a topic of debate for a long time, but it is recently, with the advent of the information era, that it has risen to new importance. Privacy has been described as the right to be left alone, protection of physical self and injury of feeling from the disclosure of personal facts, as well as the exercise of selective control to self with awareness of the consequences of exercising that control. (Spiekermann & Cranor, 2009). With the ubiquitous nature of data and its collection in today's world, threats to someone's privacy have transformed from a limited scope of traceable human sources to a much larger threat of unwanted use of personal information by companies or third parties that legally or illegally gain access to it (Spiekermann & Cranor, 2009).

Protecting the privacy of individuals whose data is being collected by different systems has given rise to several different principles and methods e.g., Colesky et al. (2016) and Kalloniatis et al. (2008). Perhaps one of the most prevalent ones is the concept of Privacy by Design, where the guiding principle is that privacy is embedded in the design of the software itself. Privacy by Design (PbD) is included also in data regulations such as GDPR (GDPR. 2021b) and FTC guidelines (Gürses et al., 2011). However, Privacy by Design is vague as a principle and leaves lots of questions unanswered and open to interpretation on how it should be implemented in practice (GDPR, 2021b; Gürses et al., 2011). Following it requires specific engineering expertise, as well as contextual analysis, and the balancing of multilateral security and privacy interests, as well as the interests of the company and what functional requirements there are for the software. (Gürses et al., 2011). It is a balancing act between respecting privacy and fulfilling the system functionality requirements, which often leads to engineers favouring the functionality of the system over the issues of privacy. (Senarath & Arachchilage, 2018).

This problem is amplified by the fact that engineers do not have enough education and knowledge of practical privacy design principles and how to incorporate privacy requirements into technological practices, so they end up discarding them altogether. (Senarath & Arachchilage, 2018). The felt lack of usefulness, perceived lack of compatibility with adopted engineering practices, and the lack of seeing the results of the methods implemented lead to low intention on behalf of engineers to use privacy engineering methods (Senarath et al., 2019). In fact, PbD has yet to become a common practice (Senarath & Arachchilage, 2018). Past surveys have also indicated that often engineers and/or developers consider privacy to be a secondary concern or "not their problem" (Senarath et al., 2019; Spiekermann & Cranor, 2009) which is, of course, alarming when talking of the people who develop the software and systems we are entrusting our personal data to. It would therefore be important to motivate and support engineers to consider and implement privacy aspects when developing software.

Another problem that privacy engineering faces is that the theory and practices do not always align. Current-day software development relies on agile methods and the use of service architectures. Kostova et al. (2020) found that privacy researchers systematically failed to consider the challenges this poses to actually implementing their proposed privacy solutions. This is because privacy researchers treated software engineering as a "black box" failing to consider the reality of the agile environment and activities of modern-day software development, which then causes significant challenges to the adequacy, feasibility, and potential deployment of privacy solutions that left software engineers unable to close the gap caused by the misalignment. Collaboration between these fields is therefore vital, so the methods developed actually fit the practice. However, it is a fact that current development practices that rely on data are ill-fit for privacy design and vice versa. Current service architectures mean privacy risks are abound and can in fact be the source of privacy problems, and privacy design limits the way the advantages of service ecosystems can be leveraged. (Kostova et al. 2020). This poses a difficult question, and clearly, something needs to change within both fields to correct this misalignment.

Contemporary software systems are increasingly open, existing in a hyperconnected setting that collects, processes, and disseminates massive data amounts daily. The amount of data being generated daily keeps growing, especially with technologies being embedded into objects that previously did not have any, creating the Internet of Things (e.g. smart refrigerators). This means that privacy management must extend to an ultra-large scale. Privacy design has been driven by bound systems, where the scope of privacy consideration is limited to the system in question, instead of the open network of systems the software architecture nowadays creates. The ever-evolving nature of this environment makes it difficult for software engineers to anticipate emergent and unknown privacy threats. (Anthonysamy et al. 2017)

The problem with many of the methods that exist for privacy engineering is that they are reductionist, i.e. they reduce this complex issue into simple or fundamental components. They are largely based on analysing and depicting the system's essential components and data flows. They do not fully take into account human- and ethical-centred demands, have narrow views of what privacy means, or only consider how to comply with GDPR requirements. Threats that arise from the system's environment are left unconsidered as only the system in question is considered, outside of the context of the interconnected environment where these systems exist. These models also largely require engineers to have existing privacy thinking skills. (Sarrala, 2021)

As discussed, there exists a gap between how engineers consider privacy and privacy requirements. This gap is potentially caused by the fact that privacy as a concept is well known and widely used and that there exist many tools with which privacy can be implemented on software, but the engineers do not have the needed mental privacy thinking skills to properly develop software with privacy in mind. This perceived gap led Tuisku Sarrala (2022) to develop a serious game for software engineers that aims to address it by evoking their privacy thinking and helping them to build a mental model of privacy. (Sarrala, 2022).

2.2 Scenario development

Scenario development is a problem-solving tool for breaking down a complex phenomenon into more analysable parts. It helps to expand people's thinking and to examine uncertainties related to different futures. (Schoemaker, 1993). Schoemaker (1993) defines scenarios as

> "…focused descriptions of fundamentally different futures presented in a coherent script-like or narrative fashion."

Scenario development is concerned with creating stories about possible futures. But not all descriptions of alternative futures are scenarios, only stories. There exist multiple categories of scenario techniques, resulting in two dozen techniques overall with different strengths and weaknesses. (Bishop et.al. 2007). Scenarios aim to reflect a variety of viewpoints to cover as many future possibilities as possible. Sometimes, depending on the size of the problem being examined at hand, only a few scenarios are needed to sufficiently examine the issue, other times, numerous scenarios are needed. (Schoemaker, 1993). Scenario development also helps to counter any psychological biases people may have. Scenarios accommodate comprehension by weaving intentional and causal accounts around strands of otherwise disparate and hard-to-remember pieces of evidence. This mode of thinking also accommodates the integration of new evidence and further inquiry. People, according to Schoemaker (1993) seem to best relate to these concrete and causally coherent narratives. The decomposition of complex problems into more understandable states also aids the human mind, which can only deal with a limited amount of complexity. (Schoemaker, 1993).

The scenario development method as described by Schoemsaker (1997) goes as follows: it is a multi-stage process that involves identifying the problem at hand, whom it affects, and what trends or other elements, such as uncertainties, will affect the variables involved. After identifying the elements involved, the construction of scenarios can begin. The scenarios are then assessed for plausibility and internal consistency, and how stakeholders would behave in them. Scenarios that are deemed implausible or incredible are eliminated and new scenarios are created during the evaluation process, to cover a wide variety of uncertainties and outcomes. (Schoemaker, 1993). However, this is only one of the dozen methods.

The intended benefit of scenarios is the way it expands people's thinking, both individually and collectively, whilst simultaneously focusing it on the issue at hand. Challenged to consider multiple possibilities and futures, the people performing scenario development consider a problem from multiple perspectives, especially when the scenarios are being built by broad organizational input. Scenarios also challenge biases, such as availability bias which leads to people undervaluing things that are hard to imagine or recall from memory. (Schoemaker, 1993).

2.3 Gamification

Gamification is a subject matter that has risen in popularity in recent years in both academic and business circles. (Hamari et al., 2014; Huotari & Hamari, 2012). Detering et al. (2011) define gamification as the use of game design elements in non-game contexts (Detering et al. 2011). Huotari & Hamari (2012) argued that this definition is problematic since it would include other things with some game-like elements like the stock market in its definition. Their definition of gamification, therefore, includes a service marketing perspective and goes as follows:

> "Gamification refers to: a process of enhancing a service with affordances for gameful experiences in order to support the user's overall value creation."

Affordance is defined by Huotari and Hamari (2012) as any quality that contributes to the emergence of gameful experiences. The core aspect of gamification is that a gameful experience emerges from intrinsic motivation on the part of the user and is, therefore, a voluntary experience that cannot be driven by designer attempts to affect the players' decision-making. (Huotari & Hamari, 2012). The aim of gamification is to motivate and engage users to perform certain tasks, and it has been demonstrated to show a positive effect (with the caveat that it depends on the context of the gamified service/experience). (Hamari et al., 2014).

Intrinsic motivation refers to motivation to want to do something because it is inherently interesting or enjoyable, whereas extrinsic motivation is motivation to do something because it leads to a separate desirable outcome. The quality of experience and performance between these two types of motivation can be very different between people who experience one type over the other. For example, intrinsic motivation results in high-quality learning and creativity, which is why it is important in phenomena for learning and education. However, intrinsic motivation cannot always be relied on which is why it is important to foster active and volitional extrinsic motivation in learners for a successful learning experience. (Ryan & Deci, 2000). This applies also when considering what motivational aspects to include during the process of gamification.

Gamification has also been used in educational circles as a new tool for teaching academic and professional skills to both children and adults (Girard et al., 2012). There exist, of course, different types of games, that can take place in a physical reality or a virtual one. The current generation of learners referred to as digital natives or the Net generation by some (Prensky 2001; Bekebrede & Warmelink, 2011), however, have grown up with different digital technologies their entire life so it would seem logical to maintain continuity between the tools used in education and the ones they use in their everyday activities. (Girard et al. 2012). A review by Bekebrede and Warmelink (2011), however, showed no particular differences between learning styles between different generations, but that people generally preferred collaborative and technology-rich learning and deemed games a valuable teaching method (Bekebrede & Warmelink, 2011), so regardless of a learner's age, a gamified approach seems valuable, although it might be easier for those who have grown up and are used to such technologies, as opposed to older generations. Various studies have shown some evidence that computer-assisted learning enhances learning compared to traditional teaching methods such as face-to-face lessons and pencil-paper-based studying. (Girard et al., 2012).

Kapp (2012) complied various meta-analysis studies on the educational value of games in his book and concluded the following: instructional games have beneficial effects when the content is clearly targeted and objectives are clearly defined, instructional games can be effective for gaining knowledge and higher knowledge retention, instructional games should be embedded in instructional programs that include debriefing and feedback to further learner understanding, games yield better attitudes towards learning, instructional support to help learners to understand how to use the game increase the instructional effectiveness of the game, instructional games do not need to be entertaining to be educational, both intrinsic and extrinsic motivational aspects should be included in gamification but extrinsic motivators (i.e. points, rewards) can sometimes undermine intrinsic motivation or learning. (Kapp, 2012, p.101-103). However, despite research showing the seemingly beneficial effects of gamification on learning, more research is needed in the area, and practitioners should not be too excited. (Girard et al., 2012).

Virtual games that are designed to have a useful purpose instead of having primarily entertainment value, e.g., games for enhancing training and education, are referred to as "serious games". (Girard et al., 2012). Serious games can be digital games, simulations, virtual environments, or mixed reality/media of various types (strategy, adventure, etc.) and training various skills (academic, health, etc.), but what makes them apart from virtual games is that they are designed with the utility of purpose from the very first step, instead of that usefulness having been added to the game subsequently (Girard et al. 2012). The privacy game analysed in this study is one such game.

2.4 Privacy is not dead! game

This section presents the experimental serious game of privacy thinking that was developed by Tuisku Sarrala (Sarrala, 2022) for software engineers. Sarrala describes the idea of the game like this:

> "The game is driven by two key ideas: (1) the use of scenarios to deal with the complexity of today's systems and privacy threats, and (2) the use of serious games to develop engineers' privacy thinking skills." (Sarrala, 2022)

The current version of the game was created using PowerPoint which should mean that it is simple to use for the players. The experimental game contains five sets of virtual card decks that can be moved and placed around the virtual playing deck by using the mouse. The version of the game that the control group played had two sets of card decks as well as a "The Software" card. An example of the experimental game in action can be seen in FIGURE 1, and an example of the control group game can be seen in FIGURE 2.

