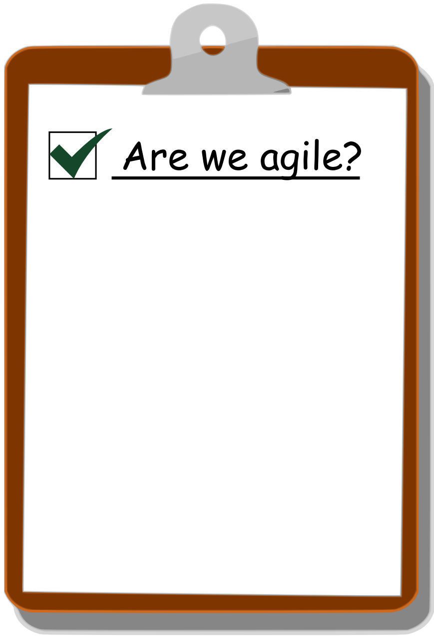
CHALMERS
UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF GOTHENBURG



Evaluating how a short survey can measure agile transformation

A field study on the survey and agile maturity model used by Volvo Cars during their transformation

Master's thesis in Computer science and engineering

Johannes Gustavsson
Pontus Lindblom

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2022

Master's thesis 2022

Evaluating how a short survey can measure agile transformation

A field study on the survey and agile maturity model used by Volvo Cars during their transformation

Johannes Gustavsson, Pontus Lindblom

UNIVERSITY OF GOTHENBURG

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Computer Science and Engineering
Chalmers University of Technology
University of Gothenburg
Gothenburg, Sweden 2022

Evaluating how a short survey can measure agile transformation
A field study on the survey and agile maturity model used by Volvo Cars during their transformation
Johannes Gustavsson, Pontus Lindblom

Supervisor: Lucas Gren, Department of Computer Science and Engineering
Advisor: Anna Sandberg, Volvo Cars
Examiner: Regina Hebig, Department of Computer Science and Engineering

Master's Thesis 2022
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: A one-question survey.

Typeset in LaTeX
Gothenburg, Sweden 2022

Evaluating how a short survey can measure agile transformation
A field study on the survey and agile maturity model used by Volvo Cars during their transformation
Johannes Gustavsson
Pontus Lindblom
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg

Abstract

With more companies moving towards agile, there is a need for ways to measure the progress. There are several agile maturity models for this purpose, but they are accompanied by a large set of indicators to be measured. Volvo Cars created a model with only ten questions in their shift towards the agile ways of working, creating an alternative and time-efficient measuring method. The purpose of this study is to evaluate to what degree one can use the 10-question survey to measure a large automotive company's agile transformation. To investigate how much of agile the model covers, a literature review was conducted to collect different agile definitions, values, and principles and then map them onto the questions. The model was validated with exploratory factor analysis and by plotting means with confidence intervals. Lastly, a thematic analysis uncovered the participants' opinions on the model.

What was found is that the survey does not cover all aspects of agile. Instead, it is composed of a mix of agile and important success factors for agile transformations. The statistical analysis did not support the proposed grouping of the questions in the model since there were too few measured variables. What could be seen is that the participants matured according to the staircase proposed. Lastly, we discovered that 8.4% responded with their opinion about the survey. Of those, 13.5% were positive, 38.2% were neutral, and 47.3% were negative. People generally had many opinions about what questions to include or exclude in the survey.

In conclusion, the scope of what can be covered with ten questions is limited. All questions except two can be connected to important aspects of agile or agile transformation, but they did not map to all included aspects in this thesis. Since all aspects of agile and the transformation cannot be covered, adapting such a survey and model to the targeted organization becomes essential. We see that grouping the questions when they are so few might hide important information. Not all aspects the employees wished for could be included with only ten questions. There also seemed to have been a dissonance between the type of questions included in the survey and what the employee expected, even though its purpose was communicated when sent out.

Keywords: Agile, Large-Scale Agile, Agile Transformation, Survey, Automotive, Agile Maturity Model, Scaled Agile Framework (SAFe).

Acknowledgements

First and foremost, we want to thank our academic supervisor Lucas Gren for his tremendous efforts in helping us during the whole process, from proposal to a finished thesis. We would also like to thank Volvo Cars and our industrial advisor Anna Sandberg for giving us access to their survey data so that we could conduct our study and for their support in our endeavor. Lastly, we would like to thank our examiner Regina Hebig for providing feedback and support at the different milestones.

Johannes Gustavsson and Pontus Lindblom, Gothenburg, June 2022

Contents

List of Figures

List of Tables

1
Introduction

The agile approach in software development has become increasingly widespread among large and small companies. Increased responsiveness to change, increased customer satisfaction, and reduced time to market are some of the benefits of agile [7].

Even though frameworks such as Scrum and Kanban exist, being agile is more than just following a framework with bundled practices [8]. It also includes a certain mindset and culture that need to be adopted within the company, and its practices [9]. The process of changing to an agile practice, mindset, and culture is called an agile transformation [10]. Having a guide through this process and measuring the progress could be helpful for a successful transformation.

There are multiple ways of measuring agility in an organization. One way to measure agile transformation is to look at the organization's agile maturity through an agile maturity model. These models guide the process with their questionnaires and maturity levels in different categories of agile aspects while letting the organization understand how far they have come in their agile transformation [11]. Even though there are multiple methods for measuring agile transformation, only a few of those tools have been statistically validated [12].

This study evaluates the possibility of measuring agile transformation through a 10-question survey. Access to the data and the model for this study was gained through Volvo Cars. The evaluation was done by first mapping the survey questions to different agile sources found in the literature. Secondly, exploratory factor analysis was applied to the answers to statistically validate the relationships between the questions in the survey and the proposed steps in the agile maturity model. To further validate the model, the mean and confidence interval of the data points were compared between the different points in time the data was collected. This comparison was made to see if the organization's maturity follows the steps proposed by Volvo Cars. Lastly, the free-text answers were analyzed to see the employees' thoughts on using the 10-question survey to measure agile transformation.

1.1 Statement of the problem

The industry wants quick ways to measure agility, but it is uncertain if such an approach is feasible. The measurement design should ensure it can measure agility in a meaningful way. At the same time, it needs to be written so that people from different disciplines answer them.

With cars becoming increasingly digital, Volvo Cars strives toward becoming a software company. Agile methods stem from software development, and most teams at Volvo Cars are now a mix of hardware and software developers since the software is central in the car. Understanding the applicability of these methods of measuring the agile transformation in such a large company is therefore important.

One challenge with measuring agile aspects is that there are many definitions of agile, and being agile according to one definition does not per se translate into being agile according to another. Even with an agreed-upon definition of agile, one still needs to understand how to measure if a team or organization is agile or not. There are several strategies to choose from when deciding how to measure agile. One way is to measure if common agile success factors, such as on-time delivery, have been improved with the agile transformation [13]. Another way is to, for example, use a tool like the Comprehensive Agility Measurement Tool (CAMT) that captures the ten most critical agility enablers and measures the agility with Likert scales, ranging from 1-5 [14].

These are just some of the available tools and frameworks, and both Gren et al. [12] and Läppenen [15] conclude in their papers that there is a need to validate these types of tools statistically.

1.2 Purpose and research questions

The purpose of this study is to evaluate to what degree one can use the 10-question survey to measure a large automotive company's agile transformation. Since the survey is short and there is no standard definition of agile, it is essential to verify that the definitions of agile used in the study can translate to an agile transformation. Being able to boil down a company's agile state into ten questions that could be sent to employees would be useful.

The agile maturity model used by Volvo Cars was statistically validated with exploratory factor analysis to see to what extent the questions correlate according to the company's intentions. The differences in the mean of the question and question pairs at the different measurement points were compared to see if the organization matures according to the staircase type levels in the model. By doing this, potential problems with the model could be exposed. By understanding those problems, that knowledge could be used to improve the process of creating a survey and model to measure agility in the future.

Lastly, participants' acceptance of the 10-question survey were explored to see how applicable it is at measuring agile transformation. A low acceptance level can indicate some underlying issues with the survey that need further evaluation. Such information can provide valuable insights about the design of this survey, which can be of use when similar surveys are designed in the future.

Based on the purpose of the study, the following research questions were formulated:
- RQ1: What parts of agility are covered in the 10-question survey?
- RQ2: Are the question pairs reflecting the maturity steps statistically?
- RQ3: What do the participants think of the 10-question agility survey?

1.3 Delimitations

The study focuses on all agile definitions that could be identified and verified in two working days when answering what parts of agility are covered in the 10-question survey. There are likely more definitions of agile available, but identifying all definitions would be out of the time scope for this study.

Analyzing the free-text answers is limited to only covering the participants' views and opinions of the survey. There were thousands of free-text answers regarding many other aspects of the agile transformation, and analyzing them further would be an overwhelming task for the scope of this study.

2
Theory

This chapter presents the theory needed to follow and understand the study. More specifically, an introduction to Waterfall, V-model, agile, large-scale agile, SAFe, agile transformation, and how to measure that transformation.

2.1 Waterfall

The Waterfall model is the oldest process model among the traditional development approaches in software development [16]. The Waterfall model is a sequential process where one phase must finish before the next phase can begin. It was first introduced by Royce back in 1970 [17], and over the years, it has evolved to take many forms. There are many variations of Waterfall available, even if the core principles remain the same. Some of the most common Waterfall variations have between five and seven phases, and the name of the phases can differ. Figure 2.1 shows an example of a Waterfall model provided by Bassil [1].

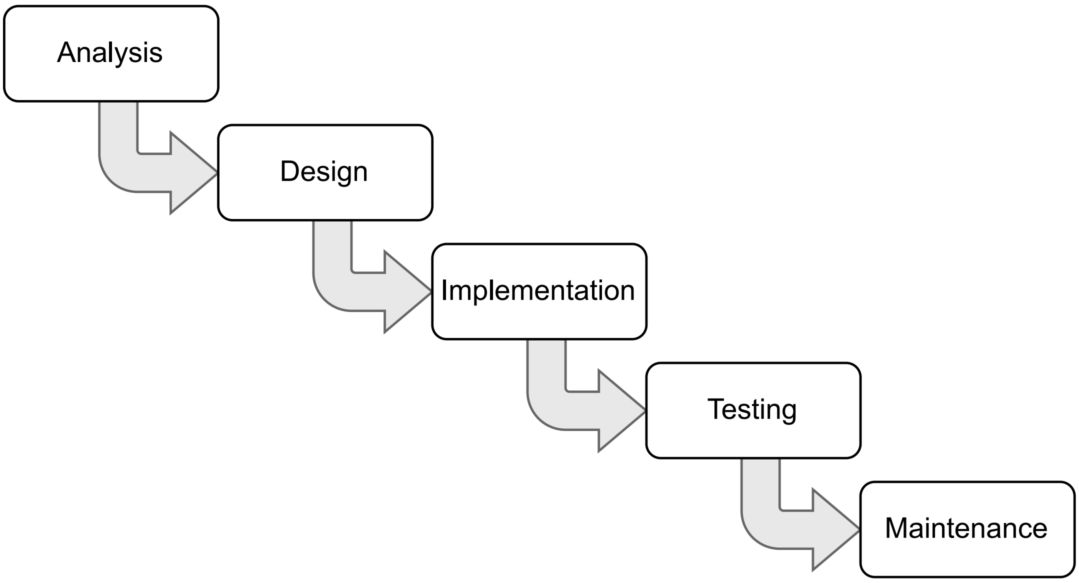


Figure 2.1: Example of a Waterfall model [1].

This specific Waterfall model from Bassil has five phases, 1) analysis, 2) design, 3) implementation, 4) testing, and 5) maintenance. A description of the different phases is shown below:

Analysis: This phase is about gathering a comprehensive description of the system's behavior that is going to be developed [1]. The customers' needs are identified and documented, and then the requirements are refined to be used in the design and implementation phase [18]. Creating clear requirements is vital as Waterfall is a sequential process, and once the design phase is reached, further changes in the requirements will not be considered [2]. The analysis phase is sometimes called the *requirement* phase.

Design: The requirements gathered from the previous phase are used to plan the design for a software solution [1]. The plan includes solutions for software architecture design, concept design, algorithm design, data structure design, graphical user interface design, and more.

Implementation: This is where the gathered requirements and created designs are converted to the production environment [1]. The implementation phase is sometimes called the *coding* or the *development* phase.

Testing: This is where the software solution is tested against the original requirements and specifications to ensure it accomplishes its intended purposes [1]. This phase is also where system glitches are found, fixed, and refined.

Maintenance: The product needs to be maintained after it has been released to the customer so that it keeps working correctly and that bugs found in the live environment are fixed [18]. Examples of maintenance tasks include error correction, performance improvements, and quality improvements.

The advantages and disadvantages of the Waterfall approach have been outlined in a paper from Balaji et al. [2]:

Advantages:

- Requirements are clear before development starts.
- Each phase is completed in a specified period before moving to the next phase.
- It is easy to implement as Waterfall is a linear model.
- Minimal resources are required to implement the model.
- The proper documentation is followed in each phase for the development quality.

Disadvantages:

- The problems in one phase are never entirely solved, and many issues arise after the phase is signed off, resulting in a poorly structured system.
- Requirements will not be changed in the current development process, even if the client requests it.

2.2 V-model

The V-model is a modification of the Waterfall model that was first proposed by Rook in 1987 [2, 19]. Like the Waterfall model, the V-model consists of many phases that one traverses from start to finish, and a phase can only start once the previous is completed. One difference between the models is that the phases of the V-model start by moving down and then bend upwards after the coding phase, as shown in Figure 2.2 [20]. The figure visualizes the relationship between the development and testing phases on the same horizontal level. In the V-model, the developer and tester work in parallel instead of doing all the tests at the end of the software development life cycle. This relationship means that acceptance tests are created while gathering the requirements. Such an approach ensures that testing is conducted at the very beginning of the project, making it possible to identify problems that could cause serious harm at a later phase [20].

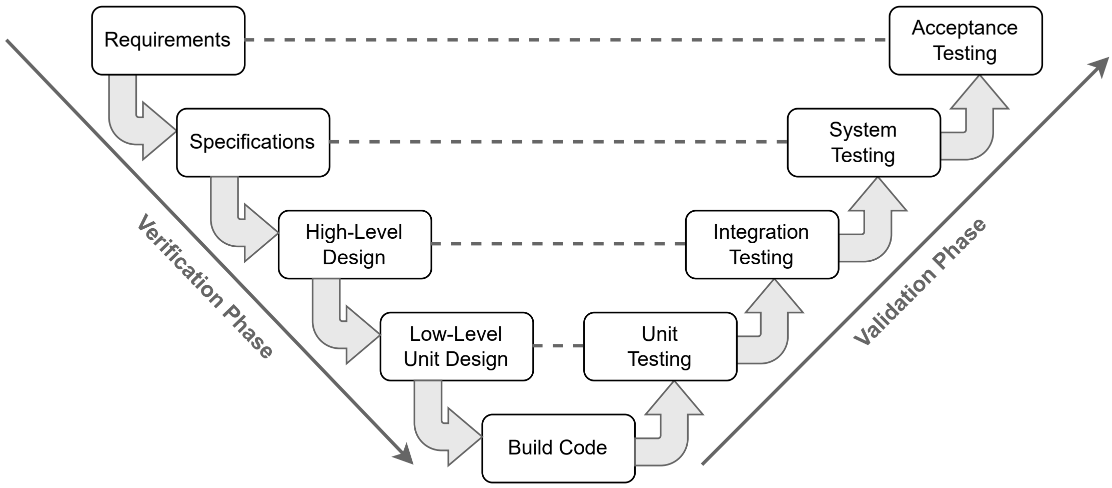


Figure 2.2: Example of a Waterfall model [2]. Dashed lines have been added between the horizontal phases to show that there is a relationship between them. Directional arrows have been added along the sides to clarify what phases are included in the verification and validation phase.

Balaji et al. outline the advantages and disadvantages of the V-model, as they did with the Waterfall model [2]:

Advantages:

- The same advantages as the Waterfall model.
- The tester role is involved in the requirements phase.
- Requirement changes are possible in any phase.

Disadvantages:

- The most significant disadvantage of the V-model is that it is very rigid and inflexible.
- If any changes happen mid-way, the requirements documents and the test documentation need to be updated.

- It is not suitable for short-term projects as it requires reviews at each stage.

The V-model was the model used at Volvo Cars before they decided to become agile. The Waterfall model and the V-model are rigid and not suitable for projects with uncertain requirements that risk changing during development [2]. Agile works very well in these situations, and one of the main reasons Volvo Cars decided to move away from the V-model to the agile way of working was that they were going too slowly compared to the changes in the automotive world [21]. There are many agile frameworks to choose from when moving to agile. As Volvo Cars was a large and complex organization, they decided to go with the large-scale agile framework SAFe (Scaled Agile Framework), which is described further in Section 2.3.2.

2.3 Agile

The agile approach in software development has become increasingly widespread among large and small companies [22]. Some of the benefits of agile are increased responsiveness to change, increased customer satisfaction, and reduced time to market [7]. The idea of agile software development was outlined in the agile manifesto in 2001 and it consists of four values (see Table 2.1) and twelve principles (see Table 2.2) [4]:

The four agile values
We are uncovering better ways of developing software by doing it and helping others do it. Through this work we have come to value: - **Individuals and interactions** over processes and tools. - **Working software** over comprehensive documentation. - **Customer collaboration** over contract negotiation. - **Responding to change** over following a plan. That is, while there is value in the items on the right, we value the items on the left more.

Table 2.1: The four agile values [4].

The twelve agile principles
- Our highest priority is to satisfy the customer through early and continuous delivery of valuable software.
- Welcome changing requirements, even late in development. Agile processes harness change for the customer's competitive advantage.
- Deliver working software frequently, from a couple of weeks to a couple of months, with a preference for the shorter timescale.
- Business people and developers must work together daily throughout the project.
- Build projects around motivated individuals. Give them the environment and support they need, and trust them to get the job done.
- The most efficient and effective method of conveying information to and - within a development team is face-to-face conversation.
- Working software is the primary measure of progress.
- Agile processes promote sustainable development. The sponsors, developers, and users should be able to maintain a constant pace indefinitely.
- Continuous attention to technical excellence and good design enhances agility.
- Simplicity–the art of maximizing the amount of work not done–is essential.
- The best architectures, requirements, and designs emerge from self-organizing teams.
- At regular intervals, the team reflects on how to become more effective, then tunes and adjusts its behavior accordingly.

Table 2.2: The twelve agile principles [4].

A few different agile frameworks have emerged, Scrum, which is the most used, and Extreme Programming (XP), to mention the two big ones [23]. There is also one method that stems from the Lean movement created at Toyota called Kanban [24]. The use of hybrid approaches is not too uncommon when looking at the data collected by digital.ai [22]. The second most used method in that report, with 7% of the respondents using it, is Scrumban, a mix of Scrum and Kanban [22].

There are many definitions of agile, and being agile according to one definition does not necessarily translate to being agile in another. To prevent a team or an organization from pulling in different agile directions, it is essential that they work with the same definition. Laanti et al. [25] collected several definitions of agile created up until 2013 when their paper was published. They conclude in their paper that the different perceptions of agility make the implementation of the agile methods complicated. A list of agile definitions in chronological order has been gathered in Table 2.3. The table includes the definitions from Laanti et al. we could verify and other definitions we found through the literature review.

Source	Agile definition
Cockburn 2001 [26]	Being effective and maneuverable. Use of light-but-sufficient rules of project behavior and the use of human and communication-oriented rules.
Highsmith 2002 [27]	Ability to both create and respond to change in order to profit in a turbulent business environment.
Anderson 2003 [28]	Ability to expedite.
Larman 2004 [29]	Rapid and flexible response to change.
Schuh 2004 [30]	Building software by empowering and trusting people. Acknowledging change as a norm, and promoting constant feedback. Producing more valuable functionality faster.
Lyytinen 2006 [31]	Discovery and adoption of multiple types of Information Systems Development innovations through garnering and utilizing agile sensing and response capabilities.
Subramaniam and Hunt 2006 [32]	Uses feedback to make constant adjustments in a highly collaborative environment.
Ambler 2007 [33]	Iterative and incremental (evolutionary) approach to software development which is performed in a highly collaborative manner by self-organizing teams with "just enough" ceremony that produces high-quality software in a cost-effective and timely manner which meets the changing needs of its stakeholders.
Kruchten 2013 [34]	The ability of an organization to react to changes in its environment faster than the rate of these changes.
Denning 2016 [35]	An umbrella term for a set of management practices–including Scrum, Kanban, and Lean– which enable offering requirements and solutions to evolve through collaboration between self-organizing, cross-functional teams. It promotes adaptive planning, evolutionary development, early delivery and continuous improvement, and it institutionalizes rapid and flexible response to customer input. Pioneered by software development teams, its ideals were first widely promulgated in the agile manifesto of 2001 [4]. Agile practices have now spread to management generally.
Gren et al. 2020 [36]	Responsiveness to change.

Table 2.3: The agile definitions chosen for this study.

Abrahamsson et al. created one definition of agile by asking, "What makes a development method an agile one?" [37]. This question creates a narrow focus on the agile connection to methods and not the whole company, and the definition is presented as the following:

- Incremental (small software releases, with rapid cycles).
- Cooperative (customer and developers working constantly together with close communication).
- Straightforward (the method itself is easy to learn and to modify, well documented).

- Adaptive (able to make last moment changes).

It is important to stress that to be agile, there is more to it than implementing a framework. “Agile is more than a set of practices used by IT requiring wide ranging change to work patterns” [8], and a key success factor in being agile is the culture [38, 22].

A recent take on the definition of agile was created by Gren et al. Their definition intends to provide an unambiguous way to discuss and measure agility, and that definition is that agile is responsiveness to change [36]. This definition has several advantages, where one of the most powerful strengths is that one can ask oneself, “Does this change make us more responsive to change?” to answer if a change makes you more agile.

2.3.1 Large-scale agile

Based on the number of teams, a definition for large-scale agile development is proposed to be agile development efforts with two or more teams [39]. To successfully scale agile beyond a single team, the organization must support communication between the teams [38]. This adds another layer of complexity compared to working within the agile team itself. Large-scale agile frameworks such as SAFe and Scrum of Scrums emerged to manage this complexity [40]. Volvo Cars chose SAFe to scale agile across their organization.

2.3.2 SAFe - Scaled Agile Framework

Scaled Agile Framework (SAFe) is a system for scaling agile across teams of teams, business units, and organizations. It implements Lean, Agile, and DevOps practices at scale. According to Scaled Agile Inc., using this approach typically results in 1) faster time-to-market, 2) improvements in quality, 3) increase in productivity, and 4) increase in employee engagement [5].

SAFe is constantly growing and evolving. It is currently in its fifth iteration, SAFe 5, and the framework is continuously updated due to increased industry expertise, more customer data, and leading-edge management practices that improve. SAFe is also the most used framework for scaled agile, where 37% of the respondents to the 15th Annual State of agile Report stated that they use SAFe [22]. SAFe is growing on a year-over-year basis, as they got 35% the year before and 30% the year before that [41]. This increase can be compared to the second most used framework for scaled agile, Scrum@Scale/Scrum of Scrums, where 9% stated they used it compared to last year when 16% said they did [41].

SAFe is built around the seven core competencies of the Lean Enterprise that enable organizations to respond to changes in the market in an agile way [3]. The core competencies are outlined in Figure 2.3.