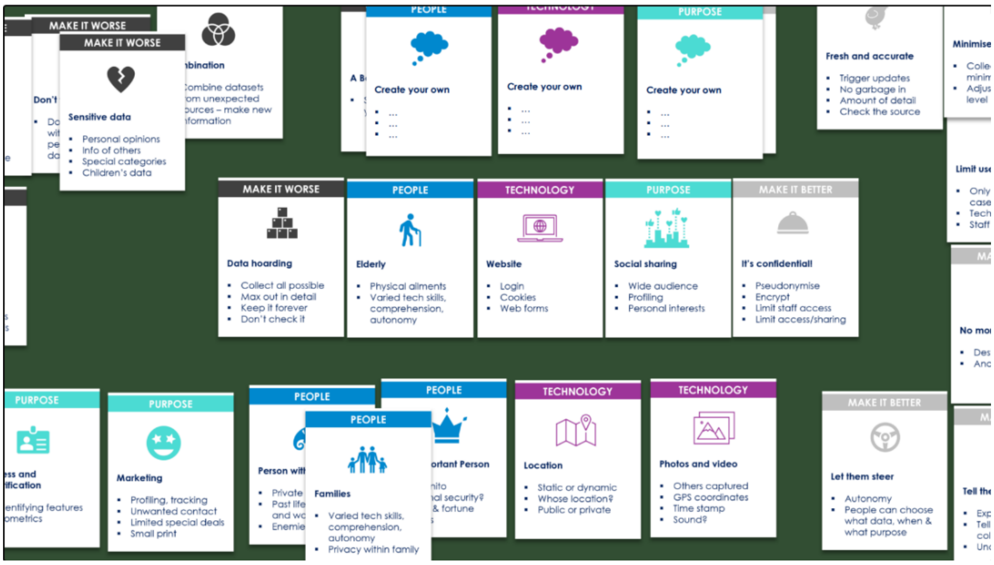


FIGURE 1 The experimental version of the game

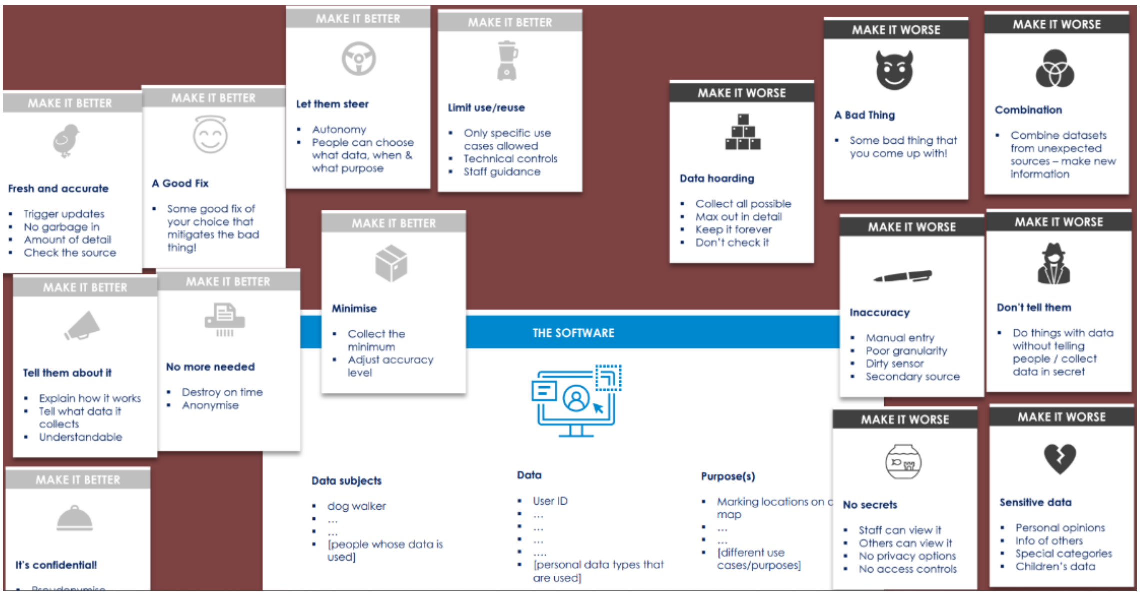
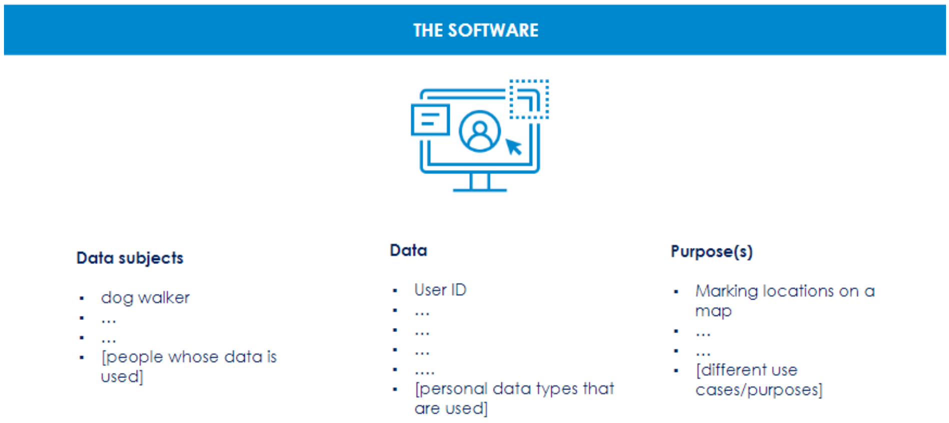


FIGURE 2 The control group version of the game

The sets of cards are as follows; the scenario cards are coloured teal, purple and blue and are used to create a privacy scenario. A scenario is made up of three cards, one of each colour. The teal-coloured cards, which are named "Purpose"-cards, contain different high-level use cases and functionalities for software. The "Technology" cards, which are purple coloured, contain different types of high-level technological solutions used in the software. And finally, the blue cards, called "People", contain different types of people who may be affected by software and privacy issues. During the game, the players select all the scenario cards that may be relevant to their software and develop new scenarios on each round of the game by replacing one or many of the cards. Players also have the chance to place additional scenario cards on the board. As well as scenario cards, there are dark grey and light grey cards, which can be added to the scenario to make the scenario worse in regard to privacy (these cards are dark grey) or to make it better for privacy (these cards are light grey). The content on these grey cards largely aims to follow the privacy principles and anti-principles described in the GDPR (European Union, 2016). These requirements are, for example, the requirement for transparency, minimization, et cetera. The players who play the experimental game use all the different coloured cards, whereas the control group only has the light grey and dark grey cards to use during the game, as well as the Software card. The Software card is shown in FIGURE 3.



THE SOFTWARE

Data subjects

- dog walker
- ...
- ...
- [people whose data is used]

Data

- User ID
- ...
- ...
- ...
- ....
- [personal data types that are used]

Purpose(s)

- Marking locations on a map
- ...
- ...
- [different use cases/purposes]

FIGURE 3 The software card given to control players

Before the game commences, the players are explained that the purpose of the game is to help them to spot and describe privacy problems, explain why they matter, and then think of possible ways to mitigate them. The definition of a privacy problem is given as follows:

> "an event where a person's private life or their rights to their data has not been respected, their personal data has not been protected, their personal data has not been processed in a fair way, or their personal data is processed in a way that causes unjustified negative effects to people."

The players are given a description of the cards they are given and then asked to familiarise themselves with the cards. They are then asked to document any threats they find during the game in a table called "catalogue of privacy problems". Players are also told that the cards are merely meant to give them ideas, and not restrict them, so the players are not necessarily bound to what is written on the cards.

The gameplay goes as follows for the groups that play the experimental game. A privacy scenario is created using the three colourful scenario cards, one of each colour. These cards are then laid out on the game board. The players can use the ready-made scenario cards or play the "Create your own" card to make new scenarios. The players split into two teams: "Baddies" and "Goodies". The Baddies aim to make the privacy of the scenario software as worse as possible by playing "make it worse" scenario cards. The Goodies aim to mitigate the damage caused by the Baddies by playing "make it better" scenario cards that place privacy controls on the software in the scenario. On each round, the Baddies start by playing their card and describe verbally how this would make the scenario worse for privacy. Then the Goodies play their own card to mitigate the damage caused by the Baddies, and verbally describe how their card achieves this in a believable way. A point is then given to the group that won the round by either successful mitigation or successfully making the scenario worse for privacy. The point and the scenario are recorded on the privacy catalogue. Then the players move on to the next round, with the winning team changing the scenario by swapping one of the cards. When the players have had enough and run out of ideas, they count the points and declare which team won the game, the Baddies or the Goodies.

The control group is given a "The Software" card where they are to fill in the missing information of the software from the point of view of privacy at the start of the game. *Data subjects* are the different types of people whose personal data is being used in the software. *Data* represent the types of personal data being used in the software. *Purposes* describe the different ways the data is being used in the software. The Software card contains the scenario being played for the entire game; it is not changed after the game starts. After the software card is filled, the gameplay goes otherwise similarly to the group that plays the experimental game.

To research the potential of this game as an educational and practical tool, an action learning approach was taken by Sarrala (2022). Action learning is a concept that can be difficult to define, especially since the inventor of the concepts related to action learning, Reginald Revans, avoided giving it a definition. The problem is compounded by the fact that action learning can take many forms. For this thesis, the definition given by Marquardt (1999, p.4) is used:

> "…action learning is both a process and a powerful program that involves a small group of people solving real problems while at the same time focusing on what they are learning and how their learning can benefit each group member and the organization."

The primary aim of action learning is ultimately learning, with problem-solving being the facilitator of the learning process. Usually, action learning is performed in small groups of five to six individuals to facilitate easy communication. (Dilworth, 1998). For this reason, the group sizes for data collection for the research ranged from three to six players.

Action learning introduces new perspectives and makes the most of the pooling of the intellectual capital of the group members. An important part of action learning is placing the participants into a new setting to enable them to come up with fresh perspectives and re-examine old problems. Another important part of action learning is reflection on the learning experience. This can be done through documentation during the process and questionaries afterward, for example. (Dilworth, 1998). Action learning is well suited to complex work in a rapidly changing environment (Dilworth, 1998), which makes it an ideal fit for privacy design that is an ever-evolving and complex area as mentioned in the previous chapter on privacy.

The first cycle of the action research was performed during a five-week software engineering course in order to collect empirical data on the game in use. The players played the privacy game thrice, two times in a group and once individually. To properly investigate the matter, data was collected from a group of players that played the experimental game and a control group that played another version of the game.

3 RESEARCH DESIGN

This chapter discusses the research methods and setting, data collection, and analysis used in this study.

3.1 Research questions and hypothesis

The research questions and hypothesis this thesis attempts to answer concern the players' experience of the game and go as follows: *RQ01* Do the players find the game effective?, *RQ02* How was the players' experience of the game?, *RQ03* What differences are there between the groups that played the experimental game vs the group that played the control game?, *RQ04* Does the game affect how much interest the developer has in the topic of privacy? The hypothesis is that the group that played the experimental game will have found the game more effective than the control group.

3.2 Research setting

Data for this study were collected from software engineering students studying at a Finnish university, the University of Jyväskylä, during a five-week software development course. The course had sixty-seven participants that played the game. Of those sixty-seven, thirty-six were in the group that played the experimental game, and thirty-one played the control game. Majority of the students had either no experience or less than a year of experience in software development (71%), Agile development (81%), or SCRUM (88%). Only two people had more than 10 years of experience in software development, while 26% of the players had 1-5 years of experience. To protect the privacy of the players all data was anonymised for analysis. This thesis was done in conjunction with the research group led by Tuisku Sarrala which is researching the game developed by her.

3.3 Research method

The research method for this thesis is a case study. The case study is based on the first action learning cycle of Sarrala's research into the Privacy is not dead! -game that was performed among engineering students attending a software engineering course at the University of Jyväskylä.

Case studies are used when one wants to closely examine a particular real-life phenomenon happening within a specific context. (Zainal, 2007). Case studies and other qualitative methods are used to gain a comprehensive understanding of social/human phenomena because such things are complex and require a holistic approach to researching them to gain an in-depth understanding of the phenomena at issue (Gagnon, 2010, p. 1-2). A case study was an appropriate method for this study because it concerns a very specific phenomenon (the experience and effectiveness of a serious game intended for learning) in use in a specific context (a university course attended by future software engineers), and its purpose is to gain an understanding of the players' experiences of the game. This was achieved by having the players fill out a post-course survey that asked them about their experiences of the game.

Case studies have also a high internal validity, meaning that they reflect the reality of the phenomena well. This, however, comes with a downside since case studies can also have low external validity and generalizability. (Gagnon, 2010, p. 2-3). But these issues were deemed acceptable since a case study is the most appropriate method in this instance when the subject under investigation is so specific and occurring in a specific context.

3.4 Data collection

This thesis is linked to a research project led by Tuisku Sarrala. The research team involved in that project developed the data collection methods that were used in this thesis, as well as prepared the data to be ready to analysed. For this research project, data was collected from the group project materials, such as user stories and privacy catalogues, pre-and post-course questionnaires, interviews of players, and recordings of the gaming sessions. All data was anonymized and permission for recording was asked. However, in this study, only data from the post-course questionnaire and the privacy catalogues were used due to the limited scope of this graduate thesis. The post-course questionnaire was a 5-point scaling questionnaire, whereas the privacy catalogues were documents filled in by the players during the game, with the type and number of scenarios they came up with during the game. The data collected from the post-course questionnaire was inputted into an Excel sheet, and the privacy catalogues were collected, anonymised, and put into a file.

3.5 Data analysis

First, the data was sorted into two groups, the players who played the experimental game, and the control group, to compare the differences between their experiences of the game.

The process of quantitative analysis of the post-survey data went as follows. A Cronbach's alpha coefficient was calculated for each of the groups to see if the data sets were internally consistent. For internal consistency to be on an acceptable level, the result has to be ≥0.7. For the experimental-game group, the alpha coefficient was 0.90 and for the control group, the alpha was 0.89. This result indicates that the internal consistency of both groups' answers was in the range of good to excellent. (Cronbach, 1951). Averages for each post-questionnaire question were calculated for the experimental-game players and the control group, as well as the overall average. Then a standard deviation, standard error of the mean, and confidence mean upper and lower values were calculated. To calculate the 95% confidence level values the standard error of the mean was first multiplied with coverage factors that were calculated to account for the small group sizes. The coverage factor was 2.030 for the experimental-game group and 2.042 for the control group. The multiplied numbers were then added to the average to get the confidence mean upper value and to get the confidence mean lower value the average was deducted from the numbers. To compare the two groups and to see if there were statistically significant differences between them a Student's t-test was performed for each group. For a two-tailed t-test, as a general rule, it is considered that a p-value of 0.05 ($p < .001$) is sufficient evidence that there is a statistically significant difference in means. A p-value of 0.05 means it can be said with 95% confidence that there is a statistically significant difference. Because the sample size for this group was not very large and no t-test p-value came back with a result of 0.05 or lower, it was deemed to be acceptable to use lower confidence levels, starting with 0.30 which translates to a confidence level of 70% and higher. To compare the ability of the players to identify privacy threats in each group, the average number of scenarios each of the two groups came up with on each of the two rounds was calculated, as well as the overall average of both groups. These averages were then compared to the players' self-identified ability to identify privacy problems to see if their evaluation of their own ability was accurate.

In the post-course questionnaire, two questions were asked that the player could provide optional free-form answers to. The questions were as follows: "Any other words that you would describe the Privacy game with?" and "Any other comments to the researchers?". To gain more insight into the players' experiences of the privacy game, a thematic analysis was performed on the answers to these questions. Thematic analysis is a process for finding meaning in research by identifying and encoding patterns that appear in the data (Braun & Clarke, 2013).

The process of the thematic analysis went as follows. First, the answers were separated into those provided by the group that played the experimental game, and those who had played the control version. Then, the answers were grouped by the player, so if a single player had answered both questions, there would be no duplicate codings if they had given a similar answer in both questions. One player's answers were not included in the analysis because they simply included encouragement for the researchers. No *a priori* codes were created, the codes were created as the data was analysed. After first familiarising with the data from the players who had played the experimental game, codes were created as the data was gone through. Then, the codes that were created were analysed to see if they could be compiled into themes or if the code worked as a theme by itself. The data was then placed into a Word document, where it was sorted by themes. A table was then created, where the number of times a theme manifested was inputted. A similar process was then performed on the data from the control-group players. However, this time, the codes created during the analysis of the players who played the experimental game were used as *a priori* codes, to look for similarities in themes between groups. When the data had been sorted into applicable themes, themes were created for the data that did not fit into the *a priori* themes, and the number of times they manifested, was counted.

4 RESULTS

In this chapter, the results from the quantitative analysis and thematic analysis will be shown; sorted into sections reflecting what particular questions were attempting to research.

4.1 Experience

This section will go through questions that were deemed to relate to the players' experience of the game and are therefore important to answer the first research question.

4.1.1 How would you describe the Privacy game? Difficult to learn-Easy to learn

The first question related to players' experience of how easy they found the game, from difficult to learn to easy to learn. In GRAPH 1, the spread of answers from both groups and the overall spread of answers is shown. The spread of answers is focused on the middle of the scale.

GRAPH 1 The spread of answers within both groups related to the easiness of learning

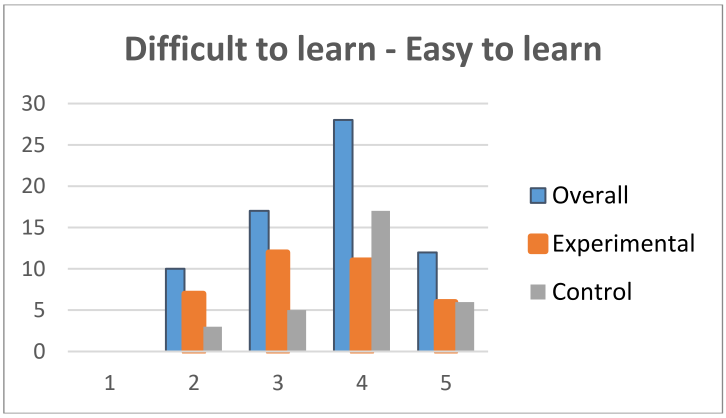


TABLE 1 Analysis of player experience; easiness of learning

Group	Experimental game	Control
Average	3.44	3.84
Standard deviation	1.00	0.86
Standard error of the mean	0.17	0.15
MAX	5	5
MIN	2	2
Standard error x2.030/2.042	0.34	0.32
95% Confidence Mean Upper Value	3.78	4.15
95% Confidence Mean Lower Value	3.11	3.52
T-test p-value		0.087

TABLE 1, displays the results of the analysis of the data related to this question. The average in both groups was over 3 with it being slightly higher in the control group. The control group also had a slightly lower standard deviation, and with the p-value of the t-test being 0.087, it can be said with close to a 90% level of confidence that the true difference in means is not equal to zero. The control group, therefore, experienced the game as somewhat easier to learn than the experimental-game group.

4.1.2 How would you describe the Privacy game? Unplayable as a game-Playable as a game

The second question is concerned with the players' experiences of the game, specifically relating to how playable they felt it was; from unplayable to playable as a game. The spread of answers within both groups and overall is displayed in GRAPH 2. The spread of answers is focused on the middle of the scale.

GRAPH 2 The spread of answers within both groups related to playability.

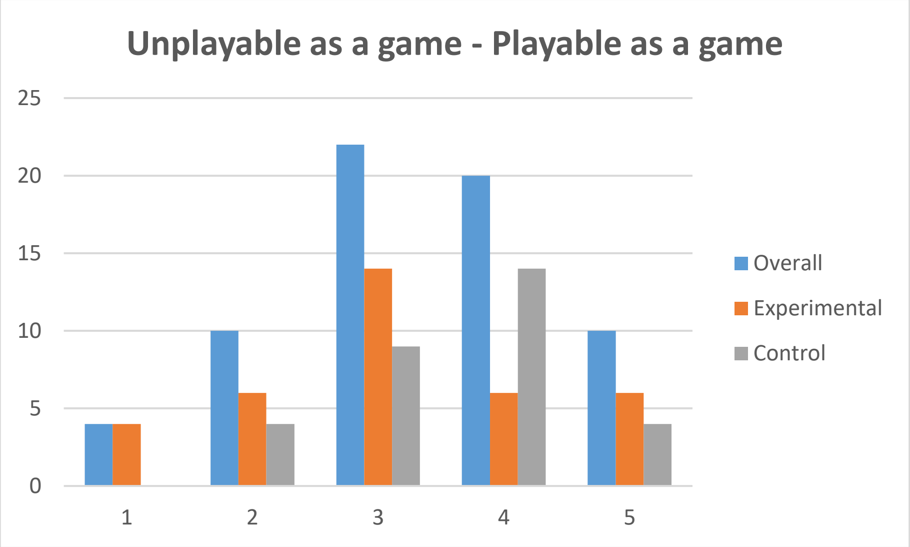


TABLE 2 Analysis of player experience; playability

Group	Experimental game	Control
Average	3.11	3.58
Standard deviation	1.21	0.89
Standard error of the mean	0.20	0.16
MAX	5	5
MIN	1	2
Standard error x2.030/2.042	0.41	0.32
95% Confidence Mean Upper Value	3.52	3.91
95% Confidence Mean Lower Value	2.70	3.26
T-test p-value		0.073

The results of the analysis of this question are displayed in TABLE 2. Again, both averages were above 3, with an experimental-game group being closer to 3, with a standard deviation rate of 1.21, whereas the control group was closer to 4, with a deviation rate of 0.89. The t-test p-value was 0.073. The null hypothesis can therefore be rejected with nearly 90% degree of confidence, and it can be said that there is a high chance that the difference in means is statistically significant, meaning that the control group found the game more playable than the group that played the experimental game.

4.1.3 How would you describe the Privacy game? Tedious-Fun, enjoyable

GRAPH 3 displays the spread of answers related to the players' experience of how enjoyable or tedious they found the game, from tedious to fun/enjoyable. The spread of answers was focused on the lower end of the scale on this question.

GRAPH 3 The spread of answers within both groups related to the enjoyability

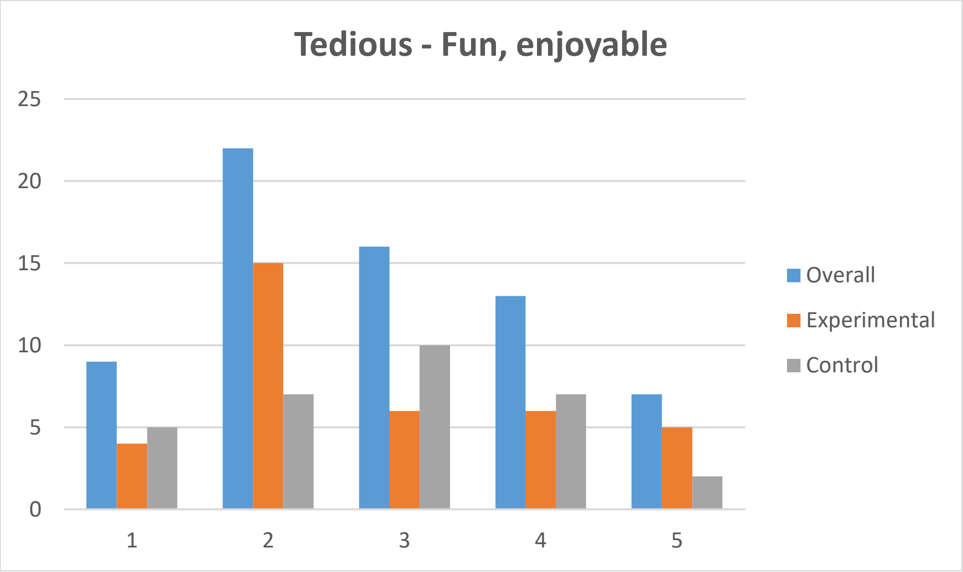


TABLE 3 Analysis of player experience; enjoyability

Group	Experimental game	Control
Average	2.81	2.81
Standard deviation	1.26	1.17
Standard error of the mean	0.21	0.15
MAX	5	5
MIN	1	1
Standard error x2.030/2.042	0.43	0.43
95% Confidence Mean Upper Value	3.23	3.23
95% Confidence Mean Lower Value	2.38	2.38
T-test p-value		0.998

TABLE 3 displays the results of the analysis on the question related to enjoyability. The averages for both groups are equal, with a slight difference in standard deviation, and the t-test value is almost 1.0. The null hypothesis is accepted in this case, meaning that the difference in means is almost equal to zero. Both groups felt that the game was equally as tedious as the other.

4.1.4 How would you describe the Privacy game? Unclear purpose-Clear purpose

This question relates to the players' experience of how clear or unclear the purpose of the game seemed to them. The spread of answers is displayed in GRAPH 4. The spread is focused more on the upper end of the scale.

GRAPH 4 Spread of answers within both groups related to the clearness of purpose

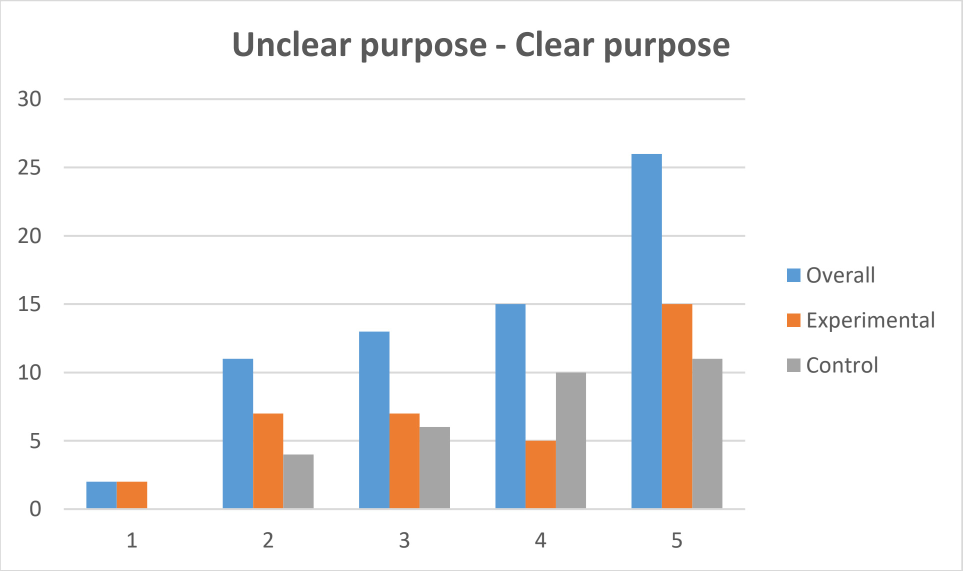


TABLE 4 Analysis of player experience; clearness of purpose

Group	Experimental game	Control
Average	3.67	3.90
Standard deviation	1.35	1.04
Standard error of the mean	0.23	0.19
MAX	5	5
MIN	1	1
Standard error x2.030/2.042	0.46	0.38
95% Confidence Mean Upper Value	4.12	4.29
95% Confidence Mean Lower Value	3.21	3.52
T-test p-value		0.423

TABLE 4 displays the results of the analysis on the question related to the players' experience of the game's clearness of purpose. Both groups' averages are near point 4 of the scale, with the experimental-game group having a standard deviation of 1.35, and the control group having 1.04 deviation. The Student's t-test p-value is 0.423, meaning that there is not a statistically significant difference between the two groups' experiences of the game's clearness of purpose.