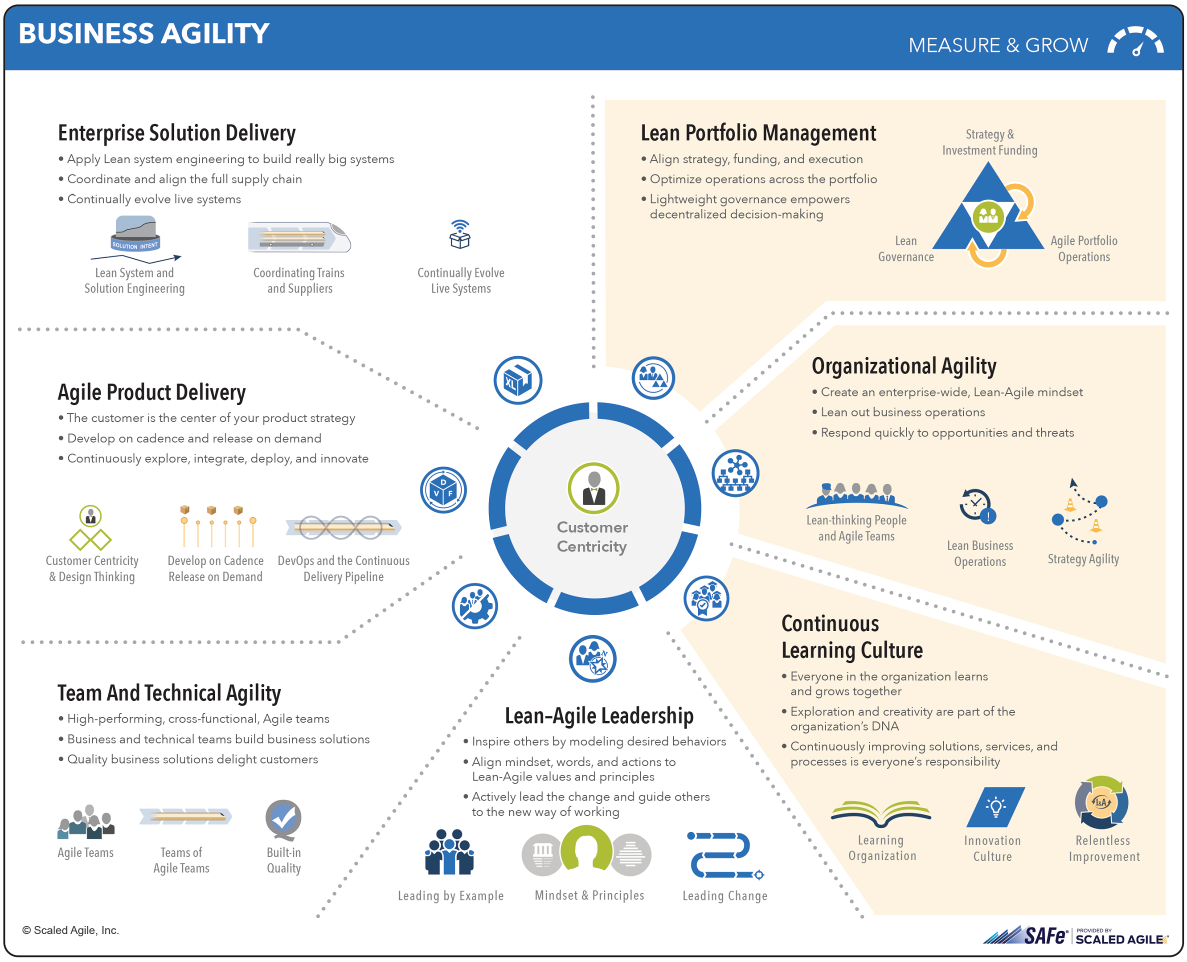


Figure 2.3: The SAFe core competencies [3].

There are four different configurations in SAFe: 1) Essential SAFe, 2) Large Solution SAFe, 3) Portfolio SAFe, and 4) Full SAFe [3]:

1. “**Essential** is the most basic configuration of the framework, and it provides the minimal elements necessary to be successful with SAFe.
2. **Large Solution** is for enterprises building large and complex solutions, which do not require the portfolio level's constructs.
3. **Portfolio** provides Strategy and Investment funding, agile Portfolio Operations, and Lean Governance.
4. **Full** is the most comprehensive configuration. It supports building large, integrated solutions that typically require hundreds of people to develop and maintain.”

An image of Full SAFe, which clearly outlines all the configurations of SAFe, can be seen in Figure 2.4.

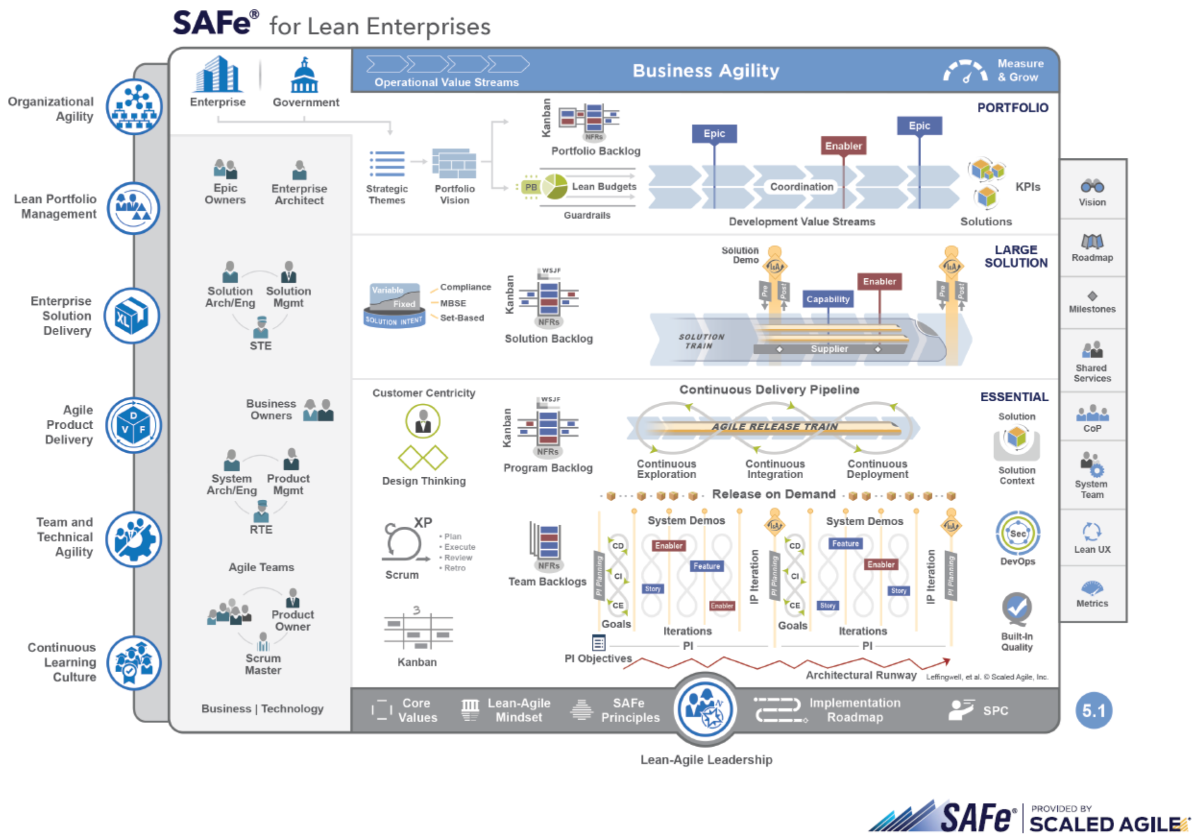


Figure 2.4: The Full SAFe configuration [3].

SAFe provides a step-by-step implementation roadmap that has been proven to be effective in successfully implementing the framework. The roadmap can be seen in Figure 2.5, and it shows the twelve steps necessary to complete the transformation.

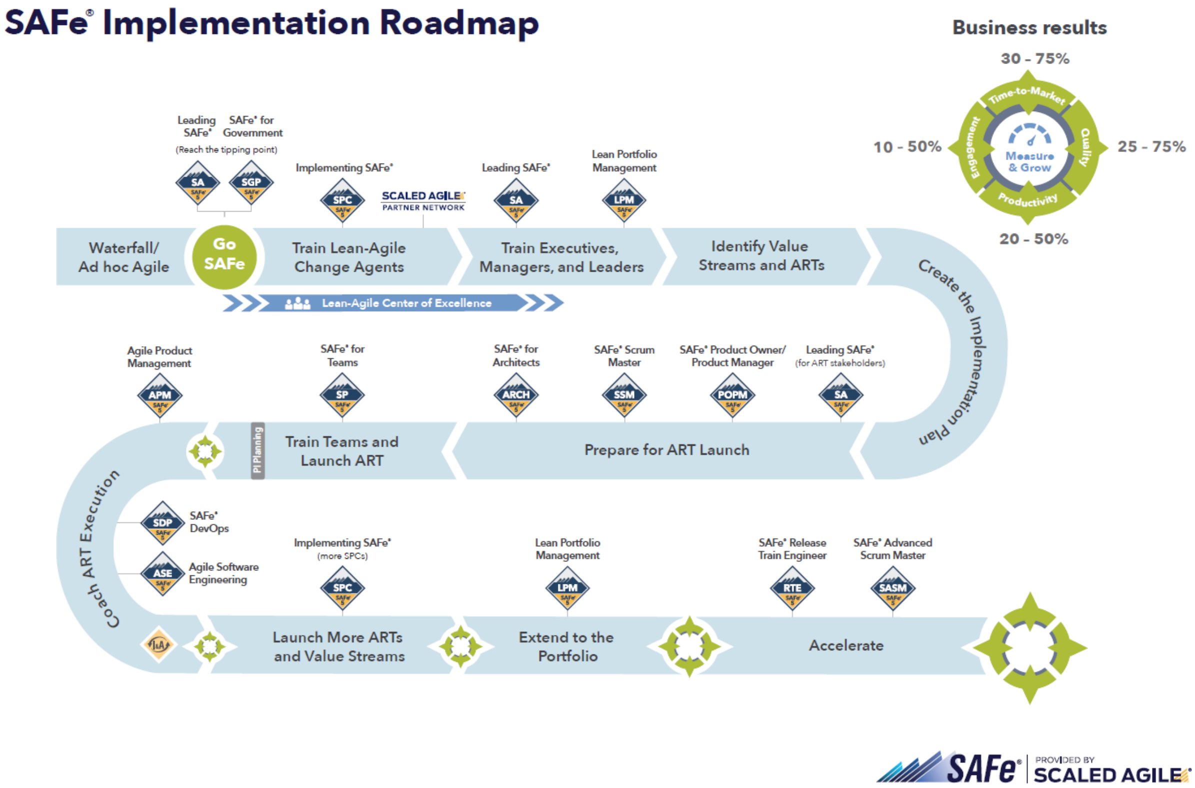


Figure 2.5: The SAFe implementation roadmap [3].

As scaled agile is complex, and all organizations are unique, there is no one-size-fits-all solution. An organization might encounter a situation where some SAFe practices do not suit their way of working, and if that is the case, SAFe provides ten principles to follow [5]. The principles are outlined in Table 2.4, and they work as guidance in the implementation of SAFe in any context.

#	**SAFe principle**
1	Take an economic view.
2	Apply systems thinking.
3	Assume variability; preserve options.
4	Build incrementally with fast, integrated learning cycles.
5	Base milestones on objective evaluation of working systems.
6	Visualize and limit WIP, reduce batch sizes, and manage queue lengths.
7	Apply cadence, synchronize with cross-domain planning.
8	Unlock the intrinsic motivation of knowledge workers.
9	Decentralize decision-making.
10	Organize around value.

Table 2.4: The ten SAFe principles [5].

2.4 Agile transformation

A paper report from the first international workshop on agile transformation [10] states that agile transformation is to implement and adopt agile practices in an organization. The paper also presents a few other definitions collected from managers, scientists, and others that attended the event:

- *"a path from adopting agile practices to establishing agile culture"*
- *"shift towards practices that enable organizational responsiveness"*
- *"agile – iterative, incremental, collaborative, effects/results/outcomes-driven transformation – continuous improvement from where you are towards the agile values and principles"*

As stated by Gandomani [42], the values in agile methods throughout the company are different from traditional ones, and therefore an agile transformation impacts the whole organization. One can further read that these transformations have multiple challenges, but the main ones are organizational culture, management, people, and processes [42].

When moving to an agile development process, there is a need to change the organizational culture from the traditional process-centric way to a people-centric one [43, 9]. No matter the size, all companies have their own organizational culture, as stated by Tolfo et al. [44]. By only looking at the visual aspects, the company can get a false image of their culture and miss potential obstacles in the planned transformation. The company needs to talk, discuss and observe people to get a proper idea of its culture. Furthermore, the authors point out that the strategic context in the company, the stakeholders with their values and principles, needs to be taken into account, as well as the tactical and operational context when evaluating the culture [44].

A critical success factor for the transformation is management buy-in and support [45]. The management team must have a logical reason for the change, create clear goals for the transformation, and be committed to the cause. Another challenge in management is that with the move to agile, there is a change in leadership style from a traditional commanding style to a servant type of leadership [45]. There is also a change in the decision-making that makes it more decentralized.

Conboy et al. [46] states numerous challenges for people during an agile transformation. One of the challenges is to deal with the fear and feeling of being inadequate that developers can get when their work becomes more transparent. Another one is that as a developer in agile, one is required to have multiple skills aside from one's niche. If the organization only looks at individual performance, employees might feel at a disadvantage when competing for a promotion [46].

Introducing agile at scale comes with its challenges. One challenge is to coordinate the large number of agile teams [38]. This coordination requires an organization that supports communication between the agile teams, which becomes even more important when the organization is distributed globally. Another challenge with agile at scale is to achieve technical consistency across all teams [38].

An efficient way of measuring the agile transformation is needed to take the correct actions and keep track of the progress. Some models for this are presented in the next section.

2.5 Measuring agile transformation

One way to measure the agility in an organization is through the quantitative metrics model proposed by Olszewska et al. [47]. This model was created to provide a quantitative approach to measuring agile impact in a software organization. It proposes four questions with eight metrics, two for each question. The metrics were chosen to be generic so that they could be used in a plan-driven, agile and lean setting. Although they also state that fine-tuning of some metrics to the specific settings would most likely be needed [47].

Agile maturity models are another way to measure agile transformation by using different categories that vary between the variety of models out there. Examples of categories used in existing models are agile principles [48], code quality metrics, and automated regression testing [49]. Within these categories, maturity levels are supposed to work like a staircase. Level one is reached first, then level two, and so on, to increase the organization's agile maturity gradually [11]. Even though there are multiple suggestions for measuring an agile transformation, there is little statistical validation for those tools [12].

An example of a model that underwent a statistical validation study after its creation is the agile adoption model created by Sidky et al. [48]. They call this model the Sidky Agile Measurement Index (SAMI), and it comes with two components, one agile measurement index, and a four-stage process. The agile measurement index has five agility levels and five agile principles, which is a summarization of the 12 agile principles from the manifesto. The four-stage process is described as the backbone of the framework by the authors. The four stages are grouped into two objectives, and the first objective contains the first stage called discontinuing factors. This objective and stage aim to provide a way to assess if the organization is ready to begin the transformation towards agile, a go/no-go assessment. The second objective is meant to guide agile coaches towards what practices fit the organization. Underneath this, the three remaining stages are found, *project level assessment*, *organizational readiness assessment*, and *reconciliation* [48].

The statistical validation on parts of this model was done by Gren et al. They concluded that "Data did not support the division of a subset of items selected from the Agile Adoption Framework. However, the data gave new categorizations of the items in the Agile Adoption Framework." [12]. They also claim that together with their changes, this was the first partly validated maturity model to their knowledge at that time.

Another example of a maturity model and one that is specified towards SAFe is the one created by Stojanov et al. [50]. The authors suggest that there was no structured roadmap for SAFe and therefore created the SAFe MM to fulfill that role. They did this by taking existing agile maturity models as a basis and then developed the model in close collaboration with industry experts through a Delphi study. The model is composed of five categories, Embrace change to deliver customer value, Plan and deliver software frequently, Human Centricity, Technical excellence, and Customer collaboration. There were five levels with varying amounts of items within each category. For example, level one in technical excellence includes Acceptance testing, Task volunteering, Knowledge sharing, and Coding standards. To evaluate the maturity, they conducted a two-hour-long interview with one scrum master and one release train engineer, which is described as a "chief scrum master." The industry experts found the model to have practical merit and were easy to understand and use but had reservations about its necessity when it was boiled down into a single variable, agile maturity. They were worried that organizations would be too occupied chasing the maturity levels instead of enhancing team performance and cooperation, which led to the researchers having a score for each practice.

3
Method

This chapter presents the research strategy used to answer the research questions of the study, how the data was collected, how the data analysis was performed, and how research ethics was taken into consideration.

3.1 Research strategy

According to Bell et al. [51], the two categories in a research strategy are qualitative and quantitative. This study uses a combination of qualitative and quantitative research, where the qualitative research was done when answering RQ1 and RQ3, and the quantitative research was done when answering RQ2. A literature review was performed to answer what parts of agility are covered in the 10-question survey (RQ1). The data points from the survey were analyzed quantitatively to answer if the question pairs reflect the maturity steps statistically (RQ2). The free-text answers were analyzed manually to identify what the participants think of the 10-question agility survey (RQ3).

The study was conducted as a field study. Stol and Fitzgerald [52] describe it as the following: "Field study refers to any research conducted in a specific, real-world setting to study a specific software engineering phenomenon." Field studies have low generalizability and a low level of obtrusiveness, which matches the approach of this study. There is low obtrusiveness in doing a literature review and analyzing already gathered data, and the study is inherently less generalizable since the data originates from one company.

3.2 Data collection

The data was collected through a literature review, surveys from Volvo Cars and their results, and other internal documents from the company to answer the research questions. The details of the data collection methods are described in this section.

3.2.1 Literature review

The literature review primarily focused on Waterfall, V-model, Agile, Scaled Agile, Agile Transformation, and Agile Maturity Models. Knowledge in these areas is essential to understand why a company decides to undergo an agile transformation, how it can be done, and how to measure it.

Data was collected from the following four sources to answer what parts of agility the 10-question survey can cover: 1) the agile values, 2) the agile principles, 3) agile definitions, and 4) the SAFe principles. Agile as it is known today started with the agile manifesto in 2001. Since the agile values and principles are part of that manifesto, it was natural to include them as agile sources. Agile has evolved since the release of the agile manifesto, and many new agile definitions have emerged [25]. Gathering some of these definitions is essential to capture the essence of how agile has evolved since the manifesto. Lastly, the SAFe principles are included because Volvo Cars made their agile transformation from the V-model to this large-scale agile framework. Including the principles provides insight into how well the survey questions map to their agile framework.

The approach of collecting agile definitions was to gather enough of them to enable a meaningful mapping of the survey questions while at the same time not collecting more than what is manageable within the time frame of this study. The source of every identified definition was verified before getting added to the collection, and if the source could not be verified, it was not added.

The literature was mainly collected through the Chalmers Library and Google Scholar databases. A technique called backward snowballing was used and is the process of going through the reference list to identify new papers, and from those reference lists, identify more papers [53]. It was used on relevant literature to identify other publications on the subject.

The keywords and phrases used to find relevant literature were: Waterfall, V-model, Agile, Scaled Agile, Scaled Agile Framework (SAFe), Agile Transformation, Agile Definition, Survey Design, Maturity Models, Confirmatory Factor Analysis and Exploratory Factor Analysis.

3.2.2 Survey

The data was collected through Volvo Cars, and they conducted the survey four times over two years, starting in 2019 and ending in 2020. The survey consisted of ten questions and a free-text answer, introduced with the text “Write your comment here.”

3.2.3 Documents

PowerPoint slides related to the studies were accessed through Volvo Cars. These slides contained the purpose behind the surveys, a background to the survey questions, and an explanation of the survey structure.

3.3 Data analysis

This section will present the methodology used to analyze the collected data. Firstly, a description of the mapping process for the ten questions to different agile sources is given. Following that, how the exploratory factor analysis (EFA) was implemented and used on the Likert scale data is explained. Lastly, the procedure of analyzing and classifying relevant results from the free-text answers is presented.

3.3.1 RQ1: Linking the survey questions to agile

The data analysis workflow for linking the survey questions to agile is found in Figure 3.1.

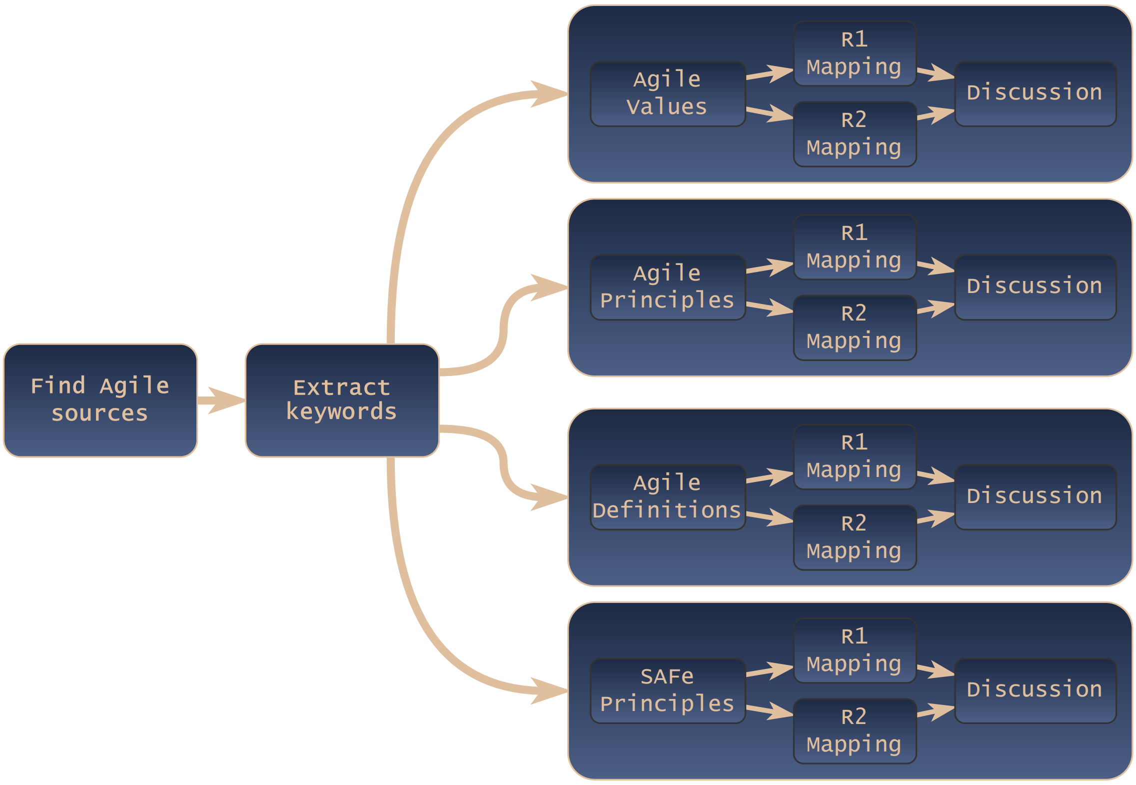


Figure 3.1: The data analysis workflow for RQ1.

The first step in the analysis is done by extracting keywords from the survey questions. These keywords were then used to map against the agile values, principles, definitions, and SAFe principles, and the authors did their mapping in parallel. When both had mapped one area, they discussed any differences between them. When deciding if an agile source mapped to the 10-question survey question, the questions and Volvo Cars' maturity model levels acted as a foundation for the decision. There were two types of connections, one where keywords were found and the other comparing the question's underlying intent with the agile value, principle, definitions, and SAFe principles. Marking the significant connections in the table was done with an X, and a parenthesis (X) is added where it was vague or the authors needed to interpret the question or object.