4.1.5 Thematic analysis

Twenty-four of the players who played the experimental game and nineteen of the control players provided answers to the optional feedback questions. Therefore, in total, forty-three of the sixty-seven players answered those questions.

The most common theme that emerged from the players that played the experimental game concerned the platform through which the game was played. The most common criticism was that PowerPoint was not an appropriate platform through which to play the game, with one player writing:

> Biggest issue with the game was the platform to play it. Why in gods green earth we have to play it in a PowerPoint ??? I understand that the game was supposed to be a cardboard game but it was really tedious to play as it is now. I highly recommend that the platform the game is being played is changed.

Like this player, many others also felt that playing the game through PowerPoint made it difficult to play, took away from the enjoyment, and that it made it intuitively difficult. One player also criticized playing the game through Zoom.

> Playing through Zoom didn't help the experience. Mostly because of this there were no deeper interaction with teammates and the game felt more boring than perhaps it was originally intended.

The players who played the control game also criticised the choice of platform, although not as much as the players who had played the experimental game. They also felt that it was not intuitive and that it was too awkward as a platform for the game to be enjoyable. One player experienced confusion due to the choice of platform.

> In the beginning a bit confusing because I was expecting it to be some online game or something. Took a while to realize that the powerpoint was the game

Originally PowerPoint was chosen as the platform for the game so it would be simple to use for the players, but it seems that, instead, the players actually felt hindered by it, and that it even took away enjoyment from the game and made it cumbersome to use. For the next cycle of research, new options should be considered, to fully benefit from gamification, instead of having the results hindered by the choice of platform.

The second most common theme that emerged from the experimental-game players and the most common theme that emerged from the control players concerned the gameplay. Many felt that either there were not enough cards or that the existing cards were not relevant to the software they were developing. This tied into a complaint about the Goodies vs Baddies system that they felt often led to Goodies winning by default, which made the game repetitive and tedious. One experimental-game player left the following comment:

> The random scenarios tended to be nonsensical, like using location for identification. They also were sometimes fairly repetitive. Also "making it better" and "making it worse" cards had direct counters to each other, like "Don't tell them" and "Tell them about it" in which case declaring anyone the winner didn't really make sense.

Whereas a control gamer had this to say:

> It did not feel as a play, It was like a document to be filled in with some helpful hints. It was a bit confusing when there were the card deck and written part, so where was the game part?

Interestingly, both groups thought that there were too few cards and that the game becomes repetitive because of it. It would be expected from those players that played the control version of the game, but it seems that the experimental-game players also found the range of options lacking.

The third most common theme that emerged with both groups was confusion. This arose either from the instructions or the game itself. An experimental-game player wrote as follows:

> I personally also wished for the instructions to be clearer or a few proper examples to be provided prior to playing.

Many players expressed that they felt that the instructions were unclear, or that they were uncertain if they had played the game correctly. As shown in the last section, the choice of the platform also contributed to this feeling of confusion with some players, who felt that it was not “game-like”. For future development, perhaps clearer instructions should be considered, as well as perhaps a demonstration round played by the instructors to help alleviate confusion.

Besides the negative themes that arose, many players from both groups also expressed that the idea behind the game was good and/or interesting and had praise for the game along with the criticisms:

> The idea behind the game is fine, but the execution is kinda bad. The instructions for the game itself could be a bit more detailed on what actually is wanted.
>
> I think the idea of gamifying the privacy aspects of software development is interesting. The usage of a powerpoint slide was quick but not necessarily the most effective/clear way of doing the game. Further development might make it more effective for learning.

With further development and taking the experience and the feedback of the players into consideration when crafting those developments, the game could be improved and become a better tool for education.

4.2 Effectiveness

This section shows the results of the questions related to how effective the game was as a tool intended for learning and spotting privacy threats.

4.2.1 How would you describe the privacy game? Slow way to spot threats-Quick way to spot privacy threats

This question wanted to find out the players' experience of how effective the game was at quickly spotting privacy threats. The scale on this question went from slow to quick way to spot threats. Graph 5 displays how the answers spread on the scale in this question. They focused on the center-high points of the scale.

GRAPH 5 Spread of answers within both groups related to the quickness of the game

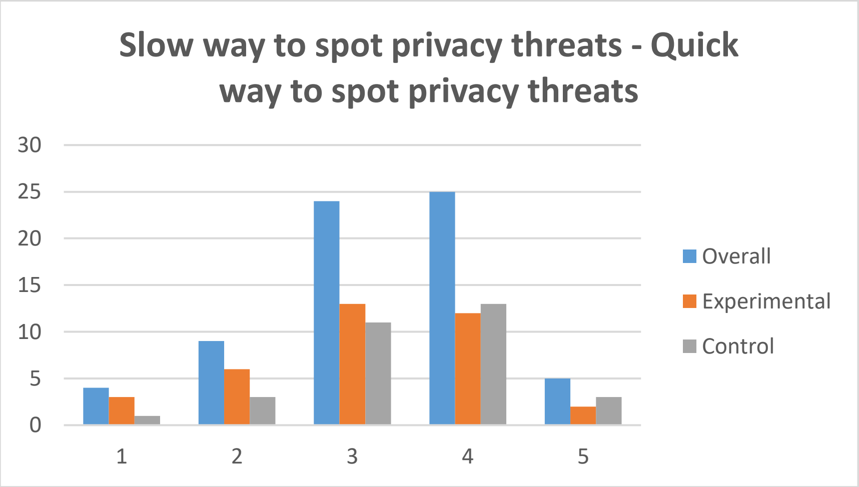


TABLE 5 Analysis of game effectiveness; quickness

Group	Experimental game	Control
Average	3.11	3.45
Standard deviation	1.04	0.93
Standard error of the mean	0.17	0.17
MAX	5	5
MIN	1	1
Standard error x2.030/2.042	0.35	0.34
95% Confidence Mean Upper Value	3.46	3.79
95% Confidence Mean Lower Value	2.76	3.11
T-test p-value		0.160

TABLE 5 shows the results of the analysis of this question. The averages were both above the middle point of the scale, with the control group having a slightly higher average. The standard deviation for the experimental-game group (1.04) was slightly higher than the control group's (0.93). The t-test p-value is 0.160, meaning that it can be said with close to an 85% level of confidence that the difference in means is statistically significant and that the control group felt more strongly that the game was a quick way to spot privacy threats.

4.2.2 How would you describe the privacy game? Ineffective, produces few threats-effective, helps to spot many threats

This question concerned how effective the groups considered the game to be at generating privacy threats. The scale ranged from ineffective, produces few threats to effective, helps to spot many threats. GRAPH 6 shows the spread of results within both groups and the overall spread of answers. The answers to this question were focused on points 3 and 4 on the scale.

GRAPH 6 Spread of answers within both groups related to the effectiveness of the game

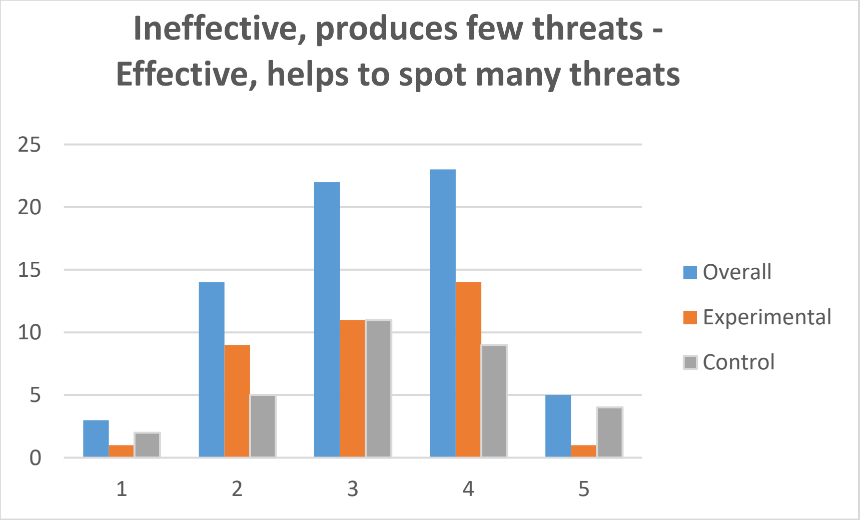


TABLE 6 Analysis of game effectiveness; finding threats

Group	Experimental game	Control
Average	3.14	3.26
Standard deviation	0.93	1.09
Standard error of the mean	0.16	0.20
MAX	5	5
MIN	1	1
Standard error x2.030/2.042	0.31	0.40
95% Confidence Mean Upper Value	3.45	3.66
95% Confidence Mean Lower Value	2.82	2.86
T-test p-value		0.636

TABLE 6 displays the results of the analysis of the data on this question. Both group averages are close to the middle of the scale, with the control group having a slightly higher average. The experimental-game standard deviation was 0.93, and the control was 1.09. With the p-value of the t-test being 0.6, it cannot be said with high certainty that the result is statistically significant. Both groups fell equally in the middle ground on this question.

4.2.3 What was the effect of the Privacy game on the privacy quality of your team's software?

This question asked if the Privacy game affected the privacy quality of the software that the teams developed during the course; as in, if the game was effective at improving their awareness of privacy issues that then translated into practice, with the scale going from no effect to significant effect. GRAPH 7 displays the spread of answers within both groups and the overall spread of answers. The answers to this question seem to tilt toward the lower end of the scale.

GRAPH 7 Spread of answers within both groups related to the impact of the game on the software

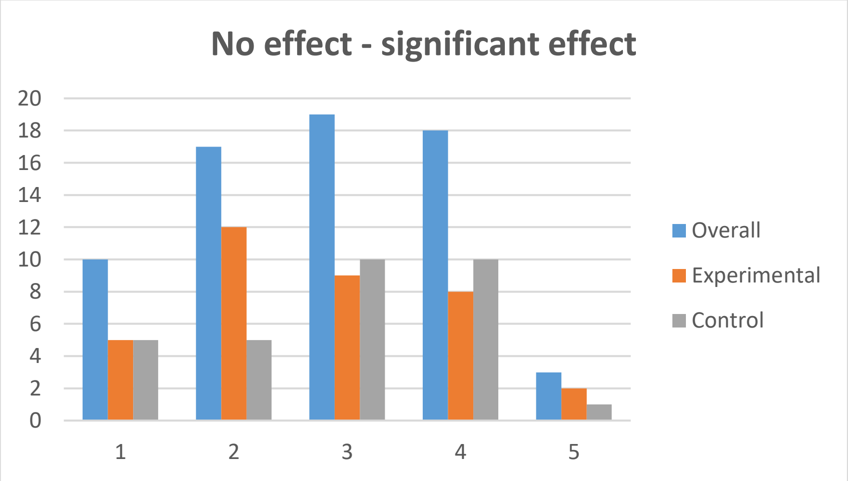


TABLE 7 Analysis of effectiveness; impact on software

Group	Experimental game	Control
Average	2.72	2.90
Standard deviation	1.14	1.14
Standard error of the mean	0.19	0.20
MAX	5	5
MIN	1	1
Standard error x2.030/2.042	0.38	0.42
95% Confidence Mean Upper Value	3.11	3.32
95% Confidence Mean Lower Value	2.34	2.49
T-test p-value		0.518

TABLE 7 displays the analysis of the data in this question. Both groups had averages that fell below the middle point of the scale. Both groups had the same standard deviation. With a 0.518 p-value, it cannot be said with high confidence that there is a statistically significant difference between the answers of both groups. Both groups deemed the game to be about equally not very effective at impacting the privacy quality of their software.

4.2.4 How would you describe the privacy game? Doesn't educate-educational

This question asked to what extent the players felt the game was educational, the scale going from doesn't educate to educational. The spread of answers is displayed in GRAPH 8. The answers to this question were focused more on the higher end of the scale, with the control group's answers placed more on the higher end than the group who played the experimental game.