3.3.2 RQ2: Model validation

The data analysis workflow for the model validation is found in Figure 3.2.

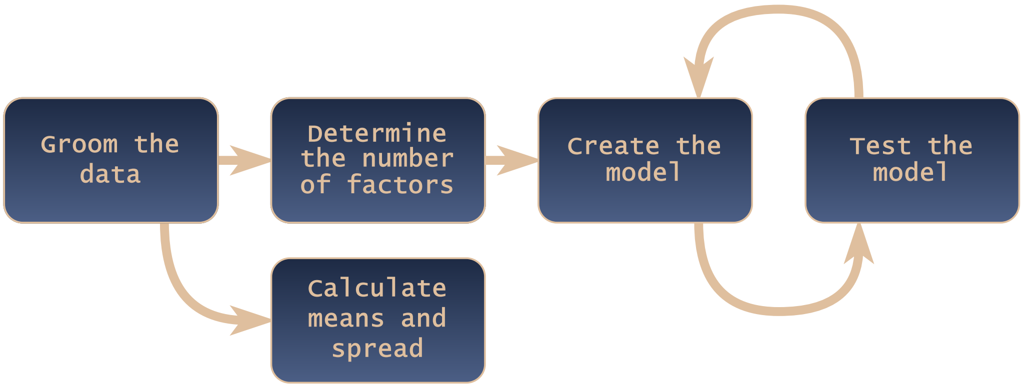


Figure 3.2: The data analysis workflow for RQ2.

For the second research question, there are two validations that together validate the maturity model proposed by Volvo Cars. The first is to statistically validate if the questions in the survey describe the created agility maturity model steps correctly. The second is to determine if the organization matures as the model proposes. This section describes why exploratory factor analysis was used and how it was implemented. It also describes how calculations of means and spread verified that the organization matured as intended.

First, the data was groomed before conducting the factor analysis. None of the survey questions were mandatory, which resulted in some data points being incomplete. As the dataset is comprehensive, list-wise deletion, where incomplete data points are removed, could be done while still having enough data for the analysis. By dividing the dataset into each survey and conducting the analysis for each survey individually, interdependency in the data was avoided.

Factor analysis is a "... set of statistical procedures designed to determine the number of distinct constructs needed to account for the pattern of correlations among a set of measures." [54]. There are three main types of factor analysis, Structural Equation Modeling (SEM), Exploratory factor analysis (EFA), and Confirmatory factor analysis (CFA).

Exploratory factor analysis is used when there is no, or only a limited, idea of the underlying structure of the correlations [54, 55]. With this method, one starts by finding how many factors to extract, and there are several ways to do this, which are described further down. When one knows how many factors to use, one runs an EFA on the data with the decided number of factors. The higher the correlation, which is called loading, is of multiple items on a single factor, the higher the chance is that they measure the same thing [54]. EFA does not tell what the factors are, that is left to the researchers to decide based on intuition and theory [54].

Confirmatory factor analysis (CFA), in contrast to EFA, is used when the researcher has a good idea, based on previous research and theory, of how the correlations are structured and what factors lie behind them [56]. Some of the use cases for CFA are psychometric evaluation of measures, construct validation, testing method effects, and testing measurement invariance [56].

Since the aim is to validate the existing model and not find a new one, CFA analysis would best fit this analysis based on the information presented above. The problem is that a prerequisite for using factor analysis is that the model should be overdetermined, which means each factor has multiple measures strongly loaded on it [56]. With the presented model from Volvo Cars, there are only two questions (measures) for each area (factor), and as such, CFA cannot be used. Therefore EFA is used to look statistically at how many factors the data support and how they group up.

The program of choice for conducting the factor analysis was JASP, and this is since it has an easy-to-use interface and is used in research. The data was split into each measurement to remove interdependence, and the analysis was run separately for each.

There are several methods to determine the number of factors, such as the Kaiser criterion, scree plot, and parallel analysis. The Kaiser criterion is to look at the eigenvalues for the correlation matrix and then choose as many factors as there are eigenvalues above one [57]. A second way is to use the scree plot [58]. Similar to the Kaiser criterion, the scree plot uses the eigenvalues of the correlation matrix, but this time they are plotted from largest to the smallest. Then the number of factors is chosen based on where the eigenvalues start to level off, forming a scree/elbow. Combining the Kaiser test and scree plot can be done by creating a horizontal line along eigenvalue one in the scree plot [58]. The third way is to use parallel analysis [59]. In this method, the eigenvalues from the correlation matrix are compared with the eigenvalues generated from simulated randomized data. The number of eigenvalues from the original correlation matrix above the ones produced by the simulated data is how many factors to retain [59]. The Kaiser test has been found to overestimate the number of factors, while the scree plot has been criticized for being too subjective [60, 61]. On the other hand, parallel analysis has been proven to have a consistent result and is what was used in this analysis [60, 62].

The two common methods for running the estimation are the Maximum-Likelihood (ML) and Principal Axis (PA). As long as the data is normally distributed, ML works well and has a more formal statistical foundation, but as nonnormality increases, it tends to overestimate [63, 64, 65]. If the data violates normality, Fabrigar et al. recommend the use of a principal factors procedure, such as principal axis factoring [63]. In the case of our data, it cannot with certainty be called normal when looking at the distribution and density plots which are included in Appendix A. Therefore Principal Axis factoring was used as the estimation method.

A rotation is applied to make the estimations easier to interpret. There are two types of rotations, orthogonal and oblique. Orthogonal rotation forces the factors to be uncorrelated [63]. All data in this survey surround the topic of agile and agile transformation. Therefore, it cannot be expected that the underlying factors are independent of each other. Thus the choice of rotation for this study was oblique, and the method Promax.

For the second part, three plots were created with the help of the R language. Firstly the mean and spread were plotted for each survey, question by question, revealing whether the organization improved between each measurement. Secondly, by plotting the result for question pairs with a confidence interval for each measuring point, the maturity levels can be evaluated, ensuring that level one increased before level two, etc. Lastly, a plot was created like the previous one, although each question was plotted separately, revealing differences inside the question pairs.

3.3.3 RQ3: Free-text analysis

The data analysis workflow for the free-text analyzation is found in Figure 3.3.

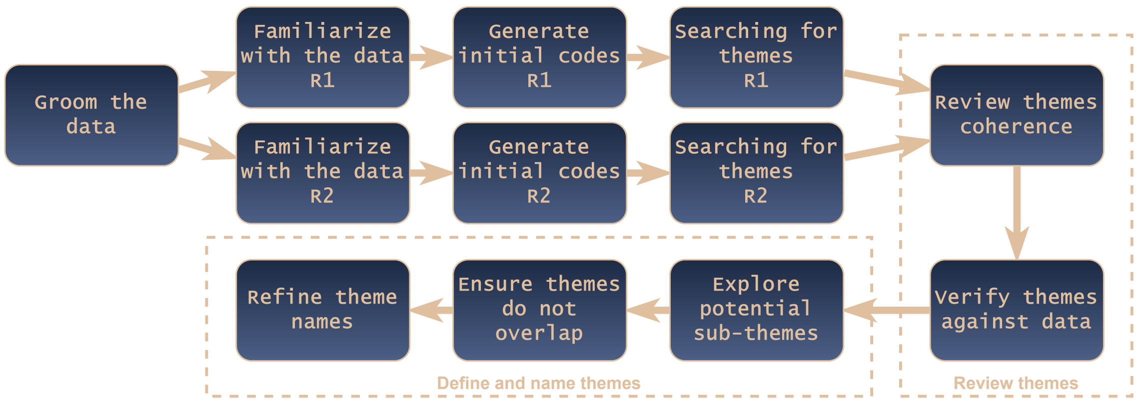


Figure 3.3: The data analysis workflow for RQ3.

The free-text answers were analyzed to answer the third research question, what do the participants think of the 10-question agility survey. Comparisons on the answers from the different measures were looked at to see changes in attitudes from the employees. If there are any changes in attitude, the free-text answers will be further analyzed to understand the reasons behind this.

The free-text answers were systematically coded with thematic analysis. Braun and Clark [66] define thematic analysis as "a method for identifying, analysing and reporting patterns (themes) within data", and the first five of their six phases were followed during the analysis. The sixth phase, *producing the report*, was not included as it merely contained guidelines on how to present the data. The removal of duplicate and empty free-text answers was made before the first phase to reduce the number of data points to be analyzed.

As described by Braun and Clark [66], the first phase was to get familiar with the dataset, which meant actively reading through the free-text answers to gather an understanding and finding patterns while taking notes. Since the dataset is extensive, the free-text answers related to the survey got marked so they could be extracted to a separate dataset. This new dataset was then used for the rest of the steps. When a basic understanding of the data has been established and an initial list of interesting insights exists, phase two can begin [66].

Further, they elaborate that in phase two, coding the essential features for each answer gives the basis for the analysis [66]. They also explain that codes are words or short sentences that summarize a piece of text. See Figure 3.1 for an example from Saad et al. [6] in their study of UX work in software startups. The coding can be done either with a theme in mind or wholly based on the data. In our case, the theme is their view on how Volvo Cars measured agility. Each answer was thoroughly read through, and codes were noted down in a document.

Original text	**Codes**
human-centered design is not fit for small companies because of **lack of resources**.	human-centered design is not fit for small companies, lack of resources
What discount **usability** and **UX methods** are promising for start-ups and Small to Medium Enterprises?	usability, UX methods
[about customer research] what to do **if feedback is negative**? How to get the information you need, **inexpensively and quickly**? **How much validation** is sufficient?	if feedback is negative, inexpensively and quickly, how much validation

Table 3.1: Example of thematic analysis codes from Saad et al. [6].

After the data were fully coded, phase three began. During the phase, the codes were grouped based on similarities that created patterns [66]. When a list of candidate themes had been established, those were to be refined, and this is what Braun and Clark [66] call phase four. They divide the phase into two levels of refinement: Level one is where the codes inside a theme are inspected, identifying if they coherently support the theme. Level two is similar, but the level of abstraction is increased to looking at the themes compared to the dataset instead of the codes [66].

During this phase, if a theme had too few data points supporting the theme or if the data were too diverse, the theme was either removed altogether or merged with another, creating a more high-level theme. If there was support for it, some themes were broken down further to smaller ones.

During level 2, the themes and their relation to the data were inspected by reading through the dataset. A few new codes can also be generated during this read-through if they have been missed in the earlier stages [66]. The result of this phase was a more defined thematic map.

The last phase is to refine the names to capture the essence of what the data describes [66]. Going through the data for each theme and figuring out what data is interesting and why, ensures that the themes align with the research questions and do not overlap too much [66]. The relations between the themes are therefore crucial to consider in this phase. During this phase, exploring potential sub-themes that emerged gave an easier understanding of complex themes [66]. This phase resulted in a more defined thematic map and clearly defined themes.

3.4 Research ethics

The only data that was kept and analyzed were the data points, the free-text answers, and information on which survey the data belonged. All other data was disregarded to keep the participants as anonymous as possible. To further strengthen the anonymity, no data point was analyzed in isolation and presented in the study, and no free-text answer was quoted. The high number of data points, in combination with the data anonymization steps above, makes it hard, if not impossible, to identify the participants of the study.

4
Results

This chapter presents the results from our data collection and the analysis of the different research questions. Firstly, the surveys and internal documentation from Volvo Cars are presented. In RQ1, tables are provided showing how the questions relate to literature. In RQ2, plots and tables show if there is statistical support for the model. Lastly, the RQ3 section includes our findings from conducting the thematic analysis.

4.1 Results of the data collection

Looking at the data obtained from Volvo Cars, the questions used in their 10-question surveys were the following:

1	I understand why my team needs to move towards Agile ways of working.
2	I understand why Volvo Cars Product Creation is moving towards agile ways of working.
3	In my team, we regularly reflect and learn to improve.
4	Leaders around my team support our improvement efforts.
5	In my team, we understand how to move towards agile ways of working.
6	I have enough knowledge about agile ways of working to do my job.
7	In my team, we are transparent with our real-time progress and status.
8	I can find real-time information needed for my work.
9	In my team we pull from a prioritized backlog according to our capacity.
10	Leaders around my team avoid pushing work over our capacity.

Table 4.1: Survey questions used by Volvo Cars to measure agility.

One question was refined, and the order of the questions was changed between the first and second surveys. In Q_2, the word “P&Q” (Product Creation and Quality) was replaced with “Volvo Cars Product Creation” as P&Q was changed after a re-organization in the company. Q_9 and Q_{10} were moved to Q_3 and Q_4, pushing the questions two steps, making Q_3-Q_8 becoming Q_5-Q_{10}. The questions were changed so that the results from the first survey showed a staircase maturity model.

Each question was measured with Likert scales of 1-6, ranging from strongly agree to strongly disagree, and there was one alternative when none of the levels were applicable. Shown below are the different response alternatives:

1. Strongly disagree

2. Disagree
3. Somewhat disagree
4. Somewhat agree
5. Agree
6. Strongly agree
- Not applicable to me

The survey was designed according to Volvo Cars' maturity model, where every pair of two questions represents one level in a maturity staircase. This design resulted in five maturity levels, and the backgrounds of the levels are shown below:

- (Q_1, Q_2) Why go agile
  - To really embrace agile we need to understand why it is important for us and understand what it means in our team.
- (Q_3, Q_4) Continuous learning
  - We need to reflect and continuously improve, and we need management to enable such efforts.
- (Q_5, Q_6) Prerequisites to go agile
  - For us to be able to move to agile we need to have knowledge and know how to change.
- (Q_7, Q_8) Real-time transparency
  - Agile ways of working are dependent on having data available in real-time, and we also need to share our own data in the same way for others to see.
- (Q_9, Q_{10}) Leveraging agile
  - Obtaining empowered teams implies we pull tasks from a prioritizes backlog and that no one else in the organization pushes us over our capacity.

The survey data consisted of 14,580 data points in total. Table 4.2 shows the number of data points in each survey and how much data remained after a list-wise deletion. Table 4.3 shows the number of free-text answers (including empty answers) and how many remained after the list-wise deletion and duplicate deletion. After the data grooming process, the remaining free-text answers were much less than the remaining data points. The average percentage of remaining free-text answers was 26.0%, compared to 88.5% of the data points.

Survey	**Data points originally**	**Data points after list-wise deletion**	**% of remaining data points**
#1	3,107	2,711	87.3%
#2	3,767	3,261	86.6%
#3	4,046	3,671	90.7%
#4	3,660	3,261	89.1%
#1-4	14,580	12,905	88.5%

Table 4.2: The number of data points originally and after list-wise deletion.

Survey	Free-text answers ***originally***	Free-text answers after ***list-wise deletion***	Free-text answers after ***duplicate deletion***	**% of remaining free-text answers**
#1	3,107	602	598	19.2%
#2	3,767	774	759	20.1%
#3	4,046	1,465	1,423	35.2%
#4	3,660	1,023	1,014	27.7%
#1-4	14,580	3,864	3,794	26.0%

Table 4.3: The number of free-text answers originally, after list-wise deletion, and after duplicate deletion.

4.2 Analysis

In this section, the research questions will be analyzed, based on the results, separately in their own subcategory.

4.2.1 RQ1: What parts of agility are covered in the 10-question survey?

The extracted keywords from the survey questions are shown below in Table 4.4.

Question	Keywords
I **understand why my team** needs to **move towards agile** ways of working.	understand why, my team, move towards agile
I **understand why** Volvo Cars **Product Creation** is **moving towards agile** ways of working.	understand why, Product Creation, moving towards agile
In my team, we **regularly reflect** and **learn to improve**.	regularly reflect, learn to improve
Leaders around my team **support** our **improvement efforts**.	Leaders, support, improvement efforts
In my team, we **understand how** to **move towards agile** ways of working.	understand how, move towards agile
I have **enough knowledge about agile** ways of working to do my job.	enough knowledge about agile
In my team, we are **transparent** with our **real-time progress and status**.	transparent, real-time progress and status
I can **find real-time information** needed for my work.	find real-time information
In my team we **pull** from a **prioritized backlog according to our capacity**.	pull prioritized backlog, according to our capacity
Leaders around my team **avoid pushing** work **over our capacity**.	Leaders, avoid pushing, over our capacity

Table 4.4: The keywords extracted from the survey questions.

The rest of this section presents the extracted keywords from the survey questions and the results from the mapping process. Table 4.5 is empty since no connections between the questions asked in the survey and the agile values could be found.

Agile values	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8	Q_9	Q_{10}
Individuals and interactions over processes and tools.										
Working software over comprehensive documentation.										
Customer collaboration over contract negotiation.										
Responding to change over following a plan.										

Table 4.5: Mapping of the agile values (Table 2.1) to the survey questions (Table 4.1).

Looking instead at Table 4.6, one can see that some connections could be made here. Out of the connections, two could clearly be seen, Q_3 towards principle 12 and Q_4 towards principle five. Q_4 also connected towards principle 12 but vaguely. Lastly, Q_9 and Q_{10} vaguely mapped towards principle eight.

Agile principle	**Q_1**	**Q_2**	**Q_3**	**Q_4**	**Q_5**	**Q_6**	**Q_7**	**Q_8**	**Q_9**	**Q_{10}**
Our highest priority is to satisfy the customer through early and continuous delivery of valuable software.										
Welcome changing requirements, even late in development. Agile processes harness change for the customer's competitive advantage.										
Deliver working software frequently, from a couple of weeks to a couple of months, with a preference for the shorter timescale.										
Business people and developers must work together daily throughout the project.										
Build projects around motivated individuals. Give them the environment and support they need, and trust them to get the job done.				**X**						
The most efficient and effective method of conveying information to and - within a development team is face-to-face conversation.										
Working software is the primary measure of progress.										
Agile processes promote sustainable development. The sponsors, developers, and users should be able to maintain a constant pace indefinitely.									**(X)**	**(X)**
Continuous attention to technical excellence and good design enhances agility.										
Simplicity–the art of maximizing the amount of work not done–is essential.										
The best architectures, requirements, and designs emerge from self-organizing teams.										
At regular intervals, the team reflects on how to become more effective, then tunes and adjusts its behavior accordingly.			**X**	**(X)**						

Table 4.6: Mapping of the agile principles (Table 2.2) to the survey questions (Table 4.1).

In mapping the questions towards agile definitions, no strong connections could be made but a few vaguer ones, which can be seen in Table 4.7. Both Q_3 and Q_4 mapped towards the definition of Schuh et al. [30]. Q_3 also mapped to both Subramaniam and Hunt [32] and Denning et al. [35] definitions.

Agile definition	$\mathbf{Q_1}$	$\mathbf{Q_2}$	$\mathbf{Q_3}$	$\mathbf{Q_4}$	$\mathbf{Q_5}$	$\mathbf{Q_6}$	$\mathbf{Q_7}$	$\mathbf{Q_8}$	$\mathbf{Q_9}$	$\mathbf{Q_{10}}$
Cockburn 2001 [26]										
Highsmith 2002 [27]										
Anderson 2003 [28]										
Larman 2004 [29]										
Schuh 2004 [30]			**(X)**	**(X)**						
Lyytinen 2006 [31]										
Subramaniam and Hunt 2006 [32]			**(X)**							
Ambler 2007 [33]										
Kruchten 2013 [34]										
Denning 2016 [35]			**(X)**							
Gren et al. 2020 [36]										

Table 4.7: Mapping of the agile definitions (Table 2.3) to the survey questions (Table 4.1).

Lastly, four questions mapped towards the SAFe principles [5]. One can see in Table 4.8 that Q_3 mapped strongly with principle four while Q_7, Q_9, and Q_{10} mapped more vaguely with principle six.

SAFe principle	$\mathbf{Q_1}$	$\mathbf{Q_2}$	$\mathbf{Q_3}$	$\mathbf{Q_4}$	$\mathbf{Q_5}$	$\mathbf{Q_6}$	$\mathbf{Q_7}$	$\mathbf{Q_8}$	$\mathbf{Q_9}$	$\mathbf{Q_{10}}$
Take an economic view.										
Apply systems thinking.										
Assume variability; preserve options.										
Build incrementally with fast, integrated learning cycles.			**X**							
Base milestones on objective evaluation of working systems.										
Visualize and limit WIP, reduce batch sizes, and manage queue lengths.							**(X)**		**(X)**	**(X)**
Apply cadence, synchronize with cross-domain planning.										
Unlock the intrinsic motivation of knowledge workers.										
Decentralize decision-making.										
Organize around value.										

Table 4.8: Mapping of the SAFe principles (Table 2.4) to the survey questions (Table 4.1).

A summary of which agile sources that did not map to any of the survey questions is shown in Table 4.9.

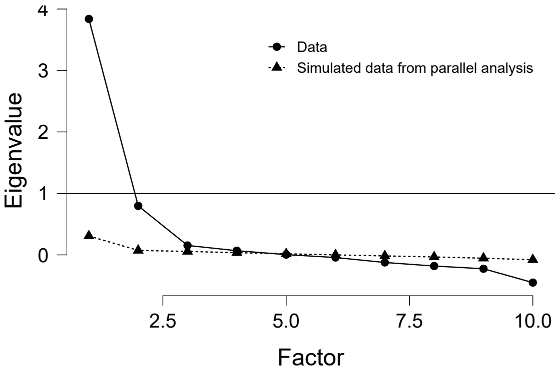
Agile source	**Did not map to any survey question**	**Statistics**
Agile values	1-4	4 of 4 (100%) did not map to any survey question
Agile principles	1-4, 6-7, 9-11	9 of 12 (75%) did not map to any survey question
Agile definitions	1-4, 6, 8-9, 11	8 of 11 (72.7%) did not map to any survey question
SAFe principles	1-3, 5, 7-10	8 of 10 (80%) did not map to any survey question

Table 4.9: The agile sources that did not map to any of the survey questions.

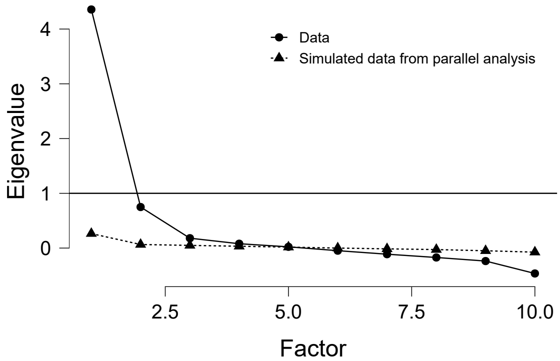
4.2.2 RQ2: Are the question pairs reflecting the maturity steps statistically?

In this section, the results of RQ2 will be presented. It will be presented with graphs such as a scree plot and text describing the outcome.

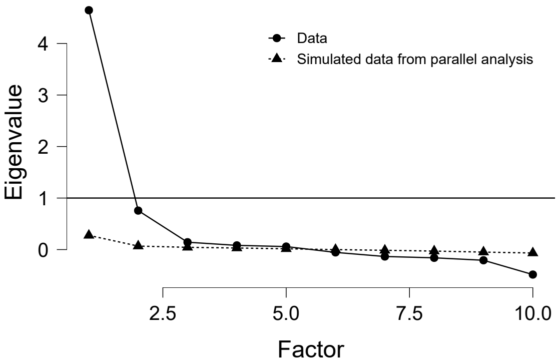
Firstly the exploratory factor analysis was conducted. The parallel analysis results, as can be seen in Figure 4.1, support four factors for measurement one, five factors for measurement two, five factors for measurement three, and five factors for measurement four.