GRAPH 8 Spread of answers within both groups related to how educational the game was

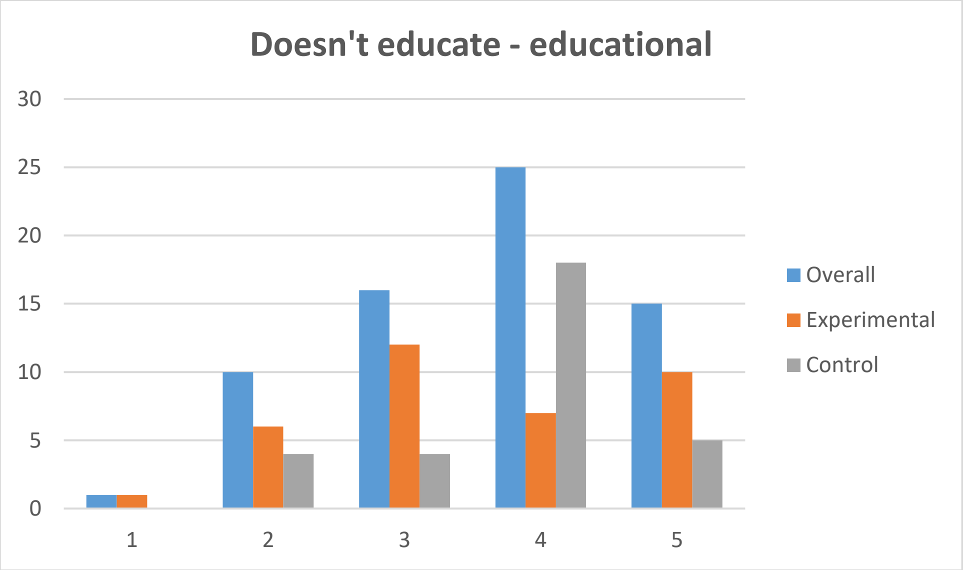


TABLE 8 Analysis of effectiveness; how educational was the game

Group	Experimental game	Control
Average	3.53	3.77
Standard deviation	1.16	0.88
Standard error of the mean	0.19	0.16
MAX	5	5
MIN	1	2
Standard error x2.030/2.042	0.39	0.32
95% Confidence Mean Upper Value	3.92	4.10
95% Confidence Mean Lower Value	3.14	3.45
T-test p-value		0.328

TABLE 8 shows the results of the analysis. Both averages were near point 4 on the scale, with the control group higher than the experimental-game group. The standard deviation for the experimental-game group was 1.16, whereas the control group had a standard deviation of 0.88. The p-value for the t-test was 0.328, which means that it can be said with a 70% of confidence value that there is a statistically significant difference in means, meaning that the control group experienced the game as somewhat more educational than the group who played the experimental game.

4.3 Motivation

This section is about how motivated the players felt that this game made them. As mentioned in the chapter on gamification, motivation is an important aspect of a gameful experience, so it felt necessary to include the players' feelings about their level of motivation during the gaming experience.

4.3.1 How would you describe the privacy game? Demotivates me to play-Motivates me to play

The first question related to motivation asked the players if the game motivated or demotivated them to play. The spread of answers from both groups and the overall spread of answers can be seen in GRAPH 9. The overall answers are focused on the lower and middle range of the scale, with the players who played the experimental game, falling more on the lower end of the scale and the players who played the control version falling more on the middle to high range of the scale.

GRAPH 9 Spread of answers within both groups related to motivation to play

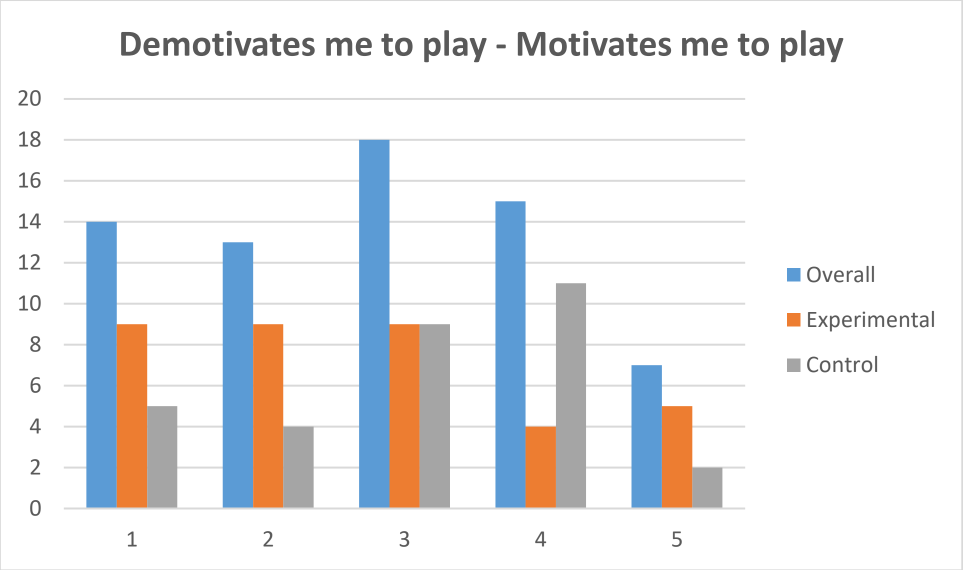


TABLE 9 Analysis of motivation; motivation to play

Group	Experimental game	Control
Average	2.64	3.03
Standard deviation	1.36	1.20
Standard error of the mean	0.23	0.21
MAX	5	5
MIN	1	1
Standard error x2.030/2.042	0.46	0.44
95% Confidence Mean Upper Value	3.10	3.47
95% Confidence Mean Lower Value	2.18	2.59
T-test p-value		0.212

TABLE 9 shows the results of the analysis of the data on this question. The average for the players who played the experimental game was below the middle point of the scale, with a standard deviation of 1.36, while the average for control players was above it, with a standard deviation of 1.20. With a t-test p-value of 0.212, it can be said with 80% confidence that the difference in means is statistically significant, meaning that the control group was somewhat more motivated to play the game than the group who played the experimental game.

4.3.2 How would you describe the privacy game? Demotivates me to seek threats-Motivates me to seek threats

This question asks about the level of motivation the game gave players to seek threats, with the scale from demotivated to motivated. The spread of answers for both groups and the overall spread of answers on the scale can be seen in GRAPH

10. The answers to this question are focused more on the higher end of the scale with both groups.

GRAPH 10 Spread of answers within both groups related to motivation to seek threats

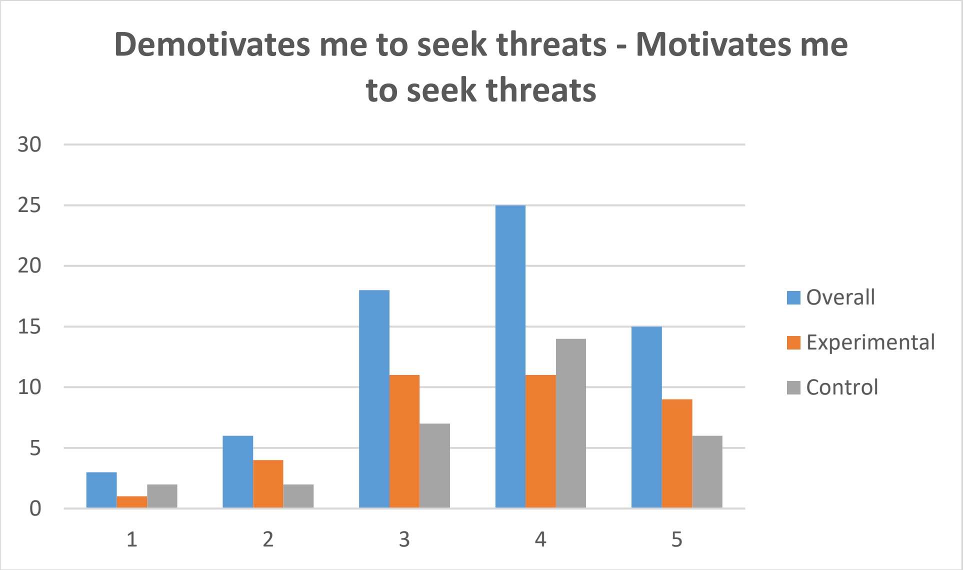


TABLE 10 Analysis of motivation; motivation to seek threats.

Group	Experimental game	Control
Average	3.64	3.84
Standard deviation	1.07	1.08
Standard error of the mean	0.18	0.19
MAX	5	5
MIN	2	1
Standard error x2.030/2.042	0.36	0.40
95% Confidence Mean Upper Value	4.00	4.04
95% Confidence Mean Lower Value	3.28	3.25
T-test p-value		0.981

TABLE 10 displays the analysis of the data related to this question. The averages for both groups were near point 4 of the scale, with almost identical rates of standard deviation. The t-test p-value is near 1.00 meaning that there is no statistically significant difference in means; both groups felt about equally motivated to seek threats by the game.

4.4 Interest in the topic

This section looks at how interested and how important the players considered the topic of privacy before and after the game.

4.4.1 How interested were you in the topic of privacy?

GRAPH 11 displays the responses to the question regarding how interesting the players found the topic of privacy, before and after the game. The scale ranged from low interest to high interest. The experimental game players' interest before the game is shown in the green bars (EX-BF), and their interest after is displayed in the red bars (EX-AF). The control group's answers are shown in the grey (control-before the game; C-BF) and yellow bars (control-after the game; C-AF). The answers to the question focused on the higher end of the scale both before and after the players played the game. Only a few players in both groups had low interest before and after the game.

GRAPH 11 The spread of answers related to the players' interest in the topic of privacy before and after the game, scaled from low to high

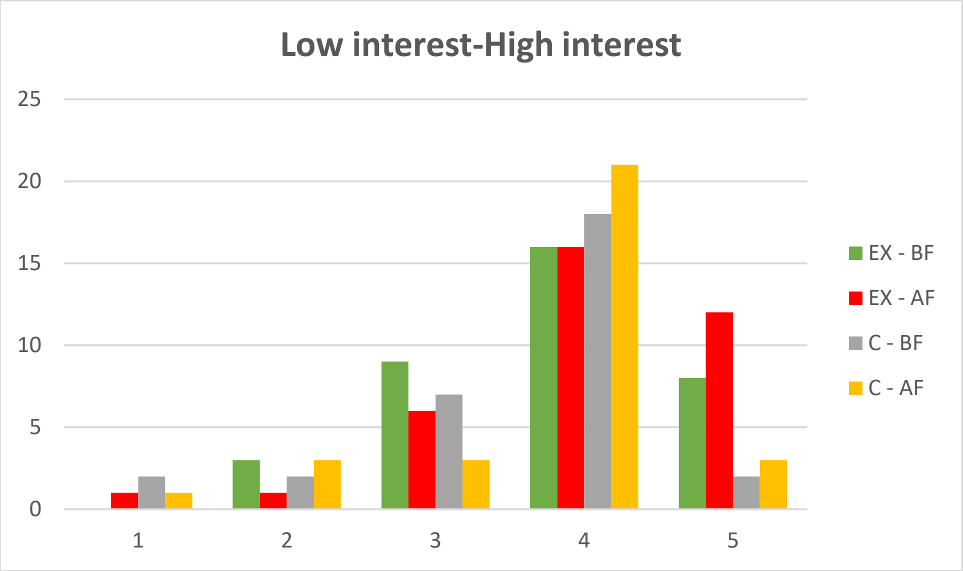


TABLE 11 Analysis of player interest in the topic of privacy, before and after the game

Group	Experimental game	Control	Experimental game	Control
	Before game	Before game	After game	After game
Average	3.81	3.52	4.03	3.71
Standard deviation	0.89	0.96	0.94	0.90
Standard error of the mean	0.15	0.17	0.16	0.16
MAX	5	5	5	5
MIN	2	1	1	1
Standard error x2.030/2.042	0.30	0.35	0.32	0.33
95% Confidence Mean Upper Value	4.11	3.87	4.35	4.04
95% Confidence Mean Lower Value	3.50	3.16	3.71	3.38
T-test p-value		0.208		0.163

TABLE 11 displays the results of the analysis. Both groups started with rather high-interest rates even before the game, but the experimental-game players had higher interest than the control group. The experimental-game group had a standard deviation of 0.89 and the control group had a standard deviation of 0.96. With the t-test's p-value being 0.208, it can be said with 80% certainty that this difference in means is statistically significant. After the game, the interest rose in both teams by about 0.20 points. The deviation levels in this question were 0.94 and 0.90 respectively. Yet again, the experimental-game group expressed higher interest than the control group, and with a p-value of 0.163, it can be said that this difference is statistically significant with a near 85% confidence level. The experimental-game group, therefore, rated their interest in the topic of privacy higher both before and after the game, and there was some rise before and after, but it was negligible. The interest rate was also quite high with the control group both before and after.

4.4.2 How important topic did you consider privacy to be?

The spread of answers that were given in the question related to how important the players felt the topic of privacy to be both before and after they played the game is shown in GRAPH 12. The players were given a number scale that ranged from low importance to high importance. Both groups' answers both before and after the game were focused on the high end of the scale, both before and after, with some rise after the game. Even fewer people fell on the lower end of the scale than in the previous question. The players therefore have and do consider privacy a very important topic. Whether this transfers to the actual privacy quality of their future softwares is another question entirely, since, as mentioned before, the gap between the concept and the execution exists.