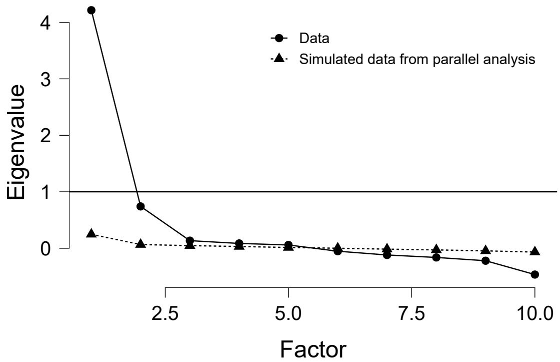
(a) First Measurement



(b) Second Measurement



(c) Third Measurement



(d) Fourth Measurement

Figure 4.1: Scree plots from the different measurement points paired with parallel analysis used to determine the numbers of statistically supported factors.

The result from the factor correlation matrix, see Table 4.10, 4.11, 4.12, and 4.13, reveals that our assumption that the factors are correlated was correct, and thus the decision to use oblique rotation is statistically supported. The correlations range from 0.304 up to 0.803 where most lie around 0.6.

	Factor 1	Factor 2	Factor 3	Factor 4
Factor 1	1.000	0.381	0.687	0.684
Factor 2	0.381	1.000	0.558	0.342
Factor 3	0.687	0.558	1.000	0.495
Factor 4	0.684	0.342	0.495	1.000

Table 4.10: Factor correlations measurement one.

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Factor 1	1.000	0.531	0.663	0.451	0.304
Factor 2	0.531	1.000	0.796	0.683	0.618
Factor 3	0.663	0.796	1.000	0.580	0.506
Factor 4	0.451	0.683	0.580	1.000	0.540
Factor 5	0.304	0.618	0.506	0.540	1.000

Table 4.11: Factor correlations measurement two.

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Factor 1	1.000	0.432	0.641	0.507	0.376
Factor 2	0.432	1.000	0.759	0.774	0.672
Factor 3	0.641	0.759	1.000	0.791	0.609
Factor 4	0.507	0.774	0.791	1.000	0.679
Factor 5	0.376	0.672	0.609	0.679	1.000

Table 4.12: Factor correlations measurement three.

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Factor 1	1.000	0.470	0.626	0.332	0.417
Factor 2	0.470	1.000	0.803	0.683	0.681
Factor 3	0.626	0.803	1.000	0.616	0.632
Factor 4	0.332	0.683	0.616	1.000	0.609
Factor 5	0.417	0.681	0.632	0.609	1.000

Table 4.13: Factor correlations measurement four.

When looking at the factor loadings table, since an oblique rotation was used, it is important to understand that 0.9 does not say that it relates 90% to that factor. Rather it should be seen as the standardized unit of increase in the measured variable for each standardized unit of increase in the common factor [54]. A cutoff at 0.4 was chosen for the rotated factor loadings to simplify the interpretation of the factor analysis table. Different sources claim that different values should be used, from 0.3 to 0.7, but 0.4 seems to be the most common [67].

In Table 4.14 the factor loading table for measurement one is presented. This first table is the only one that only supports four factors and puts Q_3, Q_4, Q_7, and Q_8 in the same factor. What also is shown in the table is that Q_9 lacks enough support in any of the factors. Q_1 and Q_2 are grouped in factor two, Q_5 and Q_6 are grouped in factor three, and Q_{10} is alone in factor four.

	Factor 1	Factor 2	Factor 3	Factor 4	Uniqueness
q1		0.875			0.127
q2		0.829			0.273
q3	0.823				0.433
q4	0.618				0.484
q5			0.592		0.360
q6			0.523		0.589
q7	0.640				0.493
q8	0.432				0.614
q9					0.537
q10				0.885	0.309

Table 4.14: Factor loadings measurement one.

In measurement two, the factor loadings table, see Table 4.15, shows that Q_1 and Q_2 highly loaded onto factor one. Q_3, Q_7, and Q_8 load onto factor two, while Q_5 alone loads onto factor three. Q_6 does not load properly onto any of the factors. Q_9 and Q_{10} nicely load onto factor four, and lastly, Q_4 loads onto factor five.

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Uniqueness
q1	0.937					0.169
q2	0.988					0.178
q3		0.541				0.402
q4					0.775	0.176
q5			1.108			0.028
q6						0.591
q7		0.995				0.361
q8		0.493				0.545
q9				0.422		0.430
q10				0.810		0.388

Table 4.15: Factor loadings measurement two.

Looking at the third measurement, the table, see Table 4.16, creates a diagonal line with Q_1 and Q_2 once again mapping strongly against their own factor, factor one. Q_3 and Q_4 load onto factor two, while Q_5 once again has its own factor. Both Q_6 and Q_9 are not loaded enough onto anything to be represented in the table. Q_7 and Q_8 pair together into factor two. Lastly, Q_{10} sits alone in factor five.

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Uniqueness
q1	0.904					0.185
q2	0.990					0.136
q3		1.066				0.126
q4		0.475				0.405
q5			1.091			0.106
q6						0.570
q7				0.528		0.425
q8				1.020		0.290
q9						0.468
q10					0.990	0.188

Table 4.16: Factor loadings measurement three.

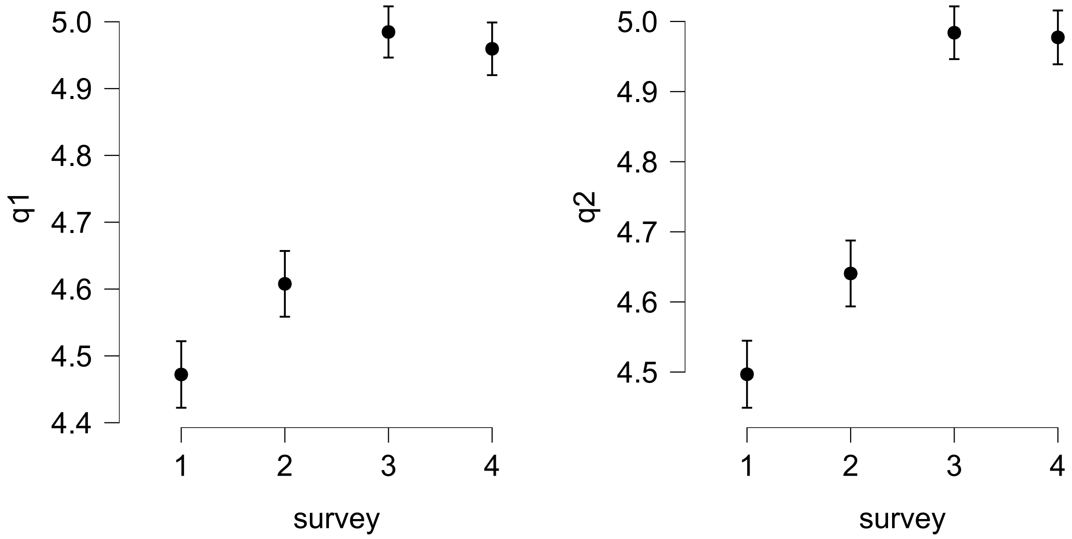
In the last measurement the factor loadings table, see Table 4.17, {Q_1, Q_2}, {Q_3, Q_4} and {Q_7, Q_8} forms one factor each. Q_6 and Q_9 once again do not have enough loading onto any factor to show in the table, while Q_5 and Q_{10} are alone to load onto their factors.

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Uniqueness
q1	0.870					0.218
q2	0.962					0.202
q3				0.411		0.387
q4				0.800		0.230
q5			0.991			0.171
q6						0.657
q7		1.013				0.308
q8		0.528				0.558
q9						0.474
q10					0.887	0.300

Table 4.17: Factor loadings measurement four.

When using an oblique rotation, a structure matrix is provided. Analyzation of these matrixes is generally not done in research, and therefore, due to time constraints, no deeper examination of them was performed [54]. Although they are included in Appendix B.

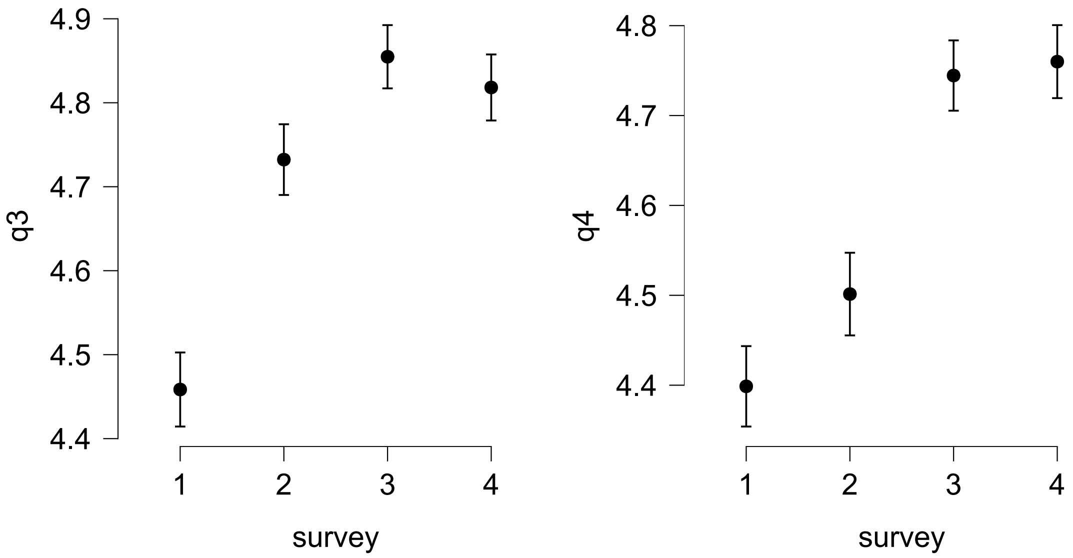
Figure 4.2, with interval plots of Q_1 and Q_2, shows that the results for survey two and three are significant compared to their previous survey.



(a) Interval plot on Q_1. (b) Interval plot on Q_2.

Figure 4.2: Interval plots on Q_1 and Q_2 across the four surveys.

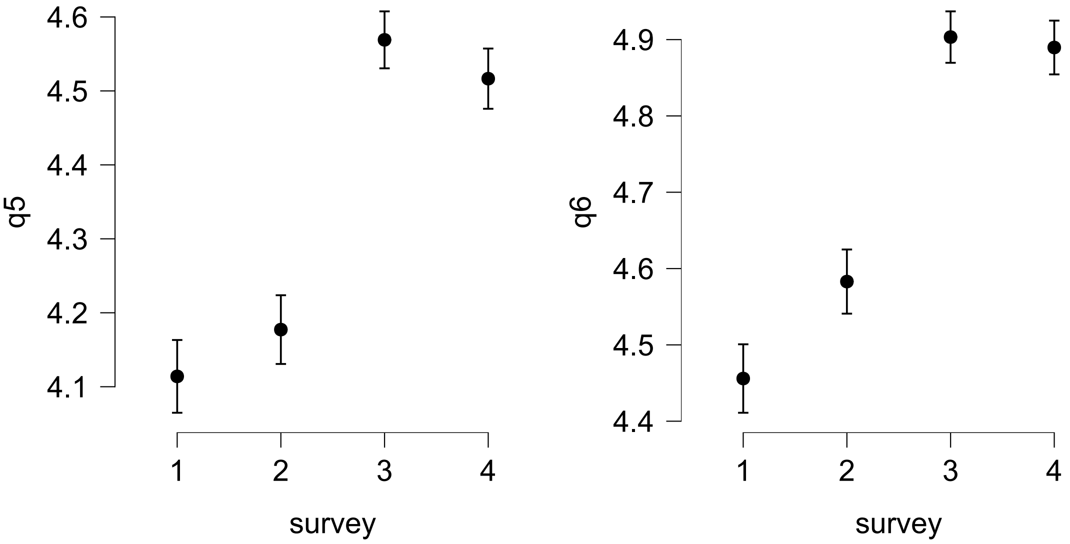
Figure 4.3, with interval plots of Q_3 and Q_4, shows that the results for survey two and three are significant compared to their previous survey. It is noted that the pattern in every question is quite similar, with the exception of Q_3 in survey two.



(a) Interval plot on Q_3. (b) Interval plot on Q_4.

Figure 4.3: Interval plots on Q_3 and Q_4 across the four surveys.

Figure 4.4, with interval plots of Q_5 and Q_6, shows that the results for survey three are significant compared to their previous survey. The second survey was only significant compared to its previous survey for Q_6, not for Q_5.

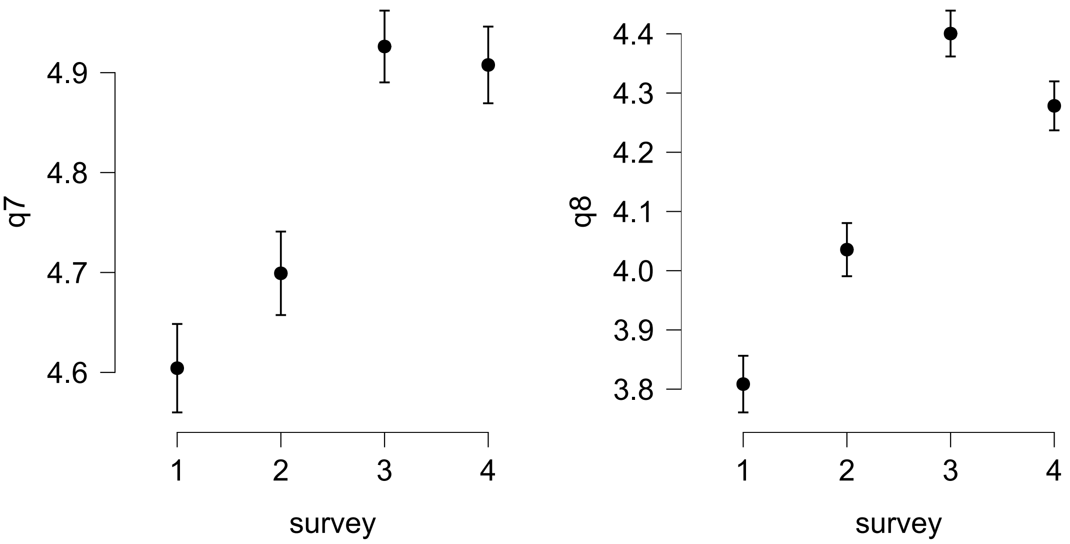


(a) Interval plot on Q_5.

(b) Interval plot on Q_6.

Figure 4.4: Interval plots on Q_5 and Q_6 across the four surveys.

Figure 4.5, with interval plots of Q_7 and Q_8, shows that the results for survey two and three are significant compared to their previous survey. The fourth survey was significant compared to its previous survey for Q_8, not for Q_7.

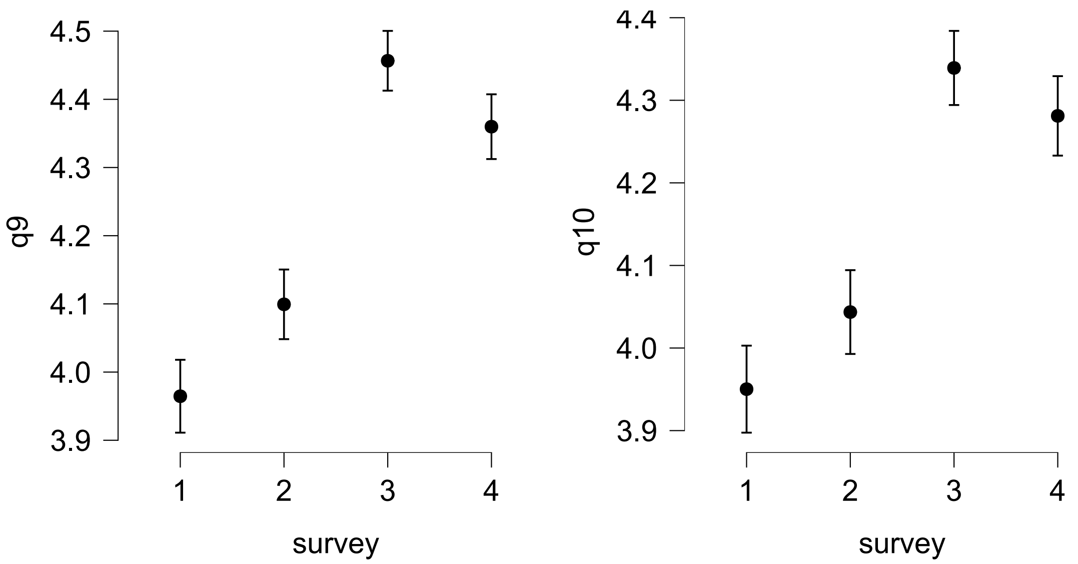


(a) Interval plot on Q_7.

(b) Interval plot on Q_8.

Figure 4.5: Interval plots on Q_7 and Q_8 across the four surveys.

Figure 4.6, with interval plots of Q_9 and Q_{10}, shows that the results for survey two, three, and four is significant compared to their previous survey in Q_9. In Q_{10}, only the third survey showed a significant result.



(a) Interval plot on Q_9. (b) Interval plot on Q_{10}.

Figure 4.6: Interval plots on Q_9 and Q_{10} across the four surveys.

Grouping the questions as intended in the proposed model and calculating the mean and 95% confidence interval will reveal if the organization matured as intended. Looking at Figure 4.7 and the zoomed in version 4.8, the question groupings generally mature quite well according to the model. In survey one, there is no significant difference between Q_1-Q_2 and Q_3-Q_4, while in survey two, both Q_1-Q_2 and Q_3-Q_4, as well as Q_5-Q_6 and Q_7-Q_8, are on the same level. Looking at the numbers which can be found in Appendix C and Table C.1, the difference between them is negligible.

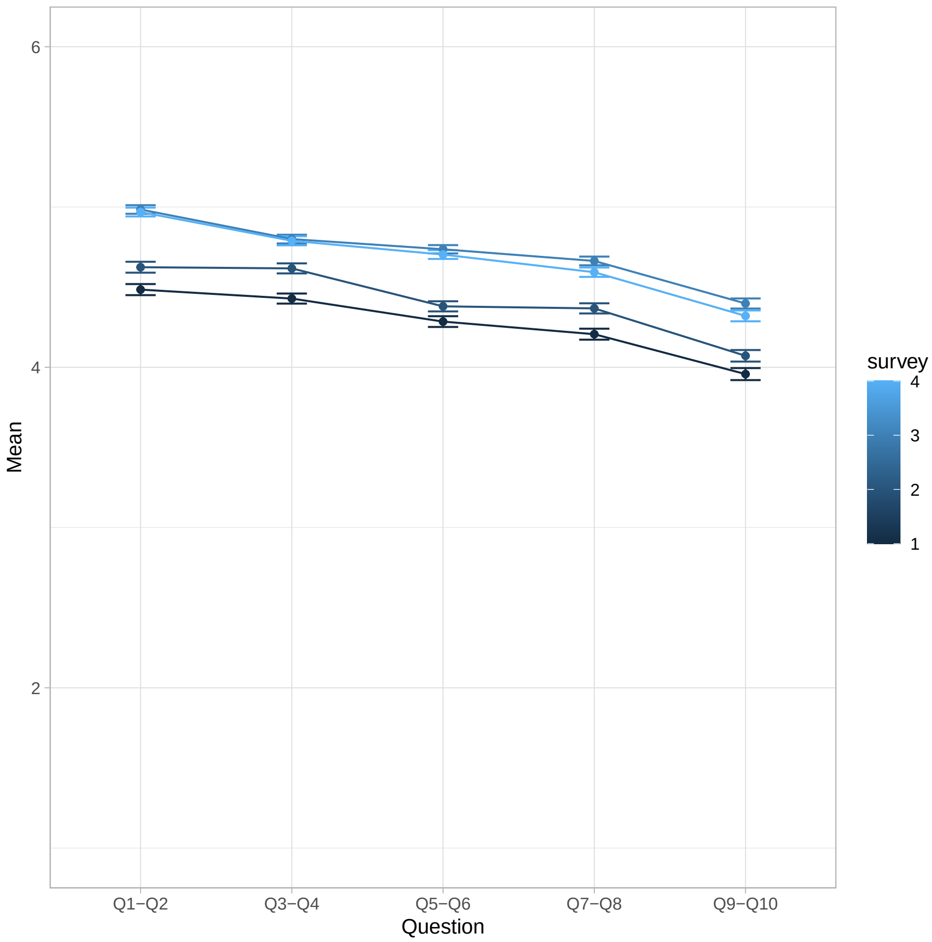


Figure 4.7: Combined Interval plot of the question groupings.

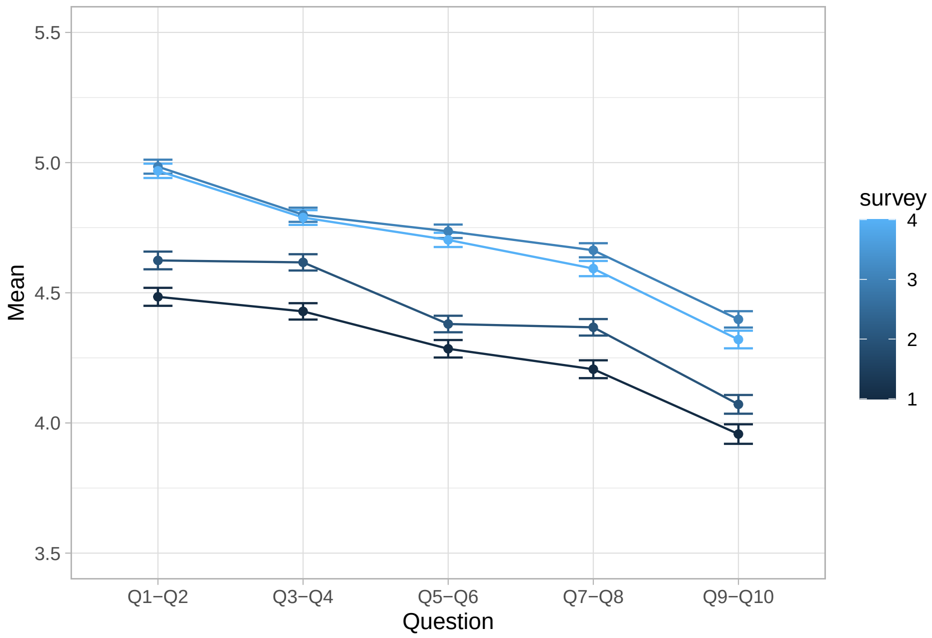


Figure 4.8: Zoomed in version combined Interval plot of the question groupings.

Figure 4.9 and the zoomed in version 4.10 can tell us a bit about the development of each question. Overall, Q_8, Q_9, and Q_{10} continuously have a lower maturity than the other questions. Q_6 and Q_7 are usually statistically significantly higher than Q_4 and Q_5. Q_5 has a low mean compared to the surrounding questions throughout all the surveys. Another remark is that the question pairs Q_7 and Q_8 have the highest difference between the pairs. From the first survey to the second, Q_3 has leaped and became the top-scoring question. From surveys two to three, Q_3 finds itself closer to its question partner Q_4.