GRAPH 12 The spread of answers related to how important the players felt the topic of privacy to be before and after the game, scaled from low to high

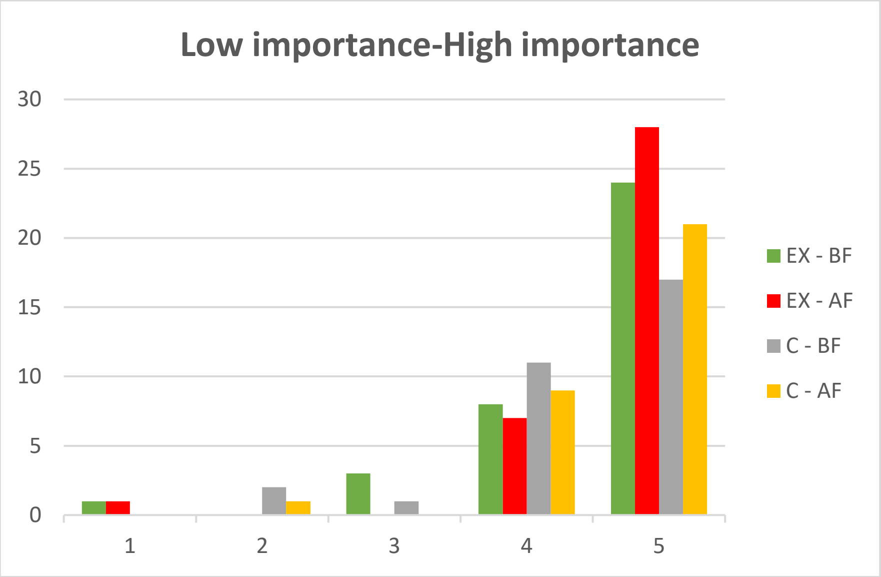


TABLE 12 Analysis of how important the players thought the topic of privacy was before and after the game

Group	**Experimental game**	**Control**	**Experimental game**	**Control**
	Before game	**Before game**	**After game**	**After game**
Average	4.50	4.39	4.69	4.61
Standard deviation	0.88	0.84	0.75	0.67
Standard error of the mean	0.15	0.15	0.12	0.12
MAX	5	5	5	5
MIN	1	2	1	2
Standard error x2.030/2.042	0.30	0.31	0.25	0.24
95% Confidence Mean Upper Value	4.80	4.70	4.95	4.86
95% Confidence Mean Lower Value	4.20	4.08	4.44	4.37
T-test p-value		0.594		0.639

TABLE 12 displays the results of this question. Both groups expressed that they thought that the topic of privacy had high importance both before and after the game. There was only about a 0.20-point rise after the game in both groups. Before the game their respective deviation levels were 0.88 and 0.84, and after the game 0.75 and 0.67. The p-value being 0.594 and 0.639 correspondingly, it cannot be said that there was a statistically significant difference between groups either before or after.

4.5 Ability

The players were asked to estimate their own ability to identify privacy threats before and after playing the privacy game from low ability to high ability. This section will analyse the results from that two-part question. The spread of answers within both groups is displayed in GRAPH 13. The answers from the group that played the experimental game are shown in grey (EX-BF: experimental-before game) and yellow (EX-AF; experimental-after game), whereas the answers from the control group will be displayed in blue (C-BF; control-before game) and green (C-AF; control-after game). The answers for the players' ability before the game are focused on the middle-to-low part of the scale, and answers for their ability after the game are focused more on the higher end of the scale. Only the group that played the experimental game gave any answers that fell on the highest point of the scale.

GRAPH 13 The player self-identified ability to identify privacy threats before and after the game, scaled from low to high

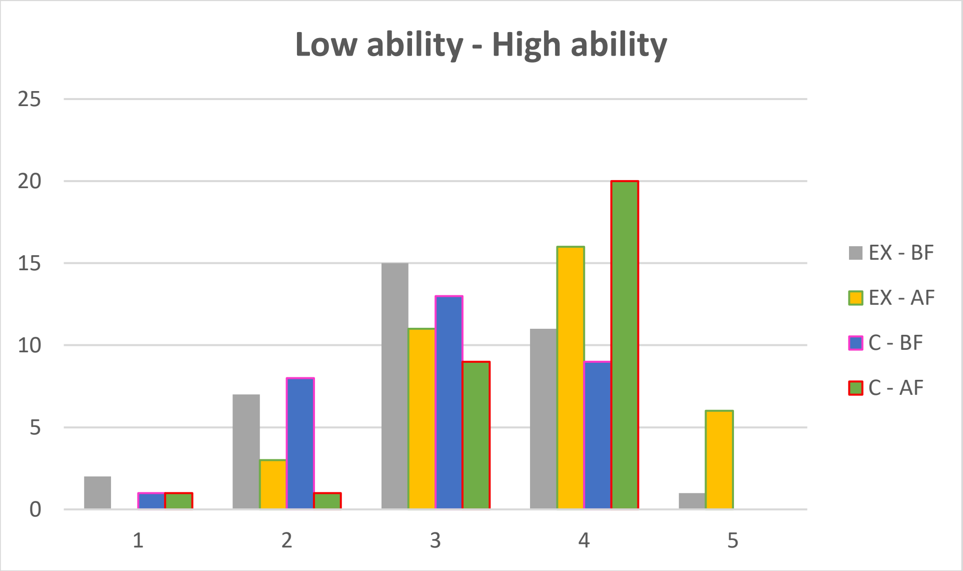


TABLE 13 Analysis of player self-identified ability to spot privacy threats before and after gaming

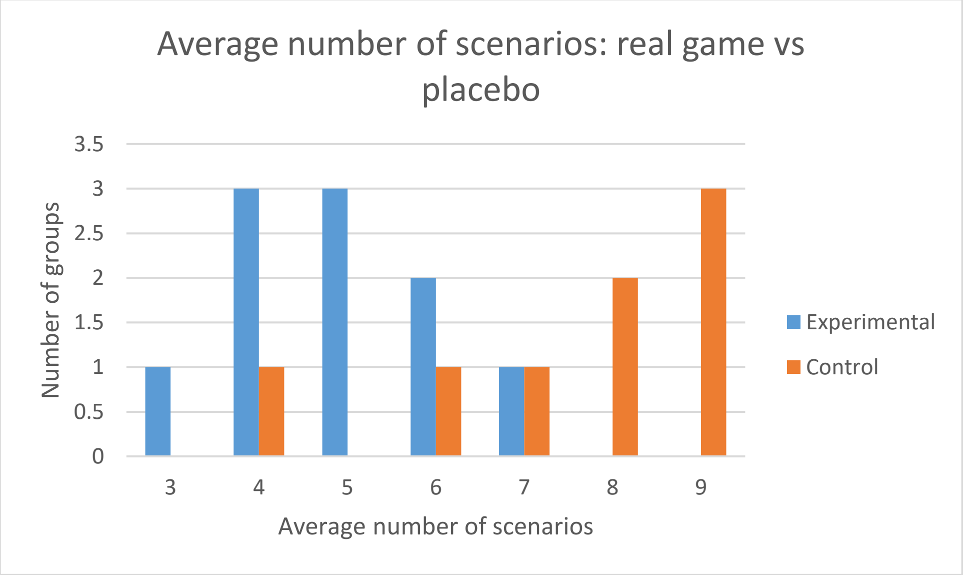
Group	**Experimental game**	**Control**	**Experimental game**	**Control**
	Before game	**Before game**	**After game**	**After game**
Average	3.06	2.97	3.69	3.55
Standard deviation	0.92	0.84	0.86	0.72
Standard error of the mean	0.15	0.15	0.14	0.13
MAX	5	4	5	4
MIN	1	1	2	1
Standard error x2.030/2.042	0.31	0.31	0.29	0.27
95% Confidence Mean Upper Value	3.37	3.27	3.98	3.81
95% Confidence Mean Lower Value	2.74	2.66	3.40	3.28
T-test p-value	0.684		0.452	

TABLE 13 shows the analysis of the data on this question. The experimental-game players' average rating of their ability to spot threats before the game was 3.06, with a standard deviation of 0.92. The control group's average rating of their ability was somewhat lower with an average of 2.97, with a standard deviation of 0.86. The t-test p-value was 0.682, meaning that it cannot be said with confidence that the difference is statistically significant. Both groups evaluated their ability to spot threats before the game averagely at the middle point of the scale.

The average for both groups rises after the game, with the experimental game having an average of 3.69, with a standard deviation of 0.86, and the control group having an average of 3.55, with a standard deviation of 0.72. The t-test value for this question was 0.452, meaning that it cannot be said with high confidence that there is a statistically significant difference between the groups. Notable is that both groups' averages rose about 0.50-0.60 points in evaluating their after-game ability, so both groups felt that the game at least somewhat improved their ability to spot privacy threats.

To compare their evaluation of their ability to spot privacy threats, to the actual number of scenarios created by each group, an average number of scenarios that were created by each group on each of the two rounds of the privacy game was calculated. GRAPH 14 shows the average number of scenarios (ranging from 3 to 9), comparing it to the number of teams in each group that created that amount. The teams in the group that played the experimental game created notably fewer scenarios than the control teams. If this result is compared to the result of the questionnaire analysis, we can see that the experimental-game group possibly overestimated their ability to spot privacy threats, whilst the control group possibly underestimated it (with no member of the control group evaluating their ability higher than 4). However, this analysis did not look at the quality of the privacy catalogues, only their quantity, which might affect the results of this analysis, if there are significant differences between the quality of the privacy catalogue contents between groups.

GRAPH 14 Average number of scenarios the groups came up with



5 DISCUSSION

This chapter will go through the research questions laid out in the introduction of this study and what the answers to those research questions were according to the results of the data analysis. After that, the implications of the study for both theory and practice will be discussed.

5.1 Answers to research questions

This section will go through the research questions from the questions that build on the main research questions of this thesis; namely, if the players found the game to be effective and what their experience of it was like.

5.1.1 What differences are there between the groups that played the experimental game vs the group that played the control game?

There surprisingly were not that many significant differences between the groups. The questions that returned a p-value with the highest confidence value (90%) that there is a significant difference between the means of the groups were concerned with how difficult/easy the players thought the game was to learn and how playable/unplayable they thought it was. The control group expressed that the game was easier for them to learn than the group that played the experimental game. They also found it more playable than the experimental group. However, both groups had an average that was over the middle point of the scale, and both had some variation within the groups, with answers ranging from the lowest point of the scale to the highest point.

The second highest confidence value (85%) came back on the questions regarding how helpful the players found the game to be in quickly spotting privacy threats, and how interested the players were in the topic of privacy after playing the game. The control players again found the game a quicker way to spot privacy threats than the experimental-game group which is perhaps surprising since instead of having ready-made scenario cards, they had to come up with scenarios themselves. However, they were less interested in the topic of privacy after gaming than the experimental-game group, but it is to be noted that their interest before the game was also lower and it rose about the same amount as the experimental-game group's.

The third highest confidence level (80%) came back from the questions that related to motivation to play and how interested the groups were in privacy before the game. The control group found that the game motivated them to play more than the experimental group, with their average above 3 and the experimental group's being 2.64, with a higher variance. However, both groups felt about equally motivated by the game to seek threats, which is a somewhat contradictory result. Additionally, the experimental group expressed that they were more interested in the topic of privacy before the game than the control group.

The lowest confidence level (70%) came back on the question related to how educational the players found the game, with the control group again finding the game more educational than the experimental-game group. However, with confidence this low, it is harder to say if the difference is actually statistically significant.

Another difference between the groups was the number of privacy scenarios they created in each game round. The control group created significantly more scenarios during the course than the experimental-game group but simultaneously rated their ability as averagely lower. However, the p-value did not return with a high confidence value so it cannot be said that the difference in their rated ability was statistically significant.

The differences between the groups manifested in how playable, easy to learn, helpful for spotting threats, and motivating the players experienced the game to be, with the control group having a higher average in all of these questions, and interest in the topic of privacy before and after the game, with the experimental-game group having more interest in the topic both before and after.

5.1.2 Does the game affect how much interest the players have in the topic of privacy?

The experimental-game group rated their interest in the topic of privacy higher than the control group both before and after the game. However, both groups rated their interest quite highly on the scale even before the game. Interest in the topic rose about 0.20 points after gaming in both groups, which is not significantly higher. Both groups also rated the importance they placed on the topic of privacy before and after the game very high (both times over point 4 on the scale), with the level of importance rising from before gaming to after gaming about 0.20 points. From this, it cannot be said that the game had any significant effect on how much interest the players had in the topic of privacy before and after playing the game.

5.1.3 How was the players' experience of the game?

The players thought that the game was not very motivating when it came to motivation to play, with the experimental-game group finding it less motivating than the control group. However, they also found that it motivated them to find threats rather well. Both groups also expressed that it was the game was closer to tedious than enjoyable, that it did not engage their interest that well, and that it did not have much of an effect on the privacy quality of the software they developed during the course. The groups found the game quite easy to learn, with the control group finding it easier to learn than the experimental-game group. The players rated the game as average for how effective it was for spotting threats, but the control group found it was a somewhat quicker method to spot threats than the experimental-game group. The playability of the game was also rated higher by the control group, with the experimental group rating the playability about the middle point of the scale. The purpose of the game was felt to be quite clear, and the game was felt by both groups to be at least somewhat educational, and that it was quite useful in software development. However, both group's had complaints about the choice of platform, the repetitiveness of the gameplay, and they also expressed that the game caused them to experience confusion, for example, by felt lack of clear instruction. Despite the negative aspects, the players did feel that the idea behind the game was good and interesting.