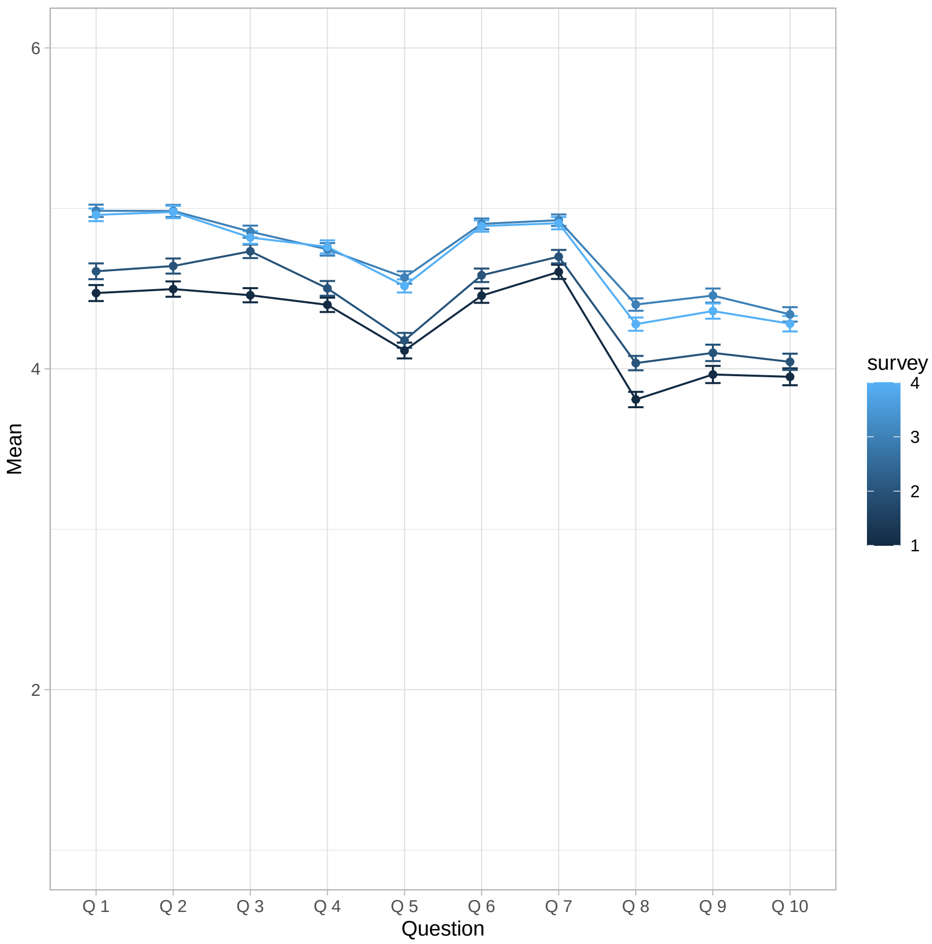


Figure 4.9: Combined Interval plot of the survey questions.

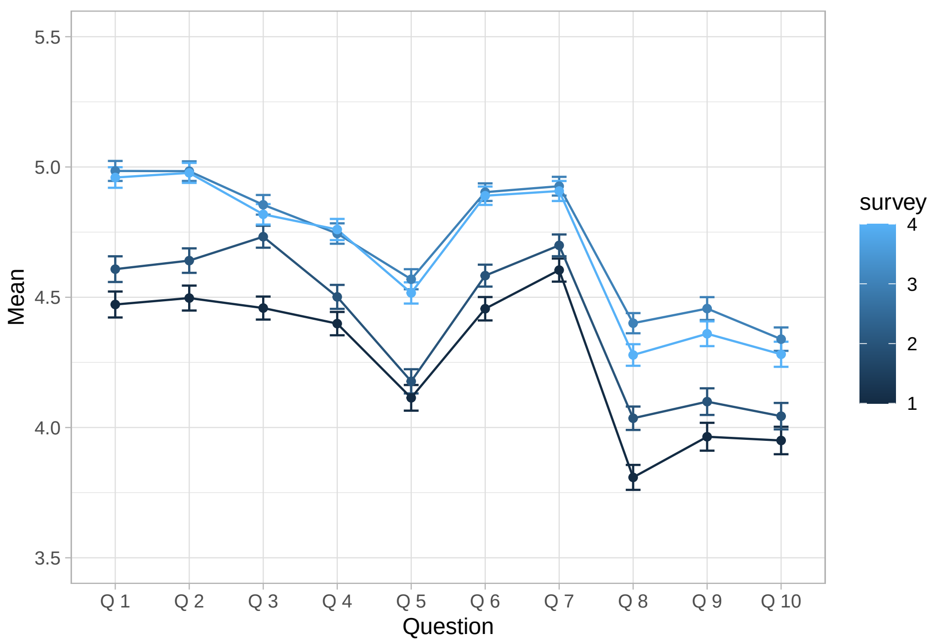


Figure 4.10: Zoomed in version of the combined Interval plot of the survey questions.

4.2.3 RQ3: What do the participants think of the 10-question agility survey?

This section presents the results of RQ3. The themes and sub-themes will be explained, and example sentences are included to clarify what types of questions were assigned to the sub-themes.

The number of total answers (displayed in Table 4.3), and how many answers were extracted from them, are shown in Table 4.18. It also shows the distribution of total and extracted answers between the four surveys. For the total answers, survey one has the least answers with 598 (15.8%), and survey three has the most answers with 1,423 (37.5%). For the extracted answers, survey one has the least answers with 48 (15.0%), and survey two has the most answers with 152 (47.6%).

Survey	**Total answers**	**% of total answers**	**Extracted answers**	**% of extracted answers**
#1	598	15.8%	48	15.0%
#2	759	20.0%	152	47.6%
#3	1,423	37.5%	67	21.0%
#4	1,014	26.7%	52	16.3%
#1-4	3,794	100%	319	100%

Table 4.18: The distribution of answers between the four surveys.

Table 4.19 shows the percentages of answers that were extracted from the total answers. The last row in the table is the combined results of all four surveys and provides a baseline to see how each survey differs from the average results. The strongest deviations are found in survey two (20.0% extracted answers) and survey three (4.7% extracted answers), compared to the average of 8.4% extracted answers.

Survey	**Total answers**	**Extracted answers**	**Extracted / total answers**
#1	598	48	8.0%
#2	759	152	20.0%
#3	1,423	67	4.7%
#4	1,014	52	5.1%
#1-4	3,794	319	8.4%

Table 4.19: The percentages of how many answers were extracted from the total.

Below are five thematic maps that show what the participants think of the 10-question agility survey. One thematic map is based on all the free-text answers, and the other four are based on the free-text answers from each survey. The thematic map based on all free-text answers (Figure 4.11), with a total of 319 answers, has 43 positive answers, 125 neutral/unclear answers, and 151 negative answers. The largest sub-theme is *wrong/missing questions*, with 47.3% of the extracted answers relating to it.

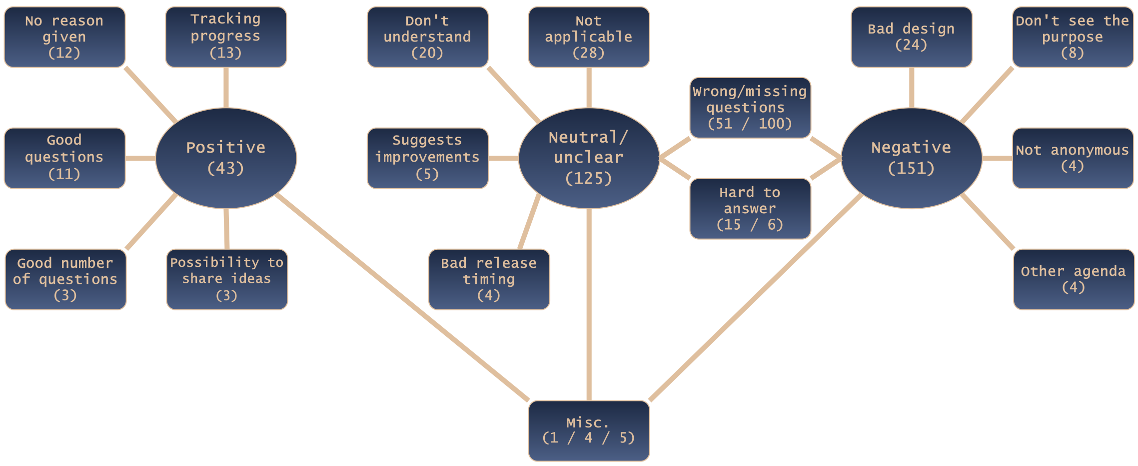


Figure 4.11: Thematic map based on all the surveys.

The thematic map based on the first survey's free-text answer (Figure 4.12), with a total of 48 answers, has three positive answers, 18 neutral/unclear answers, and 27 negative answers. The sub-themes that had no answers related to them were: good number of questions, possibility to share ideas, bad release timing, and not anonymous.

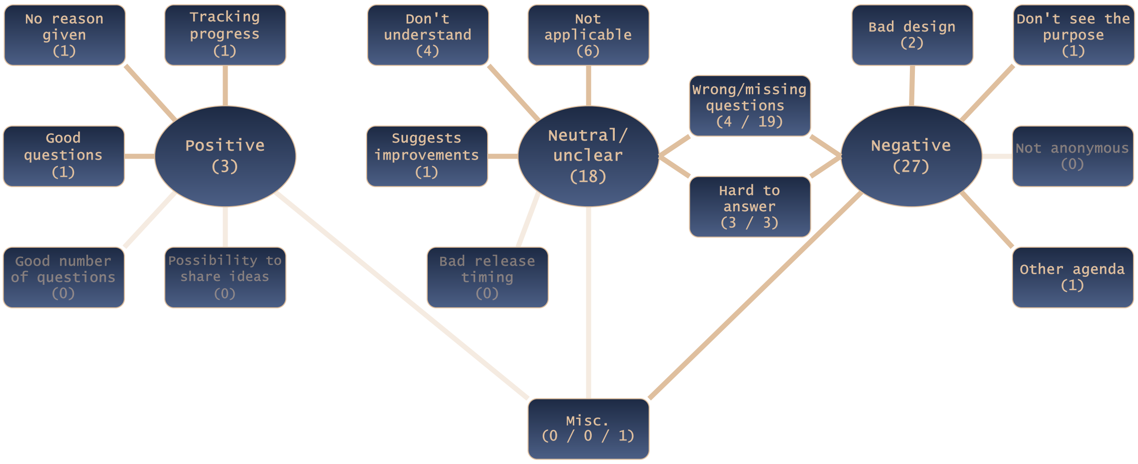


Figure 4.12: Thematic map based on the first survey.

The thematic map based on the second survey's free-text answer (Figure 4.13), with a total of 152 answers, has 28 positive answers, 60 neutral/unclear answers, and 64 negative answers. The only sub-theme with no answers related to it was: suggests improvements.

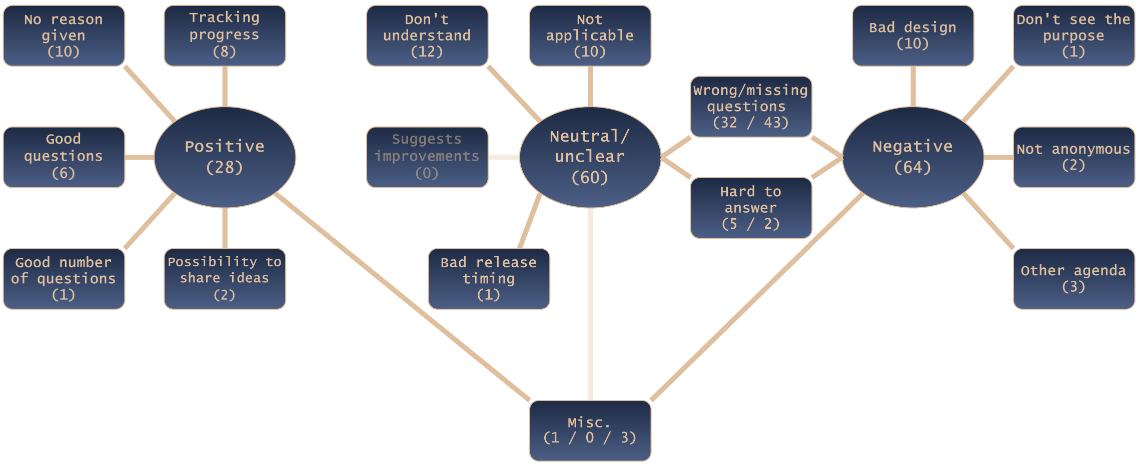


Figure 4.13: Thematic map based on the second survey.

The thematic map based on the third survey's free-text answer (Figure 4.14), with a total of 67 answers, has eight positive answers, 25 neutral/unclear answers, and 34 negative answers. The sub-themes that had no answers related to them were: no reason given, possibility to share ideas, bad release timing, misc., not anonymous, and other agenda.

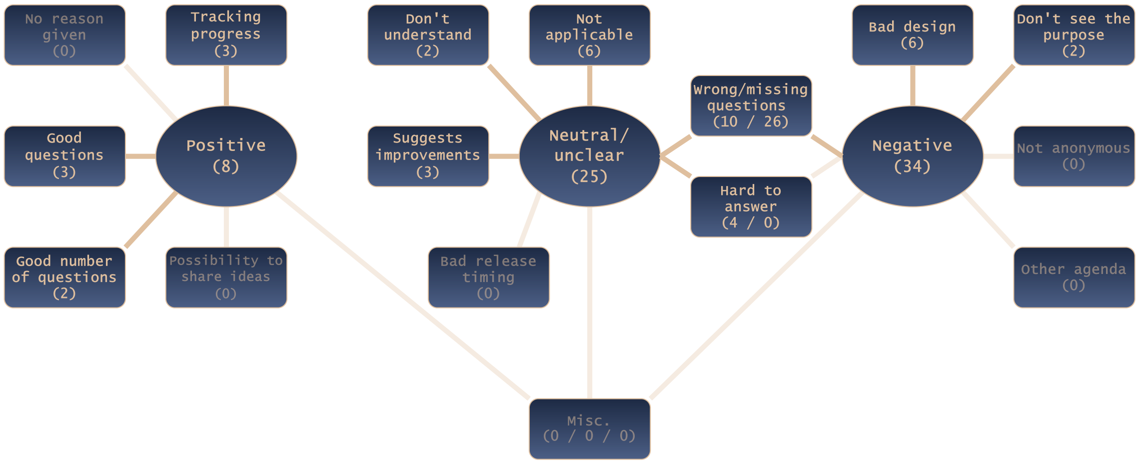


Figure 4.14: Thematic map based on the third survey.

The thematic map based on the fourth survey's free-text answer (Figure 4.15), with a total of 52 answers, has four positive answers, 22 neutral/unclear answers, and 26 negative answers. The sub-themes that had no answers related to them were: good number of questions and other agenda.

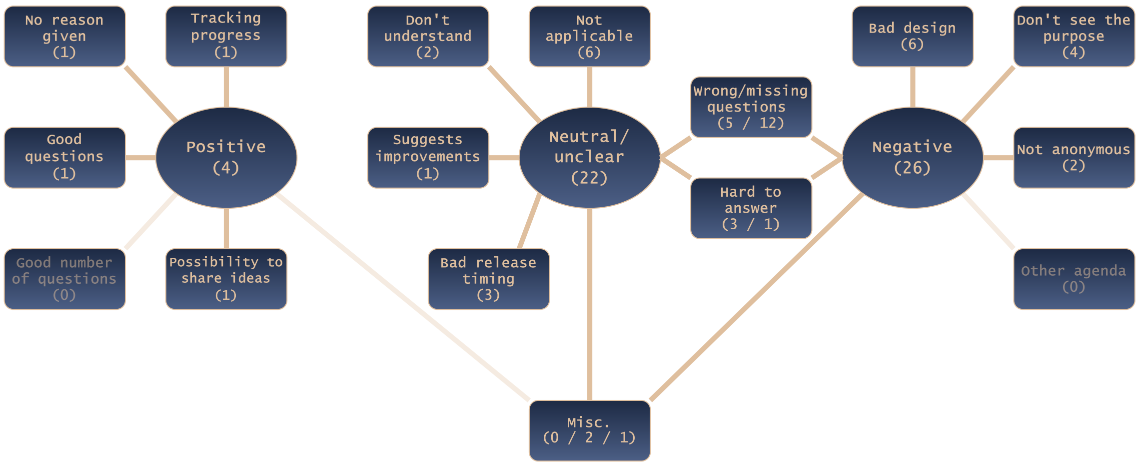


Figure 4.15: Thematic map based on the fourth survey.

A list with the details of all sub-themes is provided in Table 4.20. The table shows what theme the sub-themes belong to, how many extracted answers relate to the sub-themes, and an example sentence to understand what type of answers were included. The example sentences are made up as the free-text answers were under an NDA. No example sentences were provided to the Misc. sub-theme, as these free-text answers varied greatly and did not match any category.

Theme	Sub-theme	Extracted answers	Example sentence
Positive	Tracking progress	13	It's good to have a survey that tracks our agile progress.
Positive	No reason given	12	The survey is great.
Positive	Good questions	11	Good questions that are relevant.
Positive	Good number of questions	3	The survey is short and simple.
Positive	Possibility to share ideas	3	The survey is a good place to share our ideas.
Neutral/ unclear	Wrong/missing questions	51	I would like to see questions about X.
Neutral/ unclear	Not applicable	28	I'm part of team X, and the questions are not applicable to us.
Neutral/ unclear	Don't understand	20	I don't understand question X. What does the word Y mean?
Neutral/ unclear	Hard to answer	15	It's hard to answer when it's unclear what team I'm part of.
Neutral/ unclear	Suggests improvements	5	The survey could be better if you did X.
Neutral/ unclear	Bad release timing	4	Is this really the best time to release the survey?
Negative	Wrong/missing questions	100	The questions do not cover X. Why don't you ask us about Y?
Negative	Bad design	24	I don't see how this survey gathers any valuable information we can work with.
Negative	Don't see the purpose	8	I don't see the meaning of this survey.
Negative	Hard to answer	6	The questions are written so that I can't answer them properly.
Negative	Not anonymous	3	It's possible to identify me based on my answers.
Negative	Other agenda	3	The survey serves another purpose than measuring the agile transformation.
Positive	Misc.	1	-
Neutral/ unclear	Misc.	4	-
Negative	Misc.	5	-

Table 4.20: Details of all the sub-themes.

Below we have two tables that show the theme distribution among the total answers (Table 4.21) and the extracted answers (Table 4.22). As the number of total answers is much higher than the extracted answers, the percentages are considerably lower when compared to the total answers versus the extracted answers. Looking at the theme distribution in the extracted answers, the data shows that, among the four surveys, the first survey is the most negative, and survey two is the most positive. The first survey has the lowest frequency of positive answers (6.2%) and the highest frequency of negative answers (56.2%). On the contrary, the second survey has the highest frequency of positive answers (18.4%) and the lowest frequency of negative answers (42.1%). This can be compared to the average of 13.5% positive answers and 47.3% negative comments among the extracted answers.

Survey	**Total answers**	**Theme: Positive**	**Theme: Neutral/unclear**	**Theme: Negative**
#1	598	3 (0.5%)	18 (3.0%)	27 (4.5%)
#2	759	28 (3.7%)	60 (7.9%)	64 (8.4%)
#3	1,423	8 (0.6%)	25 (1.8%)	34 (2.4%)
#4	1,014	4 (0.4%)	22 (2.2%)	26 (2.6%)
#1-4	3,794	43 (1.2%)	125 (3.3%)	151 (4.0%)

Table 4.21: The theme distribution among the total answers.

Survey	**Extracted answers**	**Theme: Positive**	**Theme: Neutral/unclear**	**Theme: Negative**
#1	48	3 (6.2%)	18 (37.5%)	27 (56.2%)
#2	152	28 (18.4%)	60 (39.5%)	64 (42.1%)
#3	67	8 (11.9%)	25 (37.3%)	34 (50.7%)
#4	52	4 (7.7%)	22 (42.3%)	26 (50.0%)
#1-4	319	43 (13.5%)	125 (39.2%)	151 (47.3%)

Table 4.22: The theme distribution among the extracted answers.

5
Discussion

This chapter presents discussions about the research questions, what contributions this study provides practitioners of agile transformation, what significance the study has, and the identified threats to validity.

5.1 Research questions

In the following sections, the data analysis results will be discussed.

5.1.1 RQ1: Challenges with covering agile with a 10-question survey

In general, the mapping did not give much support for connecting the questions to the literature. As shown in the result section (Table 4.5), the questions did not map towards the agile values. No keywords were found in the values, and making connections by looking at the essence of the questions and values was also troublesome and gave no result. The most significant obstacle was that the values always put something over another, for example "Individuals and interactions **over** processes and tools." [4]. The survey questions answered if they were transparent, used a prioritized backlog, and worked according to their capacity, but never if they put it above something else. It could be argued since they include the questions "In my team, we regularly reflect and learn to improve.", "Leaders around my team support our improvement efforts." and "Leaders around my team avoid pushing work over our capacity." they focus on individuals. However, making that connection was a bit of a stretch and filled with assumptions.

When looking at the agile principles (AP), there were a few direct connections ($\{Q_4, AP_5\}$, $\{Q_3, AP_{12}\}$) and some that required a bit of interpretation ($\{Q_9, Q_{10}, AP_8\}$, $\{Q_4, AP_{12}\}$). The first distinct connection was Q_4 to "Build projects around motivated individuals. Give them the environment and support they need, and trust them to get the job done." and revolves around support. The question asks if "Leaders around my team supports our improvement efforts." Even though the question only covers a portion of the principle, a connection was prevalent. Q_3, "In my team, we regularly reflect and learn to improve." maps strongly toward principle 12, "At regular intervals, the team reflects on how to become more effective, then tunes and adjusts its behavior accordingly." Both are about regularly reflecting and using it to improve the team. One could interpret that the question is about improving the product since it does not specify what to reflect on and learn to improve on, but it was found highly unlikely. The mapping of Q_9 and Q_{10} towards principle 8 is not as straightforward. The two questions are about capacity, not committing to much, and the leaders should not push work on the teams. Connecting them with sustainable development and maintainable work pace does not explicitly match word for word but can be done by understanding the underlying intent. The last connection found was the one between Q_4 and principle 12. To "...reflects on how to become more effective, then tunes and adjusts its behavior accordingly." is interpreted as an improvement effort, and thus a connection could be found. This connection is vague because the question explicitly talks about leader support for the improvement efforts while the principal guides what the team should do. Thus, the essence of the two is different, and it is questionable if the connection should be made. Trying to reason why the questions mapped somewhat poorly toward the principles is more complex than the values. Some principles are relevant and concrete enough to be measured, but it might be that within the constraint of ten questions, they deemed other aspects more important.

Figure 4.7 of the agile definitions shows that Q_3 weakly connects to three agile definitions and that Q_4 weakly connects to one. Q_4 connects to the fifth agile definition as the word *empowering* in "Building software by empowering and trusting people" can be seen as a support, as empowering someone requires an action to support the person in acquiring what they need. Q_3 connects to the fifth and seventh agile definitions, as promoting constant feedback and using feedback to make constant adjustments can be seen as a way to *learn to improve*. Q_3 is also connected to the tenth agile principle, as continuous improvement can be argued to relate to the keyword *learn to improve*. However, as Q_3 talks about the team, it is only connected if the continuous improvement is about the person or team, not the software. The agile definitions vary considerably. Some are concise, others are broad, and some are vague, while others are more precise. Even if there is a substantial variety, three out of the four correlations were with the keyword *learn to improve* from Q_3. This result could indicate that most questions do not capture the core of agile, as one could otherwise expect at least some correlation to any of the available definitions. However, that does not necessarily mean that the agile definitions do an excellent job capturing the core of agile. Some of the agile definitions are vague to the extent that it becomes hard to understand what they mean in practice, and as a result, it becomes especially tricky to measure. For example, how do you measure the agile definition "Iterative and incremental (evolutionary) approach to software development which is performed in a highly collaborative manner by self-organizing teams with "just enough" ceremony that produces high-quality software in a cost-effective and timely manner which meets the changing needs of its stakeholders" [33]? It raises the question if one can only be truly agile if all of the items in this definition are satisfied. If all the items are fulfilled flawlessly, except that there is "too much" or "too little" ceremony, does it mean that one is not agile? Worth noting is that even if this definition covers many important agile aspects, none of the survey questions correlated.

Looking at Figure 4.8, there are only one strong and three weak correlations between the survey questions and the SAFe principles. Q_3 correlated strongly with the fourth SAFe principle, "Build incrementally with fast, integrated learning cycles." as *integrated learning cycles* can be clearly connected to the keyword *learn to improve*. One could also argue that *learning cycles* weakly connect to the keyword *regularly reflect*, as it is not unlikely that learning cycles include reflections, and cycles indicate that it is done regularly. Interestingly, the sixth SAFe principle, "Visualize and limit WIP, reduce batch sizes, and manage queue lengths," weakly correlates to three different questions. The number of correlations indicates that the sixth SAFe principle is more connected with the survey questions than the fourth principle, even if the connections are weak. Q_7 got a weak connection to the sixth principle as *visualizing* work can be argued to make it more *transparent*. Q_9 and Q_{10} have a weak connection to the sixth principle, as *limiting WIP* can be seen as a way to take the *team capacity* into consideration, even if it does not explicitly say so. Though, it is not a surprise that the survey questions do not correlate strongly to the SAFe principles. The survey questions measure the agile transformation, and SAFe implements Lean, Agile, and DevOps practices at scale. While some of the SAFe principles relate to agile, like the fourth principle, "Build incrementally with fast, integrated learning cycles.", some principles that do not relate as strongly, like the first principle, "Take an economic view." Another note is that the SAFe principles work as guidance in the implementation of SAFe in the entire organization, and the questions in the survey are targeted toward individuals [5]. The tenth principle, "Organize around value," makes more sense from an organization perspective. However, it might be hard for individuals to answer to what extent they feel things around them are organized around value.

When looking at the whole picture, Q_1, Q_2, Q_5, Q_6, and Q_8 did not map towards any items. It was not surprising that Q_1 and Q_2 would be hard to map since they aim to make sure the employees understand why the company is moving towards agile. Having a good reason for the transformation, which the employees understand, is not something that could be found in agile definitions or principles. It is, however, still essential for a successful agile transformation, and therefore Q_1 and Q_2 are reasonable to ask [38]. Q_5 and Q_6 did also not map towards anything, which is not too surprising either. They are big sweeping questions that aim to see if the employees feel that they know agile and how to move towards that way of working. An argument could be made that since they check the perceived general knowledge of the employee within agile, it covers important topics found in multiple agile sources, such as if the teams are working towards becoming autonomous and having regular deliveries. However, it could be argued that the questions are susceptible to the Donner Kruger effect "...in which poor performers in many social and intellectual domains seem largely unaware of just how deficient their expertise is. Their deficits leave them with a double burden—not only does their incomplete and misguided knowledge lead them to make mistakes but those exact same deficits also prevent them from recognizing when they are making mistakes and other people choosing more wisely" [68], where the persons do not understand enough about Agile to see their knowledge deficits within it. Since it is very processed-focused, Q_8 was also found hard to map onto the agile literature reviewed.

Looking instead at the questions that did map onto the literature, Q_3 had the best result with two strong and three weak. The question targets a vital part of agile: learning and adapting, and thus it only makes sense that it maps well. The other question that had a clear connection was Q_4 which mapped to agile principle five. However, the reason why it is in the survey is highly likely due to leadership support being crucial in an agile transformation [38], not its connection to that principle.

With fewer questions, the scope of those needs to become broader to cover all aspects of agile. The topic covered by Q_9 and Q_{10} are essential parts of Scrum, but they are perhaps too specific to fit in a survey with only ten questions if the aim is to cover all aspects of agile. The inclusion of these questions might be that these parts are of great importance for Volvo Cars, and they found it necessary to include these questions. If that is the case, those questions might be good for Volvo Cars, but they might not be as applicable for others.

Deciding on how many questions to include and what questions to ask is challenging. Sidky et al. [48] for example, have approximately 300 unique indicators for their model that is divided between different roles such as customers, management, and developers, while Stojanov et al. [50] had 62 indicators. Even though the number difference between them is substantial, the survey conducted by Volvo Cars is far from their numbers in measured indicators.

5.1.2 RQ2: Challenges with statistically validating so few questions

The first thing that creates problems in the factor analysis is how few measured items were used to measure the constructs. Even though the principal analysis supported four to five factors, the results had a significant problem. There was a strong dependency between the factors, indicating that the factors are not unique and statistically measure the same thing. Thus we cannot see statistical support for four to five factors.

The question pairs in the factor loadings tables mostly group up as they are meant to, but some questions, especially Q_9, do not significantly load onto any factors in some measurements. Looking at the uniqueness, how much of the variance is not explained by the factors, those questions do not have a significantly higher value. Therefore, they are not explained to a lesser extent by the extracted factors but rather by a combination of them. This insight also strengthens that it is not that there are too few factors but rather that there are too few measured variables to support four to five factors. When instead looking at the scree plot, the support is at three factors and is considered a more reasonable number based on the number of measured variables and the recommendation to have 3-5 times the number of factors from Fabrigar et al. [63].

For the second part, when the means and confidence interval were analyzed, the first plots, Figure 4.2 - 4.6, mostly followed what was expected. The most intriguing deviation was a tendency for the answers to decrease in survey four. The reason is unclear, although, in the free-text answers, it looked like there was a change in the organization to what was called R&D 2.0 sometime before the survey was handed out, which could have had an impact. Around this period, working from home due to corona could have been more prominent, which could have affected the results.

The interval plots for when the questions were combined conforms to the model. Q_1-Q_2 scores the highest, then Q_3-Q_4, etc. as expected. However, the results are much more sporadic when examining the individual questions. Creating an abstraction with the staircase and grouping only ten questions is unnecessary and leads to information being hidden. For example, the pair Q_7-Q_8 looks good when grouped up, but Q_7 is far behind Q_8. The idea behind the model is to find where resources are needed for the organization to mature, and grouping them makes it harder to identify where those resources are needed. In this regard, the plot showing the mean and confidence interval for each question better indicates how the transformation is going in this case.

5.1.3 RQ3: Discussing the results from the free-text answers

The average distributions between the themes (positive, neutral/unclear, and negative) show that the participants were clearly more negative towards the study than they were positive. Looking at the percentages, there were 13.5% negative comments, 39.2% neutral/unclear comments, and 47.3% negative comments, which means that 3.5 times more people were negative than positive. Even with such a negative result, one should not forget that looking at all the free-text answers (including those who did not comment on the survey), only 4.0% had a negative comment about the survey, as seen in Table 4.21. It is possible, and not unlikely, that the participants who disliked the survey were keener on providing that feedback than those who were positive about it.

One interesting thing to note in Table 4.22 is that the first survey was the most negative, and the second survey was the most positive. The first survey had 6.2% positive and 56.2% negative answers, and the second survey had 18.4% positive and 42.1% negative answers. There are a few theories why the first survey had the most negative responses. Firstly, it is likely that people are more negative at the start of a transformation, as not everyone is willing to change their way of working. Secondly, the abbreviation *P&Q* in Q_2 was changed to *Volvo Cars Product Creation* after the first survey. Not all participants understood what P&Q meant, which caused some frustration. Lastly, the first survey had the least free-text answers. There is a pattern in Table 4.22 which indicates that the greater the number of free-text answers, the more positive the results were. A similar pattern is found in a paper from Melián-González et al., where they found that the valance becomes more balanced, and the negative effect is mitigated when the number of reviews increases [69]. Just like this pattern could explain why the first survey was the most negative, it could also explain why the second survey was the most positive.

The second survey had by far the most extracted answers of all the surveys, as it had 47.6% of all the extracted answers, compared to 15.0% in survey one, 21.0% in survey three, and 16.3% in survey four. This means that the second survey has the highest weight, and its results affected the average results more than any other survey. If all the surveys had the same weight and the same theme distribution as shown in Table 4.22, then the average percentage of positive comments would decrease to 11.1% instead of 13.5%, and the percentage of negative comments would increase to 49.8% instead of 47.3%.

The largest sub-theme is *wrong/missing questions* with a total of 47.3% of the extracted answers (Figure 4.11) if you combine the answers from both the neutral/unclear and positive themes. As this is by far the largest sub-theme, it is important to discuss what it entails and why almost half of the participants mentioned it. The example sentences found in Table 4.20 are “I would like to see questions about X.” for the neutral/unclear theme, and “The questions do not cover X. Why don’t you ask us about Y?” for the negative theme. One reason for the high number of comments in this sub-theme could stem from the fact that a ten-question survey is limited to how many aspects it can cover. Many free-text answers expressed that some aspects they found important to the transformation were not covered. Though, in some cases, there appears to have been a gap between what questions the participants expected and what they got. The purpose of the survey was to measure the agile maturity during the transformation. However, some of the free-text answers asked for additional questions that were important to the participant but were unrelated to the agile maturity.

5.2 Contributions & significance of the study

The purpose of the study was to evaluate to what degree one can use the short survey to measure a large automotive company’s agile transformation. This study adds to the pool of ways to measure agile that have undergone statistical validation. It gives insights into the strength and weaknesses of both the 10-question survey used to gather data and the maturity model used during the agile transformation Volvo Cars underwent. This information is helpful in understanding what options there are to evaluate an agile transformation. It also provides some guidance for creating surveys for measuring an organization’s shift to agile and what to think about when designing them.

5.3 Threats to validity

Taking inspiration from the checklist presented by Lucas Green [70] we present our identified threats to validity in the following section.

5.3.1 Construct validity

Analyzing free-text comments is a challenging and time-consuming task. There are many free-text comments in the dataset, and there is no guarantee that the found aspects are the only aspects of significance. Since the authors manually analyzed the answers, they were subject to interpretations that could differ from the original intentions. The researchers did the most crucial steps separately to prevent faulty interpretations, and later on, they compared and discussed the results.

The mapping between the survey questions and the agile sources is subject to the researchers' interpretation, which introduces the risk of incorrect mapping. The researchers did the mapping individually and later discussed their results to mitigate this risk.

Only agile definitions that could be identified and verified in two working days were included in the study. There may be more definitions of agile available, so there is a risk that the survey questions map strongly to definitions that are not included. However, the number of included definitions is substantial, and the risk of missing vital definitions is considered small.

There is a risk that the *duplicate deletion* of the free-text answers removed valid and identical responses. Still, the likelihood of this happening was considered extremely low. There is also a risk that comments about the survey were not successfully extracted. This risk was mitigated by having the researchers extract the data individually and afterward compare the results.

There could have been outer influence on the data which we do not know of since the data collection was performed by Volvo Cars. For example we expect that some change in the world or in the organisation affected the results during survey four.

5.3.2 External validity

This is a field study where the accessed data regarding the 10-question survey came from a single company. Because of this, the results are context-specific and there is a risk that the they do not apply in a more generalized context.

The data was gathered from a 10-question survey released four times every sixth month. Even if this study ends up concluding that the survey does not successfully measure agile transformation in the automotive industry, it does not necessarily mean that there are no other 10-question surveys that could measure this.

6
Conclusion

Even though the survey questions did not cover the agile sources brought up in this thesis that well, there is still value in the questions asked. In Q_1 and Q_2, making sure to have clear goals and a logical reason for making the shift that the employees understand is, as mentioned in Section 2.4 necessary for a successful agile transformation. Q_3 targets the need to learn, improve and adapt, which is a huge part of agile and can be found in the twelfth agile principle. Q_4 focuses on the important topic of having leadership commitment and support. This topic is another one of the crucial aspects of a successful transformation we wrote about in Section 2.4. For Q_5 and Q_6, while there is no direct support in the literature, making sure the employees understand and have enough knowledge of how to move forward with their agile ways of working makes sense. Contrary to what is measured in Q_7 and Q_8 agile promotes face-to-face conversation. Although, in a big automotive company such as Volvo Cars, that might be hard due to the sheer amount of teams that need to collaborate. Ensuring that the teams are transparent and the information is available for all teams can be a great compromise between face-to-face conversation and time since meetings take up a lot of time. Lastly, both Q_9 and Q_{10} connect to the eighth agile principle of sustainable development. Using such a short survey, one can see that you have to select what is important for the company to measure. This selection allows the company to show the employees what is crucial through what questions they include. For example, in this case, question ten *"Leaders around my team avoid pushing work over our capacity."* could indicate to the employees that it is of high importance since it made it to the list of ten questions.

The data followed the proposed staircase model as expected. However, grouping the questions could hide important information if each question is not analyzed. With so few questions, this might not be needed to interpret the result in a useful manner. Statistical support for the number of proposed factors could not be found since the number of measured variables was too few. Instead, there was support for three factors for the given amount of measured variables. To conclude the participants' view, 8.4% of the survey respondents expressed an opinion about the survey. Of those, 13.5% had positive remarks, 39.2% left a neutral comment, and 47.3% showed a negative attitude towards it. The most significant portion (47.3%) of the extracted free-text answers belonged to the sub-theme *wrong/missing questions*. Not all aspects the employees wished for could be included with only ten questions. There also seemed to have been a dissonance between the type of questions included in the survey and what the employee expected, even though its purpose was communicated when sent out.

6.1 Future work

The industry currently uses many different definitions of agile, and some of them are vague to the point that it becomes hard to measure the level of agility. The authors of this paper see the need for a systematic literature review on the agile definitions to try and pin down what definitions cover useful aspects of agile. As industries differ from one another, it could also be beneficial to identify the most useful agile definitions for a specific industry, such as the automotive industry. Once a few useful agile definitions have been identified, it would be possible to compare them against each other in-field to see which works best and why. By converging toward a few selected and proven definitions of agile, it will be possible to remove some of today's ambiguity.

As 47.3% of the free-text answers regarding the survey ended up in the sub-theme wrong/missing questions, this raises the question of how short a survey can be before it stops being an effective tool for measuring agile transformation. Future research is needed to identify what kind of questions are the most effective to measure an agile transformation when the number of questions is limited. Research is also needed to identify the optimal number of questions to cover the essential parts of an agile transformation while at the same time keeping the survey as short as possible.

Bibliography

[1] Y. Bassil, “A simulation model for the waterfall software development life cycle,” *arXiv preprint arXiv:1205.6904*, 2012.

[2] S. Balaji and M. S. Murugaiyan, “Waterfall vs. v-model vs. agile: A comparative study on sdlc,” *International Journal of Information Technology and Business Management*, vol. 2, no. 1, pp. 26–30, 2012.

[3] “SAFe 5.0 Framework.” [Online]. Available: https://www.scaledagileframework.com/

[4] “Manifesto for Agile Software Development.” [Online]. Available: https://agilemanifesto.org/

[5] “Business Agility | Agility Planning | Scaled Agile Framework.” [Online]. Available: https://scaledagile.com/

[6] J. Saad, S. Martinelli, L. S. Machado, C. R. B. de Souza, A. Alvaro, and L. Zaina, “UX work in software startups: A thematic analysis of the literature,” *Information and Software Technology*, vol. 140, p. 106688, Dec. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950584921001452

[7] S. Sharma, D. Sarkar, and D. Gupta, “Agile processes and methodologies: A conceptual study,” *International journal on computer science and Engineering*, vol. 4, no. 5, p. 892, 2012.

[8] P. Gregory, L. Barroca, H. Sharp, A. Deshpande, and K. Taylor, “The challenges that challenge: Engaging with agile practitioners’ concerns,” *Information and Software Technology*, vol. 77, pp. 92–104, Sep. 2016. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0950584916300623

[9] S. Chandra Misra, V. Kumar, and U. Kumar, “Identifying some critical changes required in adopting agile practices in traditional software development projects,” *International Journal of Quality & Reliability Management*, vol. 27, no. 4, pp. 451–474, Jan. 2010. [Online]. Available: https://doi.org/10.1108/02656711011035147

[10] L. Barroca, T. Dingsøyr, and M. Mikalsen, “Agile Transformation: A Summary and Research Agenda from the First International Workshop,” in *Agile Processes in Software Engineering and Extreme Programming – Workshops*, ser. Lecture Notes in Business Information Processing, R. Hoda, Ed. Cham: Springer International Publishing, 2019, pp. 3–9.

[11] R. M. Fontana, R. Albuquerque, R. Luz, A. C. Moises, A. Malucelli, and S. Reinehr, “Maturity Models for Agile Software Development: What Are They?” in *Systems, Software and Services Process Improvement*, X. Larrucea, I. Santamaria, R. V. O’Connor, and R. Messnarz, Eds. Cham: Springer

International Publishing, 2018, vol. 896, pp. 3–14. [Online]. Available: http://link.springer.com/10.1007/978-3-319-97925-0_1

[12] L. Gren, R. Torkar, and R. Feldt, "The prospects of a quantitative measurement of agility: A validation study on an agile maturity model," *Journal of Systems and Software*, vol. 107, pp. 38–49, Sep. 2015. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0164121215001028

[13] "Top 10 Tips for Measuring Agile Success | Agile Alliance," Mar. 2015. [Online]. Available: https://www.agilealliance.org/top-10-tips-for-measuring-agile-success/

[14] A. Erande and A. Verma, "Measuring Agility of Organizations - A Comprehensive Agility Measurement Tool (CAMT)," *International Journal of Applied Management and Technology*, vol. 6, no. 3, Jan. 2008. [Online]. Available: https://digitalcommons.odu.edu/engtech_fac_pubs/58

[15] M. Leppänen, "A Comparative Analysis of Agile Maturity Models," in *Information Systems Development*, R. Pooley, J. Coady, C. Schneider, H. Linger, C. Barry, and M. Lang, Eds. New York, NY: Springer, 2013, pp. 329–343.

[16] Ming Huo, J. Verner, Liming Zhu, and M. Babar, "Software quality and agile methods," in *Proceedings of the 28th Annual International Computer Software and Applications Conference, 2004. COMPSAC 2004.* Hong Kong: IEEE, 2004, pp. 520–525. [Online]. Available: http://ieeexplore.ieee.org/document/1342889/

[17] W. W. Royce, "Managing the development of large software systems: concepts and techniques," in *Proceedings of the 9th international conference on Software Engineering*, 1987, pp. 328–338.

[18] K. Petersen, C. Wohlin, and D. Baca, "The waterfall model in large-scale development," in *International Conference on Product-Focused Software Process Improvement*. Springer, 2009, pp. 386–400.

[19] P. Rook, "Controlling software projects," in *Software Management*, 7th ed., D. J. Reifer and B. Boehm, Eds. IEEE, 2007, ch. 10, pp. 327–336.

[20] S. Mathur and S. Malik, "Advancements in the v-model," *International Journal of Computer Applications*, vol. 1, no. 12, pp. 29–34, 2010.

[21] S. Denning, "Why And How Volvo Embraces Agile At Scale," Jan. 2020. [Online]. Available: https://www.forbes.com/sites/stevedenning/2020/01/26/how-volvo-embraces-agile-at-scale/

[22] "15th Annual State Of Agile Report | Digital.ai." [Online]. Available: https://digital.ai/resource-center/analyst-reports/state-of-agile-report

[23] B. Meyer, *Agile! the good, the hype and the ugly*, 1st ed. New York: Springer, 2014.

[24] L. Raut, R. Wakode, and P. Talmale, "Overview on kanban methodology and its implementation," *International Journal for Scientific Research & Development*, vol. 03, pp. 2518–2521, 07 2015.

[25] M. Laanti, J. Similä, and P. Abrahamsson, "Definitions of Agile Software Development and Agility," in *Systems, Software and Services Process Improvement*, ser. Communications in Computer and Information Science, F. McCaffery, R. V. O'Connor, and R. Messnarz, Eds. Berlin, Heidelberg: Springer, 2013, pp. 247–258.

[26] A. Cockburn, *Agile software development: the cooperative game.* Pearson Education, 2001.

[27] J. A. Highsmith and J. Highsmith, *Agile software development ecosystems.* Addison-Wesley Professional, 2002.

[28] D. J. Anderson, *Agile management for software engineering: Applying the theory of constraints for business results.* Prentice Hall Professional, 2003.

[29] C. Larman, *Agile and iterative development: a manager's guide.* Pearson Education India, 2004.

[30] P. Schuh, *Integrating agile development in the real world.* Charles River Media, Inc., 2004.

[31] K. Lyytinen and G. M. Rose, "Information system development agility as organizational learning," *European Journal of Information Systems*, vol. 15, no. 2, pp. 183–199, 2006.

[32] V. Subramaniam and A. Hunt, *Practices of an agile developer: Working in the real world.* Pragmatic Bookshelf, 2006.

[33] S. W. Ambler, "Disciplined Agile Software Development: Definition," 2007. [Online]. Available: http://www.agilemodeling.com/essays/agileSoftwareDevelopment.htm

[34] P. Kruchten, "Contextualizing agile software development," *Journal of software: Evolution and Process*, vol. 25, no. 4, pp. 351–361, 2013.

[35] S. Denning, "How to make the whole organization "Agile"," *Strategy & Leadership*, vol. 44, no. 4, pp. 10–17, Jan. 2016. [Online]. Available: https://doi.org/10.1108/SL-06-2016-0043

[36] L. Gren and P. Lenberg, "Agility is responsiveness to change: An essential definition," in *Proceedings of the Evaluation and Assessment in Software Engineering.* Trondheim Norway: ACM, Apr. 2020, pp. 348–353. [Online]. Available: https://dl.acm.org/doi/10.1145/3383219.3383265

[37] P. Abrahamsson, O. Salo, J. Ronkainen, and J. Warsta, "Agile Software Development Methods: Review and Analysis," *arXiv:1709.08439 [cs]*, Sep. 2002, arXiv: 1709.08439. [Online]. Available: http://arxiv.org/abs/1709.08439

[38] K. Dikert, M. Paasivaara, and C. Lassenius, "Challenges and success factors for large-scale agile transformations: A systematic literature review," *Journal of Systems and Software*, vol. 119, pp. 87–108, Sep. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0164121216300826

[39] T. Dingsøyr, T. E. Fægri, and J. Itkonen, "What is large in large-scale? a taxonomy of scale for agile software development," in *International Conference on Product-Focused Software Process Improvement.* Springer, 2014, pp. 273–276.

[40] F. Almeida and E. Espinheira, "Large-Scale Agile Frameworks: A Comparative Review," *Journal of Applied Sciences, Management and Engineering Technology*, vol. 2, no. 1, pp. 16–29, Mar. 2021. [Online]. Available: https://ejurnal.itats.ac.id/jasmet/article/view/1832

[41] "The 14th Annual State of Agile Report is Here." [Online]. Available: https://digital.ai/catalyst-blog/the-14th-annual-state-of-agile-report

[42] T. J. Gandomani, H. Zulzalil, A. A. A. Ghani, A. B. M. Sultan, and M. Z. Nafchi, "Obstacles in Moving to Agile Software Development Methods; at a

Glance," *Journal of Computer Science*, vol. 9, no. 5, pp. 620–625, May 2013. [Online]. Available: https://thescipub.com/abstract/jcssp.2013.620.625

[43] S. Nerur, R. Mahapatra, and G. Mangalaraj, "Challenges of migrating to agile methodologies," *Communications of the ACM*, vol. 48, no. 5, pp. 72–78, May 2005. [Online]. Available: https://doi.org/10.1145/1060710.1060712

[44] C. Tolfo, R. S. Wazlawick, M. G. G. Ferreira, and F. A. Forcellini, "Agile methods and organizational culture: reflections about cultural levels," *Journal of Software Maintenance and Evolution: Research and Practice*, vol. 23, no. 6, pp. 423–441, 2011. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/smr.483

[45] D. Naslund and R. Kale, "Is agile the latest management fad? A review of success factors of agile transformations," *International Journal of Quality and Service Sciences*, vol. 12, no. 4, pp. 489–504, Jan. 2020. [Online]. Available: https://doi.org/10.1108/IJQSS-12-2019-0142

[46] K. Conboy, S. Coyle, X. Wang, and M. Pikkarainen, "People over Process: Key Challenges in Agile Development," *IEEE Software*, vol. 28, no. 4, pp. 48–57, Jul. 2011. [Online]. Available: https://doi.org/10.1109/MS.2010.132

[47] M. Olszewska (née Pląska), J. Heidenberg, M. Weijola, K. Mikkonen, and I. Porres, "Quantitatively measuring a large-scale agile transformation," *Journal of Systems and Software*, vol. 117, pp. 258–273, Jul. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016412121600087X

[48] A. Sidky, J. Arthur, and S. Bohner, "A disciplined approach to adopting agile practices: the agile adoption framework," *Innovations in Systems and Software Engineering*, vol. 3, no. 3, pp. 203–216, Sep. 2007. [Online]. Available: http://link.springer.com/10.1007/s11334-007-0026-z

[49] R. Benefield, "Seven Dimensions of Agile Maturity in the Global Enterprise: A Case Study," in *2010 43rd Hawaii International Conference on System Sciences*, Jan. 2010, pp. 1–7, iSSN: 1530-1605.