5.1.4 Do the players find the game effective?

In all the questions evaluating effectiveness, the averages for both groups stayed near the middle point of the scale. The highest averages were returned in the question regarding how educational the game was felt to be, with the control group (3.77) rating the game as more educational than the experimental-game group (3.53). Whether this difference is statistically significant is somewhat uncertain since the confidence level is merely 70%. But, overall, it can be said that the players felt that the game was more educational than not.

However, both groups rated the actual effect of the game on the final software's privacy quality quite low, with both averages being under point 3 on the scale (experimental game: 2.72, control 2.90). The impact on the software was quite possibly so low because the thematic analysis revealed that many players felt the cards were lacking and not relevant to the software they were developing. The players also rated the effectiveness of the game at spotting threats quite averagely (3.14, 3.26), as well as the quickness of the game at spotting threats (3.11, 3.45), although the control group found the game statistically significantly to be quicker.

Since the results are quite consistently average, it is difficult to say that the players felt that the game was effective, especially since the actual impact on the software was rated so low. But since the results fell on the middle ground, it is also difficult to say that the players felt that the game was not effective at all. However, the groups, especially the control group, were able to generate many scenarios during the game.

5.1.5 Hypothesis

The hypothesis for this thesis was:

> HP01 the group that played the experimental game found the game more effective than the control group.

This hypothesis surprisingly proved not to be true. The control group actually found the game to be more effective in all questions regarding effectiveness, and this difference in means was statistically significant in at least one question measuring effectiveness. Based on this, it should be perhaps evaluated if the control version of the game would not be a better option rather than the experimental version.

5.2 Implications for theory

The privacy game aimed at filling in the gap between privacy as a concept and privacy tools, by teaching engineers privacy thinking skills. A gamified approach was taken since there is evidence that gamification can improve learning. The results were not the most illuminating based on this data since this thesis merely looked at the players' experience of the game and their own evaluation of the game's effectiveness, and because the results were rather inconclusive. Further research is needed, and additional studies should be concluded on the game, and the rest of the data, including the concrete product, and the software, should be analysed to gain a more holistic view of the game's value as an educational and practical tool. The results did, however, uncover potential areas that need development to improve the game.

5.3 Implications for practice

According to previous research, a game does not have to be enjoyable to be educational (Kapp, 2012). Therefore, the privacy game could potentially be used in practice in its current form. The players indicated that the game did not have much of an impact on the privacy quality of their software, but to find out if the game actually was effective at developing the engineers' mental privacy thinking tools, further research would be needed. From the results, it is rather clear that further development of the game is needed for the full benefit of gamification and player enjoyment. Considering player feedback, a platform change should at the very least be considered, and the instruction given to players before the gameplay honed to alleviate confusion that arose in the players during this trial run. To improve player motivation, it should be considered if additional elements that generate both intrinsic and extrinsic motivation should be added. These efforts could simultaneously improve how engaging and enjoyable the players will find the game. It should also be taken into account that the control group found the game more motivating, playable, easy to learn, and a quicker tool for spotting privacy threats. It could be considered that perhaps using the control game would be more effective than using the experimental game. Alternatively, a new game could be developed with the same idea behind it while taking into account the results of this experiment in gamified approach to teaching engineers privacy thinking.

6 CONCLUSION

Privacy has been a topic of debate for a long time, but it has risen to new importance with the advent of the information era with the ubiquitous nature of technology and data in our everyday lives. The protection of individuals' data privacy has given rise to multiple different methods. However, the problem with many of these methods is that they are reductionist and consider privacy from a narrow point of view. The problem of data privacy protection is amplified by how hyper-connected today's systems are and how they generate massive amounts of data daily, which makes predicting potential vulnerabilities a difficult task. There also exists a gap between privacy as a concept and the tools used to implement protections in software. Engineers often consider privacy a secondary problem and they do not have the know-how to properly implement it. Engineers need to develop mental tools for privacy thinking to properly develop software with privacy in mind. The Privacy is not dead! -game aims to teach them to do just that.

The game is based on scenario development, which is a problem-solving tool that breaks down complex problems into manageable parts and helps to imagine different futures and uncertainties, and gamification, which is the use of game elements in non-game contexts and has promising results on the area of education, where it has been shown to improve learning. The game was played by a group of software engineering students. One team played the experimental version, and another played a control version. This thesis was a case study that examined the differences between the teams, their experiences of the game, how it affected their interest in the topic of privacy, and if they found it effective by analysing the post-course survey filled in by the players after the course.

The differences between the groups manifested in how playable, easy to learn, helpful for spotting threats, and motivating the players experienced the game to be, with the control group having a higher average in all of these questions, and interest in the topic of privacy before and after the game, with the experimental-game group having more interest in the topic both before and after.

The players thought that the game was more tedious than enjoyable, it did not engage their interest, it was not very motivating, the game was easy to learn, average as a tool for spotting threats, and it did not have much of an effect on the privacy quality of the software they developed, but that it was somewhat educational and quite useful in software development. They also had complaints about the choice of platform, the unclearness of instruction and the gameplay, which they felt was repetitive. However, the players did feel that the idea behind the game was good and interesting. The game also did not affect the players' level of interest in privacy. That remained high throughout. The results on the game's effectiveness were rather inconclusive since the effectiveness was rated averagely by players. More research is needed for the full picture.

This study had some significant limitations. Since this is a master's thesis, the scope of this study is small, and not all data that was gained from the study group was used in the analysis. The results largely relied on the post-course questionnaire data that cannot give a full picture of the game's effectiveness and value, since the questionnaire was based on self-evaluation and not the actual tangible output; the software and user stories. And although privacy catalogues were used in the analysis, this study did not consider their quality, only their quantity. The size of the research group was also quite small, as were the experimental-game group and control-game group sizes. The qualitative analysis part of the results also suffers from non-response bias, since the open-ended questions were optional, so free feedback was only received from those players that took the time to answer those questions. Bias can also come from the fact that the players are aware that their answers will be read, meaning that the players could have answered in ways that made them look good or answered in the way that they feel that is the "right answer" even though they do not actually feel that way. This particularly affects the questions regarding their level of interest in privacy and how important they consider that topic, and their ability to spot privacy threats. It is, therefore, harder to say if the game actually affected their level of interest in the topic of privacy, however, at least the question of ability can be controlled by comparing their answers to the actual number and quality of privacy threats they wrote on their privacy catalogues during the game rounds. To really grasp the value of the game in developing privacy thinking skills in software engineers, the control group should not have played the game at all, to properly see the difference in the privacy thinking skills between these groups. The questions regarding their interest in privacy should have also been asked both before and after the actual game, not simply on the post-questionnaire. This would have perhaps shown more accurately how the game affected their interest in privacy.

As already discussed in the last chapter, further research and development on this game should be considered, as well as a fuller review of all the data gathered during the research should be conducted to gain a more holistic picture of the game's effectiveness as a potential tool for developing privacy thinking in software engineers. Further research should also be conducted on the ways to combine privacy with software engineering since this appears to be a significant problem in actually applying privacy controls on software. Additionally, further research should be conducted on the development of other tools besides this game to develop engineers’ mental tools related to privacy thinking to potentially close this gap between privacy as a concept and privacy engineering tools.

SOURCES

Anthonysamy, P., Rashid, A., & Chitchyan, R. (2017, May). Privacy requirements: present & future. In 2017 IEEE/ACM 39th international conference on software engineering: software engineering in society track (ICSE-SEIS) (pp. 13-22). IEEE.

Bekebrede G., Warmelink H.J.G. & Mayer I.S. (2011) Reviewing the need for gaming in education to accommodate the net generation. Computers & Education 57, 1521–1529

Bishop, P., Hines, A., & Collins, T. (2007). The current state of scenario development: an overview of techniques. foresight, 9(1), 5-25.

Braun, V., & Clarke, V. (2013). Successful qualitative research: A practical guide for beginners. London, England: SAGE

Colesky, M., Hoepman, J.-H., & Hillen, C. (2016). A Critical Analysis of Privacy Design Strategies. 2016 IEEE Security and Privacy Workshops (SPW). https://doi.org/10.1109/spw.2016.23

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. psychometrika, 16(3), 297-334.

Deterding, S., Dixon, D., Khaled R., & Nacke L., (2011). From Game Design Elements to Gamefulness: Defining "Gamification", Proceedings of MindTrek, 2011

Dilworth, R. L. (1998). Action learning in a nutshell. Performance Improvement Quarterly, 11(1), 28-43.

European Union (2016). Regulation (EU) 2016/679, General Data Protection Regulation (GDPR). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016, on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, 2016. URL: http: //eur-lex.europa.eu/legal-content/FI/TXT/?uri=CELEX%3A32016R0679.

FTC (2019, July 24). FTC Imposes $5 Billion Penalty and Sweeping New Privacy Restrictions on Facebook. Federal Trade Commission. https://www.ftc.gov/news-events/press-releases/2019/07/ftc-imposes-5-billion-penalty-sweeping-new-privacy-restrictions

Gagnon, Y. C. (2010). The case study as research method: A practical handbook. PUQ.

General Data Protection Regulation (GDPR) (2021a, October 22). GDPR Fines/Penalties. General Data Protection Regulation (GDPR). https://gdpr-info.eu/issues/fines-penalties/

General Data Protection Regulation (GDPRb) (2021, October 22). GDPR Privacy by Design. General Data Protection Regulation (GDPR). https://gdpr-info.eu/issues/privacy-by-design/

Girard, C., Ecalle, J., & Magnan, A. (2013). Serious games as new educational tools: how effective are they? A meta-analysis of recent studies. Journal of computer assisted learning, 29(3), 207-219.

Gürses, S., Troncoso, C., & Diaz, C. (2011). Engineering Privacy by Design. https://software.imdea.org/~carmela.troncoso/papers/Gurses-CPDP11.pdf

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does Gamification Work? -- A Literature Review of Empirical Studies on Gamification. 2014 47th Hawaii International Conference on System Sciences. https://doi.org/10.1109/hicss.2014.377

Huotari, K., & Hamari, J. (2012, October). Defining gamification: a service marketing perspective. In Proceeding of the 16th international academic MindTrek conference (pp. 17-22).

Kalloniatis, C., Kavakli, E., & Gritzalis, S. (2008). Addressing privacy requirements in system design: the PriS method. Requirements Engineering, 13(3), 241–255. https://doi.org/10.1007/s00766-008-0067-3

Kapp, K. M. (2012). The gamification of learning and instruction: game-based methods and strategies for training and education. John Wiley & Sons.

Kostova, B., Gürses, S., & Troncoso, C. (2020). Privacy engineering meets software engineering. on the challenges of engineering privacy by design. arXiv preprint arXiv:2007.08613

Marquardt, M. (1999). Action learning in action: Transforming problems and people for world-class organizational learning. Palo Alto, CA: Davies-Black Publishing.

Prensky M., ed. (2001) Digital natives digital immigrants. In On the Horizon, Vol. 9 pp. 1–6. MCB University Press.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. Contemporary educational psychology, 25(1), 54-67.

Sarrala, T. (2021) Uncovering privacy threats with soft systems methodology: Development of a privacy threat modelling method for today’s needs.

Sarrala, T. (2022) Privacy thinking in software engineering: A research agenda. [Unpublished manuscript]

Schoemaker, P. J. (1993). Multiple scenario development: Its conceptual and behavioral foundation. Strategic management journal, 14(3), 193-213.