[50] I. Stojanov, O. Turetken, and J. J. Trienekens, "A Maturity Model for Scaling Agile Development," in *2015 41st Euromicro Conference on Software Engineering and Advanced Applications*. Madeira, Portugal: IEEE, Aug. 2015, pp. 446–453. [Online]. Available: http://ieeexplore.ieee.org/document/7302487/

[51] E. Bell *et al.*, *Business research methods*. Oxford university press, 2022.

[52] K.-J. Stol and B. Fitzgerald, "The abc of software engineering research," *ACM Transactions on Software Engineering and Methodology (TOSEM)*, vol. 27, no. 3, pp. 1–51, 2018.

[53] C. Wohlin, "Guidelines for snowballing in systematic literature studies and a replication in software engineering," in *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering - EASE '14*. London, England, United Kingdom: ACM Press, 2014, pp. 1–10. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2601248.2601268

[54] L. R. Fabrigar and D. T. Wegener, *Exploratory factor analysis*, ser. Understanding statistics. Oxford ; New York: Oxford University Press, 2012.

[55] A. E. Hurley, T. A. Scandura, C. A. Schriesheim, M. T. Brannick, A. Seers, R. J. Vandenberg, and L. J. Williams, "Exploratory and

Confirmatory Factor Analysis: Guidelines, Issues, and Alternatives," *Journal of Organizational Behavior*, vol. 18, no. 6, pp. 667–683, 1997. [Online]. Available: https://www.jstor.org/stable/3100253

[56] T. A. Brown, *Confirmatory Factor Analysis for Applied Research, Second Edition.* Guilford Publications, Jan. 2015, google-Books-ID: tTL2BQAAQBAJ.

[57] H. F. Kaiser, "The Application of Electronic Computers to Factor Analysis," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 141–151, Apr. 1960. [Online]. Available: http://journals.sagepub.com/doi/10.1177/001316446002000116

[58] R. B. Cattell, "The Scree Test For The Number Of Factors," *Multivariate Behavioral Research*, vol. 1, no. 2, pp. 245–276, Apr. 1966. [Online]. Available: https://doi.org/10.1207/s15327906mbr0102_10

[59] J. L. Horn, "A rationale and test for the number of factors in factor analysis," *Psychometrika*, vol. 30, no. 2, pp. 179–185, Jun. 1965. [Online]. Available: https://doi.org/10.1007/BF02289447

[60] W. R. Zwick and W. F. Velicer, "Comparison of five rules for determining the number of components to retain." *Psychological Bulletin*, vol. 99, no. 3, pp. 432–442, 1986. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-2909.99.3.432

[61] M. W. Watkins, "Exploratory Factor Analysis: A Guide to Best Practice," *Journal of Black Psychology*, vol. 44, no. 3, pp. 219–246, Apr. 2018. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0095798418771807

[62] Ő. çokluk and D. Koçak, "Using Horn's Parallel Analysis Method in Exploratory Factor Analysis for Determining the Number of Factors," *Educational Sciences: Theory and Practice*, vol. 16, no. 2, pp. 537–551, Apr. 2016. [Online]. Available: https://eric.ed.gov/?id=EJ1101205

[63] L. R. Fabrigar, D. T. Wegener, R. C. MacCallum, and E. J. Strahan, "Evaluating the use of exploratory factor analysis in psychological research." *Psychological Methods*, vol. 4, no. 3, pp. 272–299, Sep. 1999. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/1082-989X.4.3.272

[64] R. Cudeck, "10 - Exploratory Factor Analysis," in *Handbook of Applied Multivariate Statistics and Mathematical Modeling*, H. E. A. Tinsley and S. D. Brown, Eds. San Diego: Academic Press, Jan. 2000, pp. 265–296. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780126913606500112

[65] P. J. Curran, S. G. West, and J. F. Finch, "The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis." *Psychological Methods*, vol. 1, no. 1, pp. 16–29, Mar. 1996. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/1082-989X.1.1.16

[66] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, Jan. 2006. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa

[67] M. C. Howard, "A Review of Exploratory Factor Analysis Decisions and Overview of Current Practices: What We Are Doing and How Can We Improve?" *International Journal of Human-Computer Interaction*,

vol. 32, no. 1, pp. 51–62, Jan. 2016. [Online]. Available: http://www.tandfonline.com/doi/full/10.1080/10447318.2015.1087664

[68] D. Dunning, “Chapter five - The Dunning–Kruger Effect: On Being Ignorant of One’s Own Ignorance,” in *Advances in Experimental Social Psychology*, J. M. Olson and M. P. Zanna, Eds. Academic Press, Jan. 2011, vol. 44, pp. 247–296. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780123855220000056

[69] S. Melián-González, J. Bulchand-Gidumal, and B. González López-Valcárcel, “Online customer reviews of hotels: As participation increases, better evaluation is obtained,” *Cornell Hospitality Quarterly*, vol. 54, no. 3, pp. 274–283, 2013.

[70] L. Gren, “Standards of validity and the validity of standards in behavioral software engineering research: the perspective of psychological test theory,” in *Proceedings of the 12th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, ser. ESEM ’18. New York, NY, USA: Association for Computing Machinery, Oct. 2018, pp. 1–4. [Online]. Available: https://doi.org/10.1145/3239235.3267437

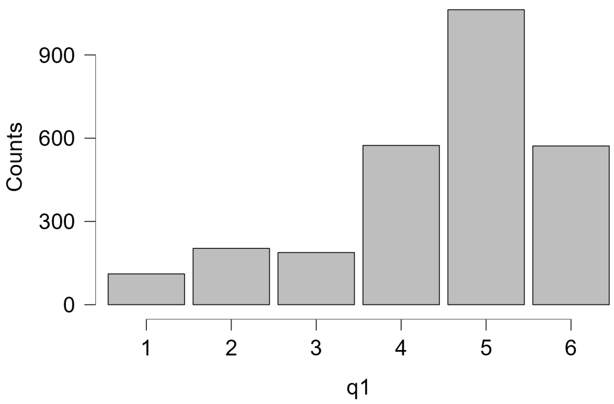
A

Normality plots

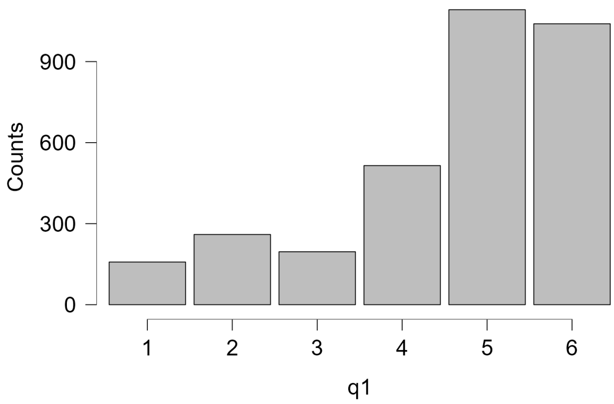
Distribution Plots

q1

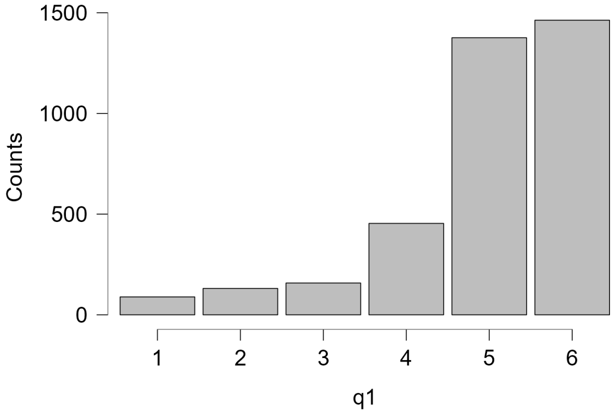
1



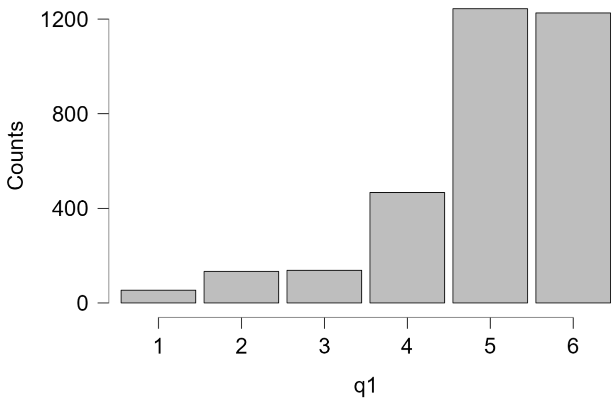
2



3

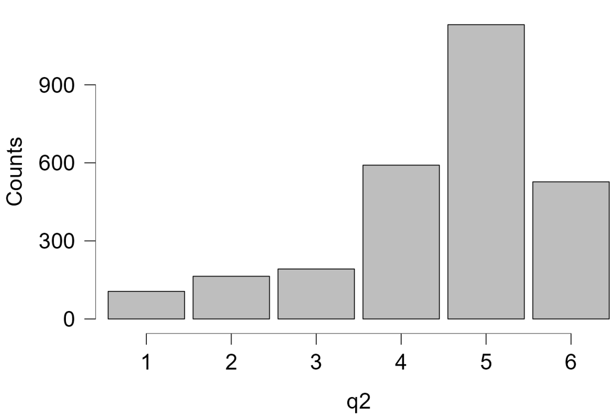


4



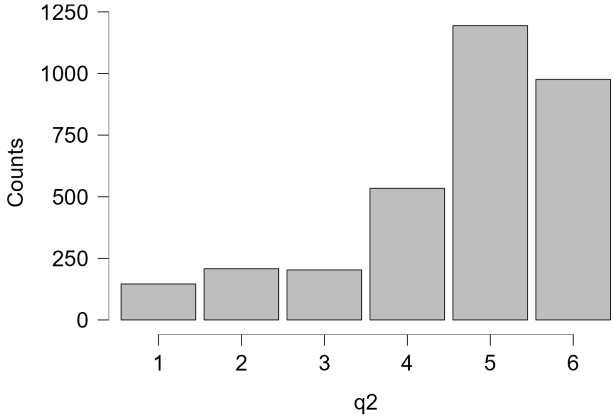
q2

1



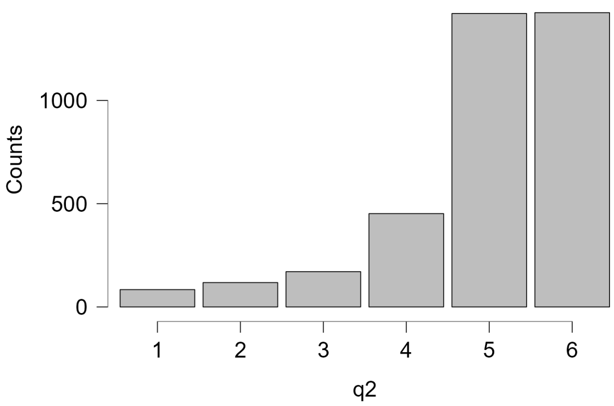
Counts
0
300
600
900
1
2
3
4
5
6
q2

2



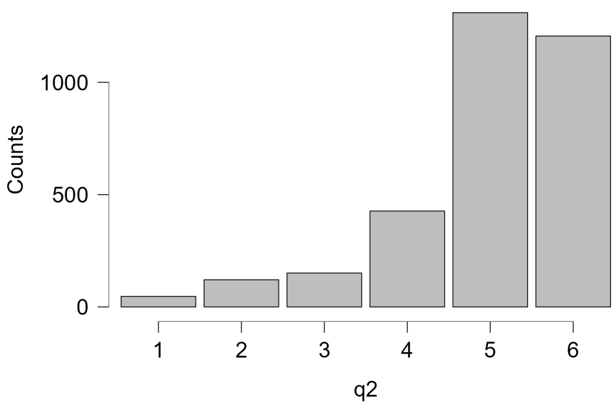
Counts
0
250
500
750
1000
1250
1
2
3
4
5
6
q2