Senarath, A., Grobler, M., & Arachchilage, N. A. G. (2019). Will They Use It or Not? Investigating Software Developers’ Intention to Follow Privacy

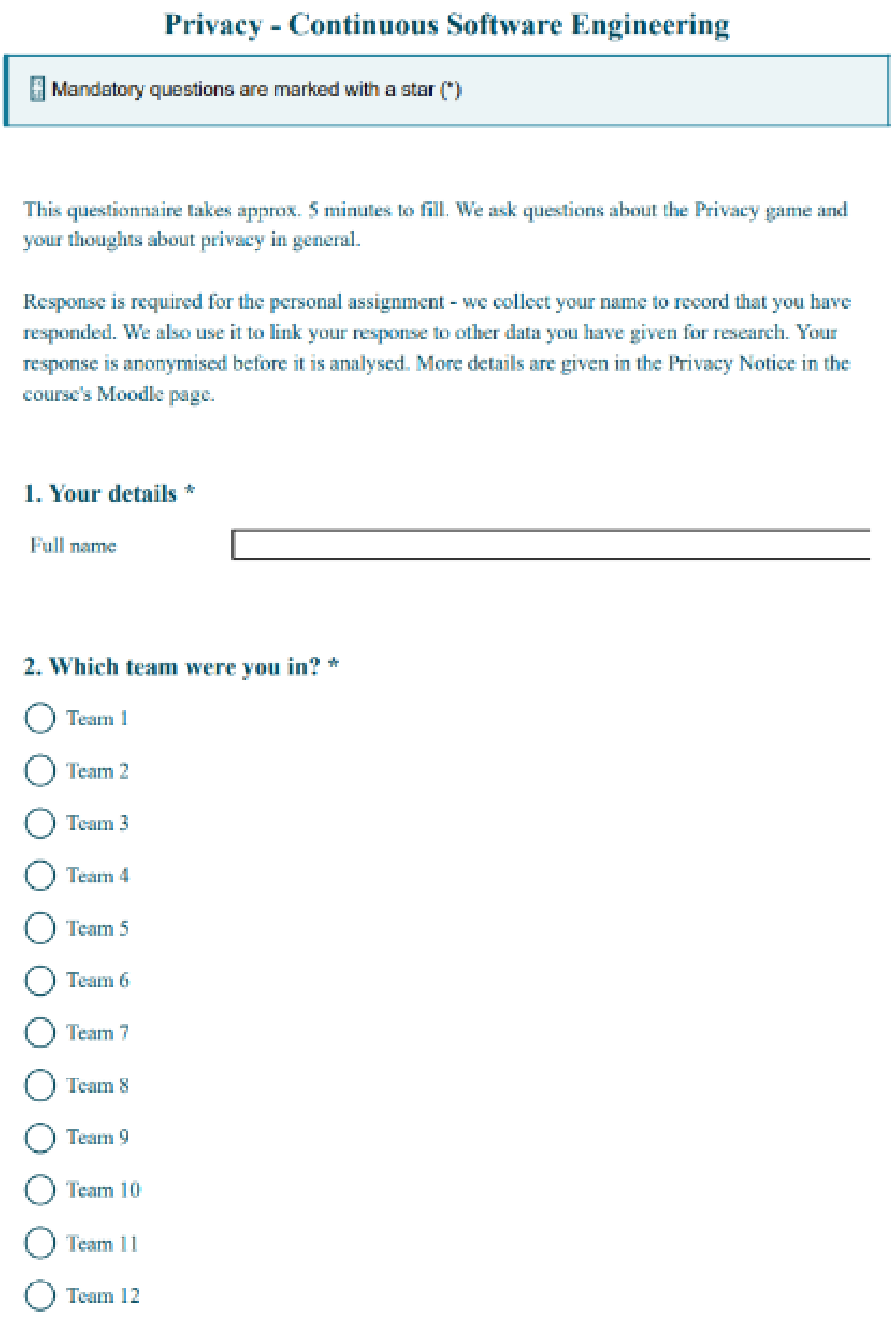
Engineering Methodologies. ACM Transactions on Privacy and Security (TOPS), 22(4). ACM Digital Library. https://dl-acm-org.ezproxy.jyu.fi/doi/abs/10.1145/3364224

Senarath, A., & Arachchilage, N. A. G. (2018). Why developers cannot embed privacy into software systems? Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering, EASE'18, 211–216. https://dl.acm.org/doi/abs/10.1145/3210459.3210484?casa_token=J-RJh3ZedI4AAAAA:Fj8gxj6vYcZDiZYw3AmhwOgd5B267veUiVqGb8RBMFVs15sB7wtvAhuzolCwwyVi83R9VXVUP9pU

Spiekermann, S., & Cranor, L. F. (2009). Engineering Privacy. IEEE Transactions on Software Engineering, 35(1), 67–82. https://doi.org/10.1109/tse.2008.88

Zainal, Z. (2007). Case study as a research method. Jurnal kemanusiaan, 5(1).

ATTACHMENTS 1 POST-COURSE QUESTIONNAIRE



Privacy - Continuous Software Engineering

Mandatory questions are marked with a star (*)

This questionnaire takes approx. 5 minutes to fill. We ask questions about the Privacy game and your thoughts about privacy in general.

Response is required for the personal assignment - we collect your name to record that you have responded. We also use it to link your response to other data you have given for research. Your response is anonymised before it is analysed. More details are given in the Privacy Notice in the course's Moodle page.

1. Your details *

Full name

2. Which team were you in? *

○ Team 1

○ Team 2

○ Team 3

○ Team 4

○ Team 5

○ Team 6

○ Team 7

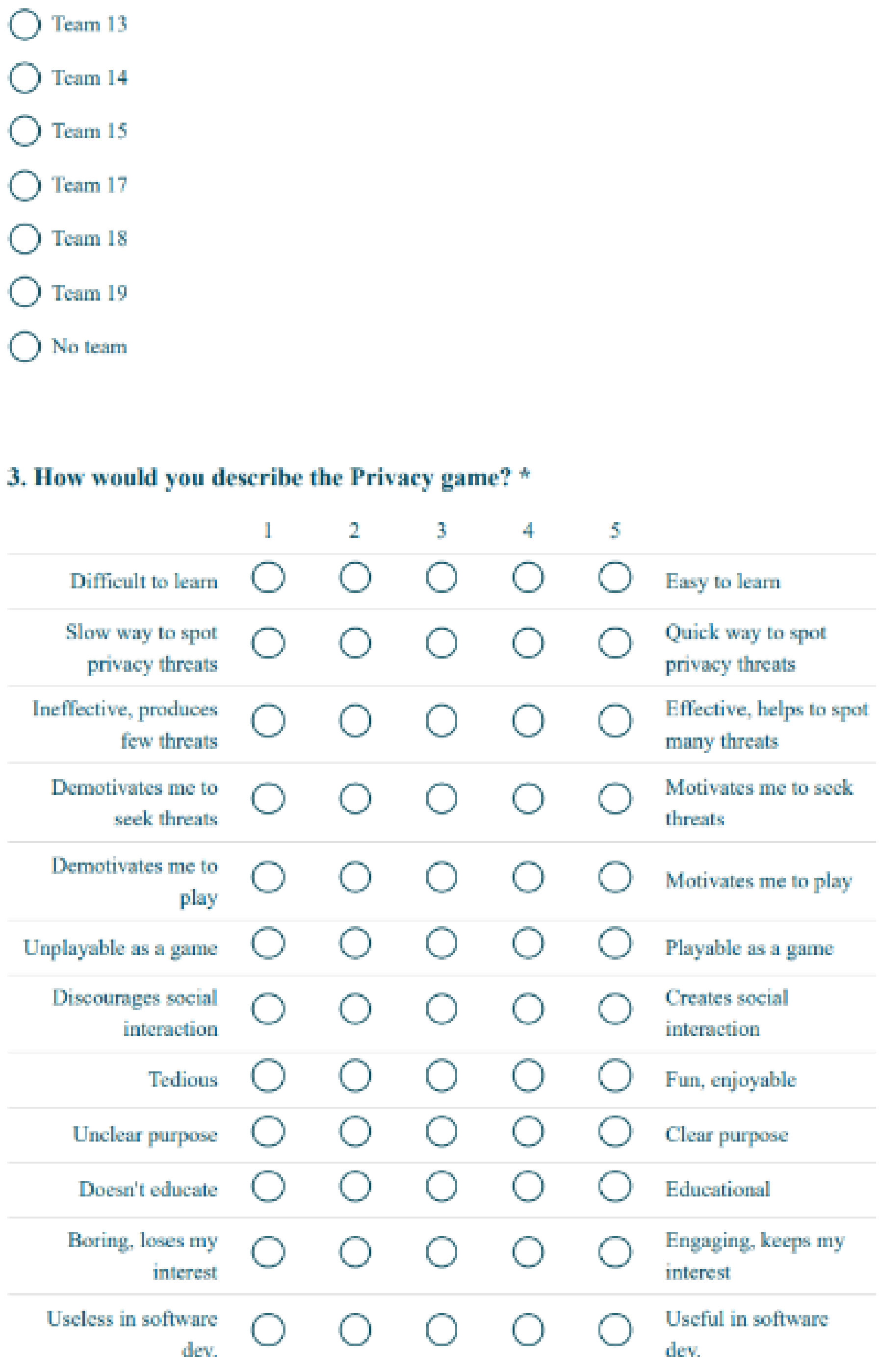
○ Team 8

○ Team 9

○ Team 10

○ Team 11

○ Team 12



○ Team 13

○ Team 14

○ Team 15

○ Team 17

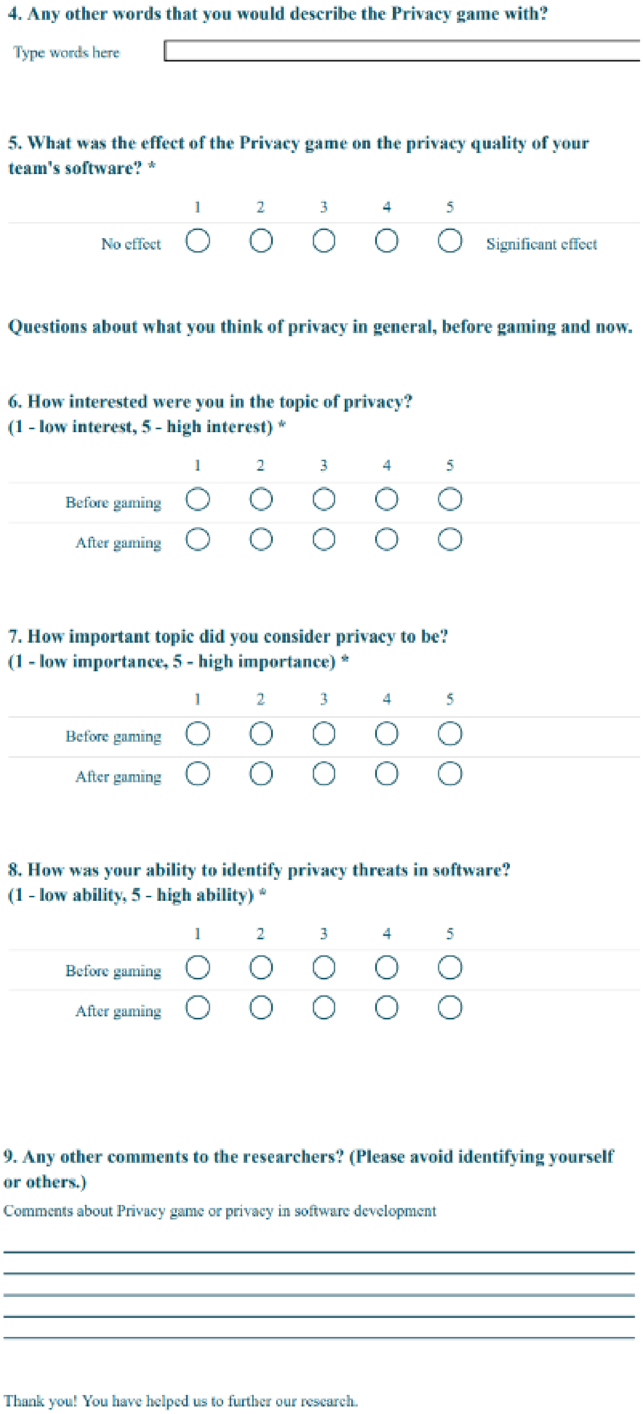
○ Team 18

○ Team 19

○ No team

3. How would you describe the Privacy game? *

	1	2	3	4	5	
Difficult to learn	○	○	○	○	○	Easy to learn
Slow way to spot privacy threats	○	○	○	○	○	Quick way to spot privacy threats
Ineffective, produces few threats	○	○	○	○	○	Effective, helps to spot many threats
Demotivates me to seek threats	○	○	○	○	○	Motivates me to seek threats
Demotivates me to play	○	○	○	○	○	Motivates me to play
Unplayable as a game	○	○	○	○	○	Playable as a game
Discourages social interaction	○	○	○	○	○	Creates social interaction
Tedious	○	○	○	○	○	Fun, enjoyable
Unclear purpose	○	○	○	○	○	Clear purpose
Doesn't educate	○	○	○	○	○	Educational
Boring, loses my interest	○	○	○	○	○	Engaging, keeps my interest
Useless in software dev.	○	○	○	○	○	Useful in software dev.



4. Any other words that you would describe the Privacy game with?

Type words here

5. What was the effect of the Privacy game on the privacy quality of your team's software? *

	1	2	3	4	5	
No effect	○	○	○	○	○	Significant effect

Questions about what you think of privacy in general, before gaming and now.

6. How interested were you in the topic of privacy?
(1 - low interest, 5 - high interest) *

	1	2	3	4	5
Before gaming	○	○	○	○	○
After gaming	○	○	○	○	○

7. How important topic did you consider privacy to be?
(1 - low importance, 5 - high importance) *

	1	2	3	4	5
Before gaming	○	○	○	○	○
After gaming	○	○	○	○	○

8. How was your ability to identify privacy threats in software?
(1 - low ability, 5 - high ability) *

	1	2	3	4	5
Before gaming	○	○	○	○	○
After gaming	○	○	○	○	○

9. Any other comments to the researchers? (Please avoid identifying yourself or others.)

Comments about Privacy game or privacy in software development

Thank you! You have helped us to further our research.