3



Counts
0
500
1000
1
2
3
4
5
6
q2

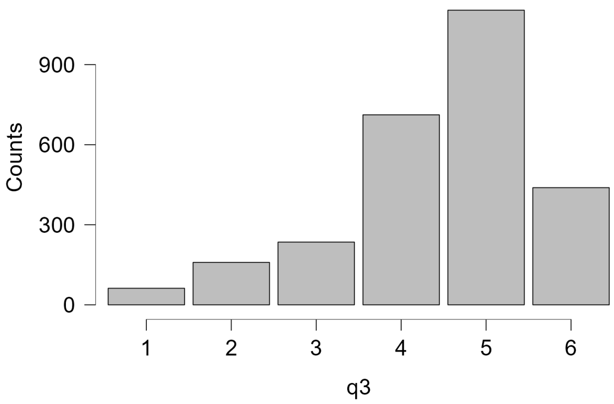
4



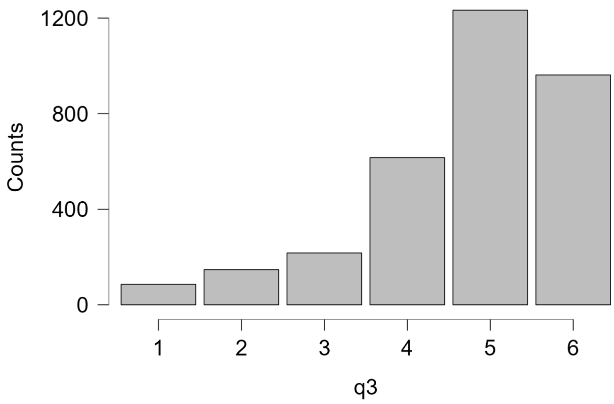
Counts
0
500
1000
1
2
3
4
5
6
q2

q3

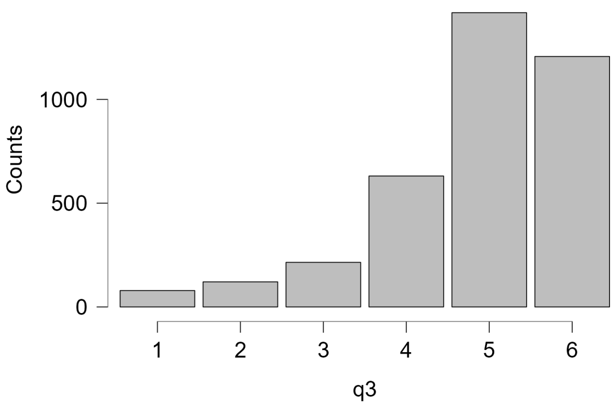
1


Counts
0
300
600
900
1
2
3
4
5
6
q3

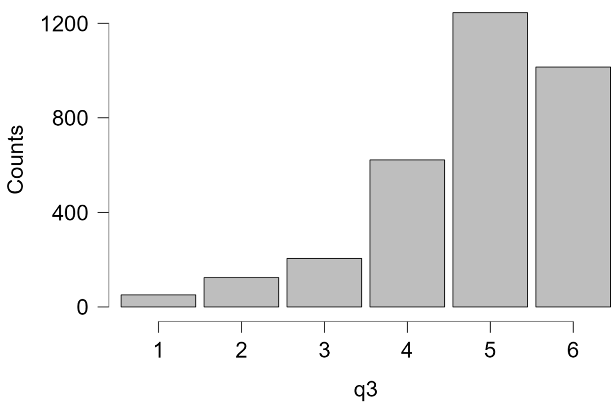
2


Counts
0
400
800
1200
1
2
3
4
5
6
q3

3

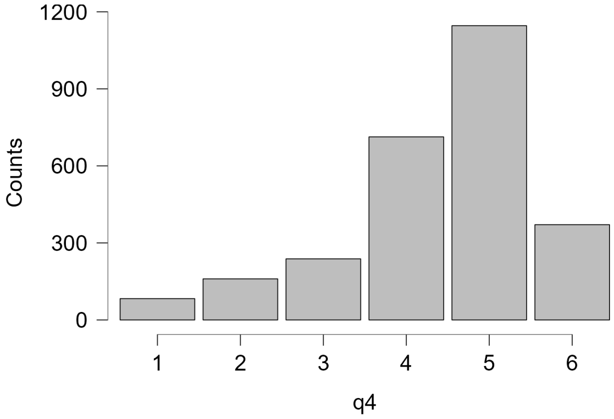

Counts
0
500
1000
1
2
3
4
5
6
q3

4

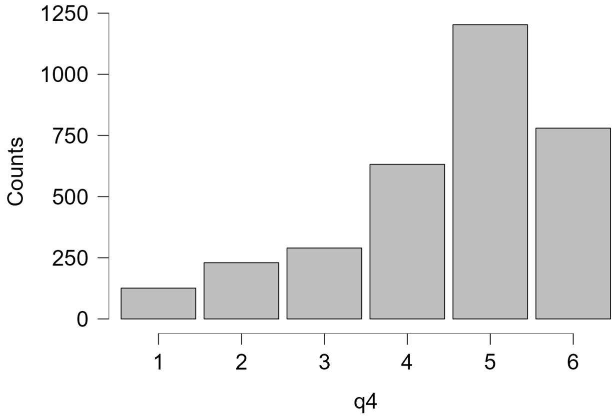

Counts
0
400
800
1200
1
2
3
4
5
6
q3

q4

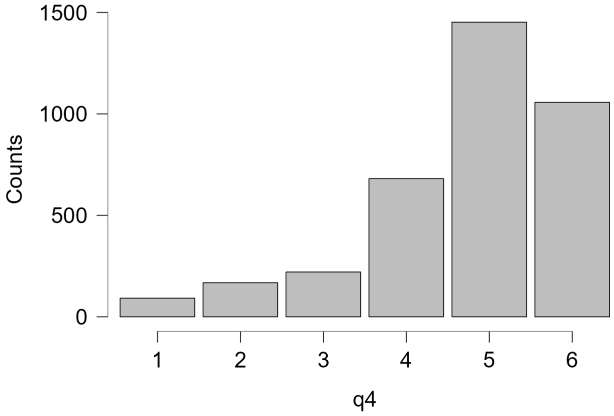
1


1200
900
600
300
0
Counts
1 2 3 4 5 6
q4

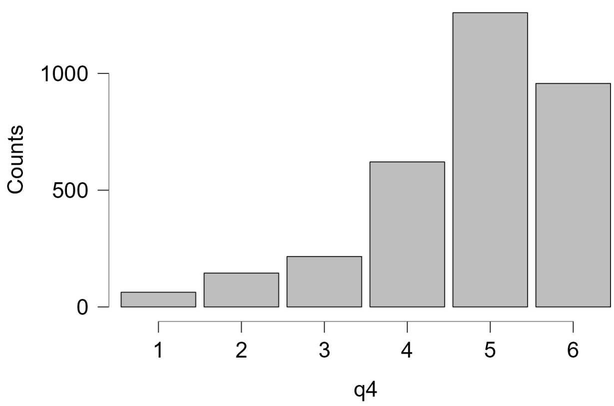
2


1250
1000
750
500
250
0
Counts
1 2 3 4 5 6
q4

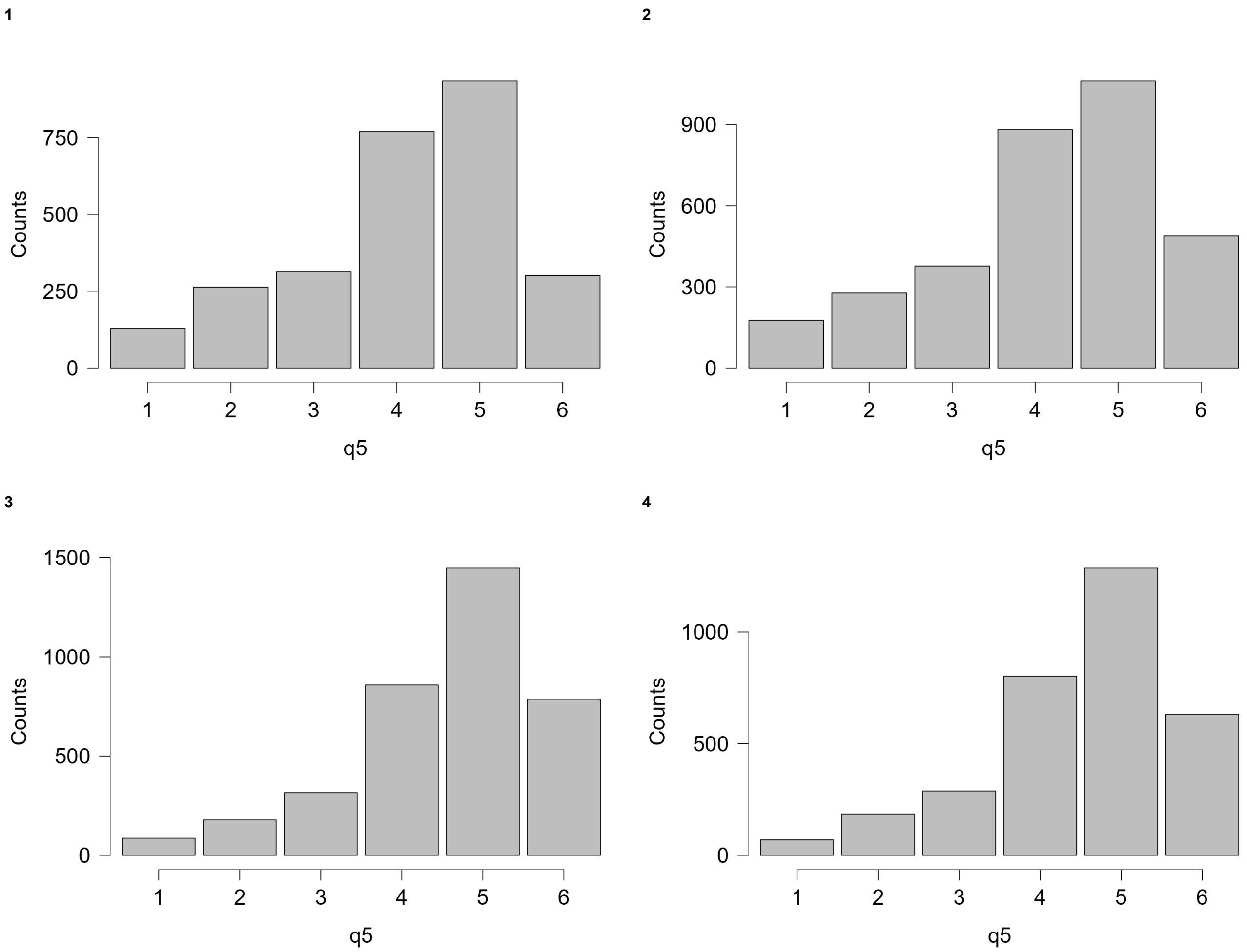
3


1500
1000
500
0
Counts
1 2 3 4 5 6
q4

4

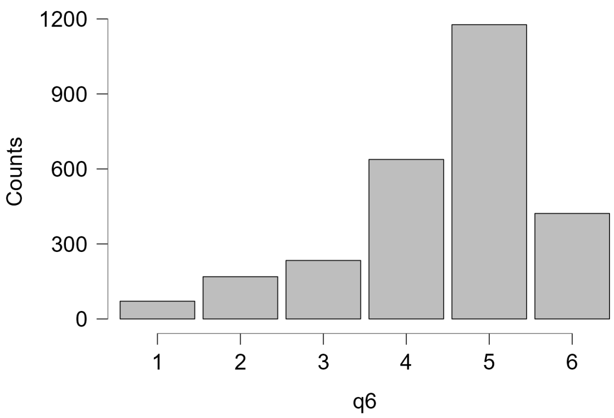

1000
500
0
Counts
1 2 3 4 5 6
q4

q5

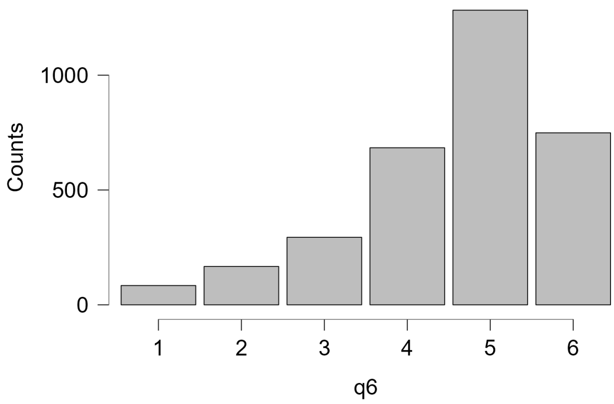


q6

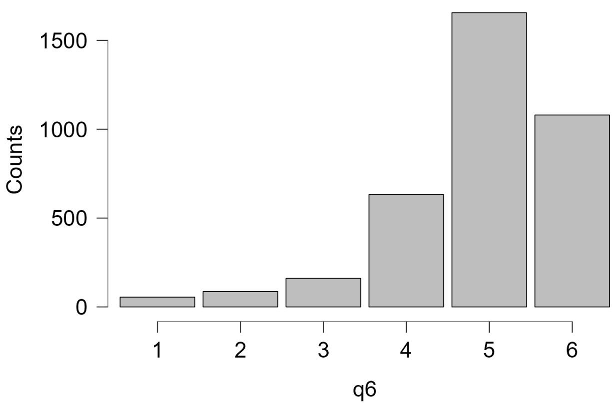
1


1200
900
600
300
0
Counts
1 2 3 4 5 6
q6

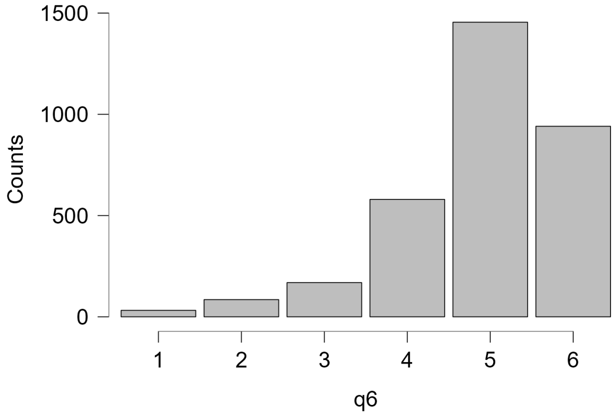
2


1000
500
0
Counts
1 2 3 4 5 6
q6

3

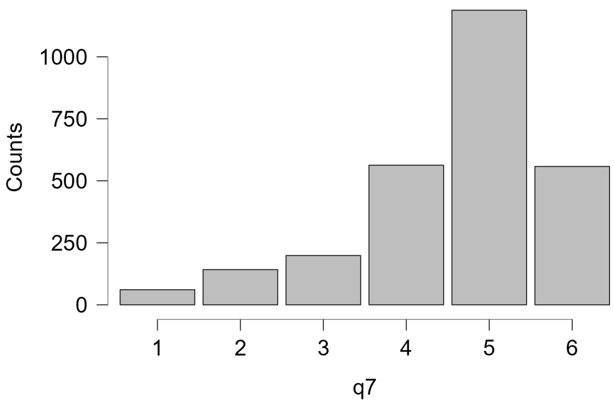

1500
1000
500
0
Counts
1 2 3 4 5 6
q6

4

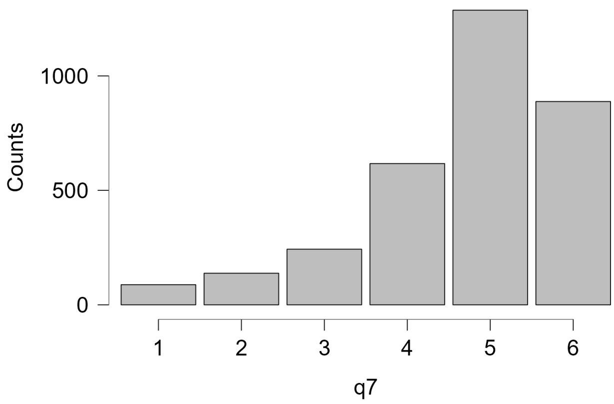

1500
1000
500
0
Counts
1 2 3 4 5 6
q6

q7

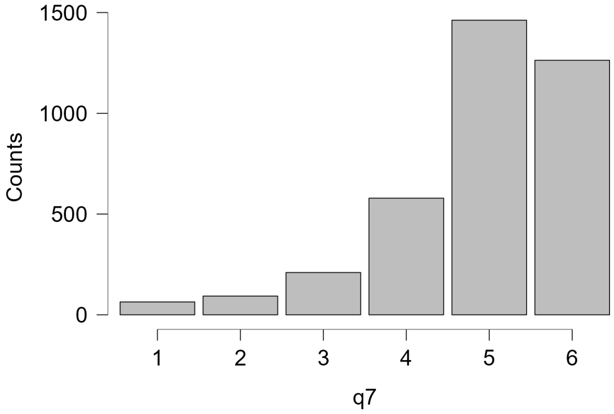
1


Counts
0
250
500
750
1000
1 2 3 4 5 6
q7

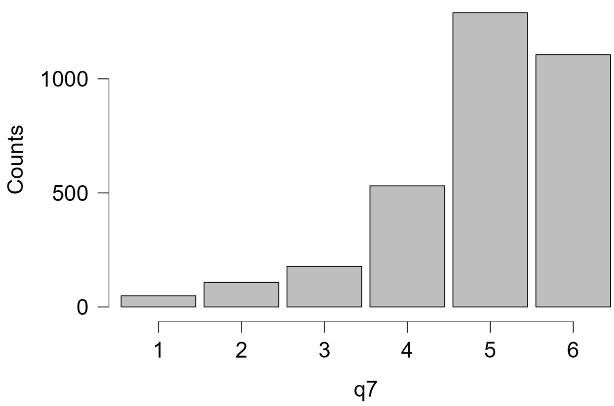
2


Counts
0
500
1000
1 2 3 4 5 6
q7

3

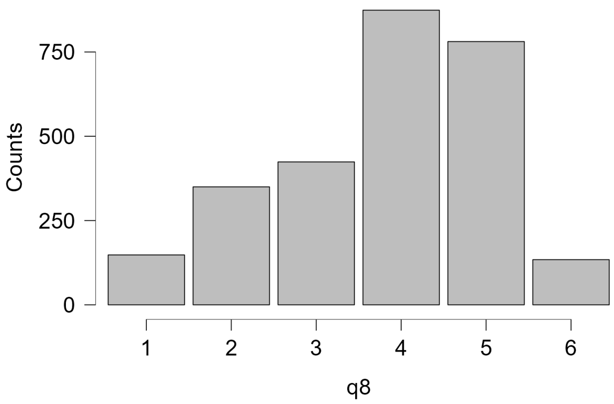

Counts
0
500
1000
1500
1 2 3 4 5 6
q7

4


Counts
0
500
1000
1 2 3 4 5 6
q7

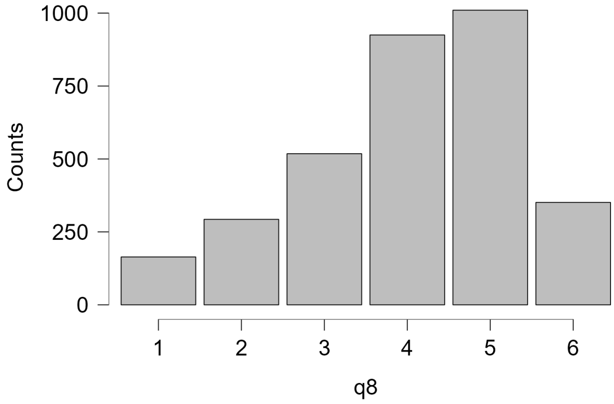
q8

1



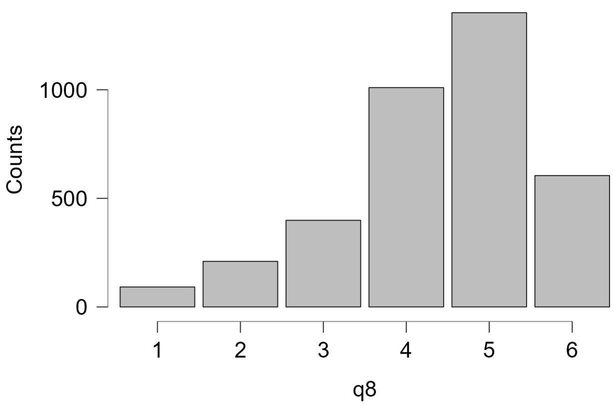
Counts
0
250
500
750
1
2
3
4
5
6
q8

2



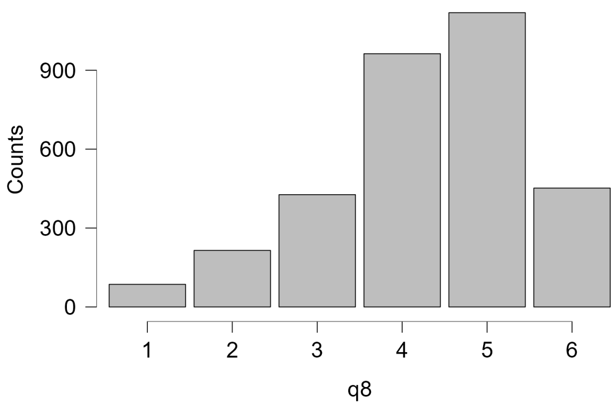
Counts
0
250
500
750
1000
1
2
3
4
5
6
q8

3



Counts
0
500
1000
1
2
3
4
5
6
q8

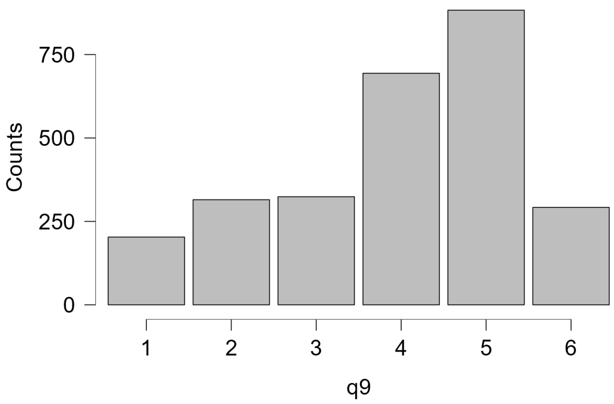
4



Counts
0
300
600
900
1
2
3
4
5
6
q8

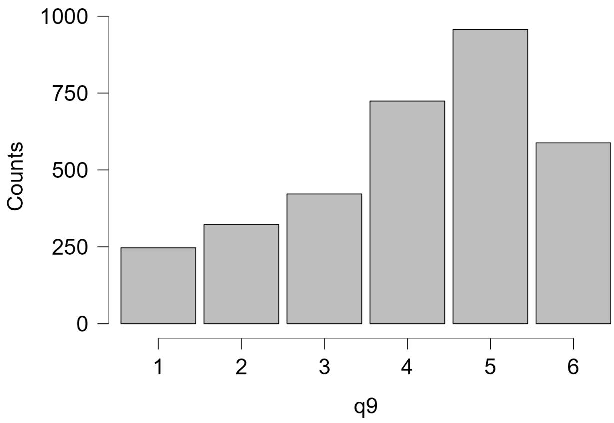
q9

1



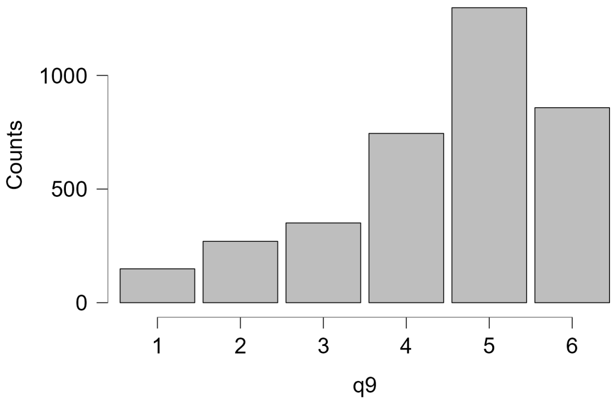
Counts
0
250
500
750
1
2
3
4
5
6
q9

2



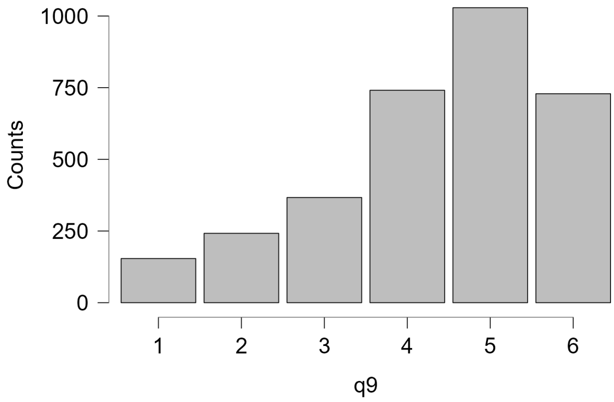
Counts
0
250
500
750
1000
1
2
3
4
5
6
q9

3



Counts
0
500
1000
1
2
3
4
5
6
q9

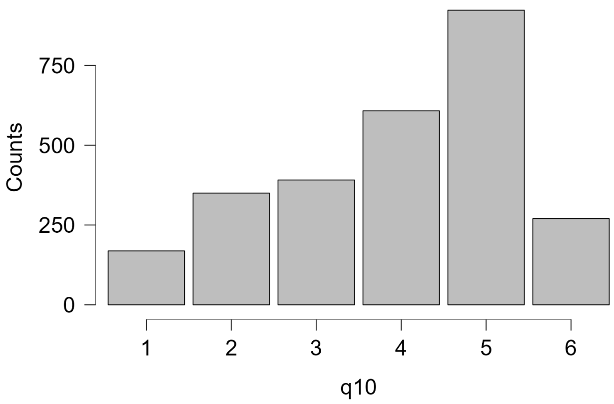
4



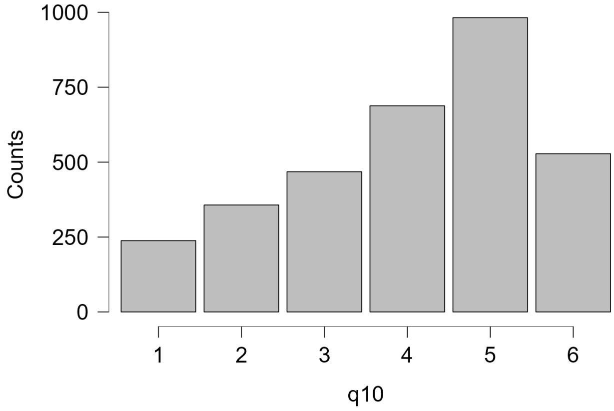
Counts
0
250
500
750
1000
1
2
3
4
5
6
q9

q10

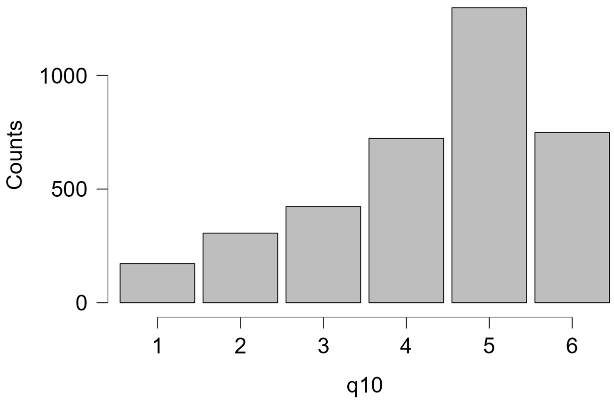
1


Counts
750
500
250
0
1
2
3
4
5
6
q10

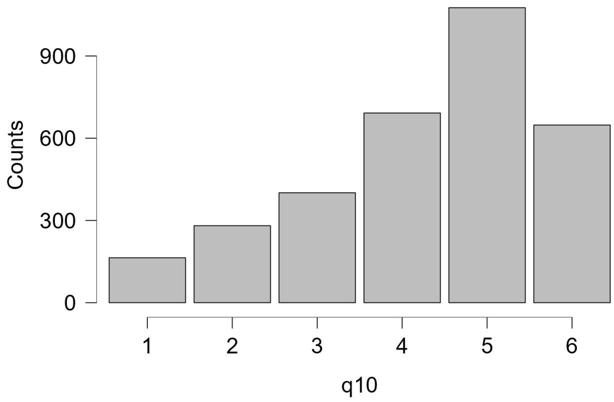
2


Counts
1000
750
500
250
0
1
2
3
4
5
6
q10

3


Counts
1000
500
0
1
2
3
4
5
6
q10

4


Counts
900
600
300
0
1
2
3
4
5
6
q10

B

Structure matrixes

	Factor 1	Factor 2	Factor 3	Factor 4
q1		0.929	0.587	
q2		0.851	0.506	
q3	0.751		0.489	0.483
q4	0.690		0.429	0.584
q5	0.608	0.552	0.781	0.462
q6	0.435	0.464	0.625	
q7	0.682		0.565	
q8	0.601		0.525	0.431
q9	0.599		0.577	0.584
q10	0.528			0.828

Table B.1: Structure matrix measurement one.

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
q1	0.910	0.439	0.581		
q2	0.901	0.418	0.528		
q3		0.726	0.614	0.463	0.635
q4		0.642	0.543	0.614	0.892
q5	0.611	0.743	0.982	0.554	0.498
q6	0.543	0.539	0.592		
q7		0.787	0.554	0.497	0.460
q8		0.654	0.529	0.565	0.442
q9	0.456	0.679	0.617	0.687	0.419
q10		0.494	0.408	0.777	0.457

Table B.2: Structure matrix measurement two.

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
q1	0.902		0.580	0.438	
q2	0.926		0.537	0.430	
q3		0.928	0.677	0.671	0.544
q4		0.739	0.620	0.666	0.644
q5	0.569	0.697	0.939	0.684	0.541
q6	0.551	0.445	0.610	0.536	
q7		0.672	0.642	0.738	0.524
q8		0.591	0.584	0.833	0.547
q9	0.408	0.622	0.655	0.646	0.640
q10		0.535	0.485	0.557	0.896

Table B.3: Structure matrix measurement three.

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
q1	0.883	0.403	0.564		
q2	0.888		0.490		
q3		0.698	0.675	0.701	0.495
q4		0.626	0.573	0.870	0.610
q5	0.556	0.697	0.908	0.567	0.549
q6	0.510	0.441	0.538		
q7		0.821	0.601	0.525	0.481
q8		0.650	0.533	0.490	0.533
q9	0.400	0.655	0.654	0.488	0.637
q10		0.520	0.473	0.512	0.832

Table B.4: Structure matrix measurement four.

C

Mean with 95% interval

Question	Mean	Lower	Upper	survey
Q1−Q2	4.484508	4.449979	4.519036	1
Q3−Q4	4.428624	4.397209	4.460039	1
Q5−Q6	4.284950	4.251342	4.318559	1
Q7−Q8	4.206381	4.172119	4.240644	1
Q9−Q10	3.957396	3.919875	3.994917	1
Q1−Q2	4.624195	4.590166	4.658225	2
Q3−Q4	4.616835	4.585531	4.648140	2
Q5−Q6	4.380098	4.348376	4.411820	2
Q7−Q8	4.367372	4.335666	4.399078	2
Q9−Q10	4.071450	4.035448	4.107453	2
Q1−Q2	4.984337	4.957479	5.011194	3
Q3−Q4	4.799646	4.772465	4.826827	3
Q5−Q6	4.736175	4.710296	4.762055	3
Q7−Q8	4.663307	4.636229	4.690385	3
Q9−Q10	4.397848	4.366448	4.429248	3
Q1−Q2	4.968424	4.940920	4.995929	4
Q3−Q4	4.789086	4.760811	4.817362	4
Q5−Q6	4.703096	4.675795	4.730397	4
Q7−Q8	4.593041	4.563816	4.622266	4
Q9−Q10	4.320509	4.286691	4.354327	4

Figure C.1: Mean, lower bound and upper bound of the 95% interval for the grouped questions.

Question	Mean	Lower	Upper	survey
Q 1	4.472150	4.422315	4.521986	1
Q 2	4.496865	4.449033	4.544697	1
Q 3	4.458502	4.414385	4.502620	1
Q 4	4.398746	4.354011	4.443481	1
Q 5	4.113980	4.064757	4.163204	1
Q 6	4.455920	4.411030	4.500810	1
Q 7	4.604205	4.559926	4.648484	1
Q 8	3.808558	3.760715	3.856400	1
Q 9	3.964589	3.911140	4.018037	1
Q 10	3.950203	3.897496	4.002910	1
Q 1	4.607789	4.558547	4.657031	2
Q 2	4.640601	4.593601	4.687602	2
Q 3	4.732291	4.690148	4.774433	2
Q 4	4.501380	4.455397	4.547363	2
Q 5	4.177246	4.130746	4.223747	2
Q 6	4.582950	4.540904	4.624996	2
Q 7	4.699172	4.657371	4.740973	2
Q 8	4.035572	3.990668	4.080476	2
Q 9	4.099356	4.048281	4.150431	2
Q 10	4.043545	3.992782	4.094307	2
Q 1	4.984745	4.946408	5.023082	3
Q 2	4.983928	4.946286	5.021570	3
Q 3	4.854808	4.817137	4.892479	3
Q 4	4.744484	4.705354	4.783613	3
Q 5	4.569055	4.530521	4.607589	3
Q 6	4.903296	4.869578	4.937014	3
Q 7	4.926178	4.890255	4.962102	3
Q 8	4.400436	4.361720	4.439151	3
Q 9	4.456551	4.412709	4.500394	3
Q 10	4.339145	4.294242	4.384048	3
Q 1	4.959534	4.920110	4.998958	4
Q 2	4.977315	4.938933	5.015696	4
Q 3	4.818210	4.778875	4.857544	4
Q 4	4.759963	4.719337	4.800590	4
Q 5	4.516554	4.475814	4.557294	4
Q 6	4.889638	4.854409	4.924868	4
Q 7	4.907725	4.869339	4.946111	4
Q 8	4.278357	4.236989	4.319725	4
Q 9	4.359902	4.312401	4.407403	4
Q 10	4.281116	4.232981	4.329251	4

Figure C.2: Mean, lower bound and upper bound of the 95% interval for the questions.