Investigating the Reasons and Expectations for Adopting Scaling Agile Frameworks

Shun Long Hong

Abstract Agile methodologies emphasize change tolerance, collaboration, and the involvement of customers during the software development process. Since these methodologies prove to be successful for small, co-located team environments, large organizations aim to scale the underlying methods onto an enterprise-level to harness the same benefits, such as shorter time to market, and increased responsiveness to change. To support this agile transformation, a growing number of companies rely on scaling agile frameworks, e.g., Scaled Agile Framework (SAFe) or Large-Scale Scrum (LeSS). Despite a growing attention from the industry and increasing amount of scientific research, literature investigating the reasons and motivation for companies to choose particular frameworks is still scarce. Also, there is no scientific literature focusing on the satisfaction with scaling agile frameworks. This paper aims to fill these gaps by providing a quantitative study analyzing the reasons to choose specific scaling agile frameworks and the fulfillment of expectations. We collected and statistically analyzed data from more than 20 countries, including USA, Germany, Denmark, Brazil, Japan, and New Zealand. Our sample indicates that 75.4% of survey participants are satisfied with their framework. Contrary to existing expectations in literature, we could not confirm that wide adoption is the primary reason for organizations to adopt scaling agile frameworks. We divided the sample data into different groups depending on organizational characteristics, e.g. size of development organization, distribution of teams and development sites. Using inferential analysis we compared different sizes of development organizations and found that there is no significant difference regarding the agreement towards documentation or support as relevant framework selection reasons. Also, respondents where more organizational areas were included within the scaling agile framework were significantly more satisfied than organizations that included less areas. SAFe and LeSS practitioners agreed significantly more with documentation and definition, and available support (e.g., training, coaching, and certifications) of frameworks being relevant for framework selection than organizations that used Spotify or internally created frameworks.

Shun Long Hong
Technical University of Munich
D-85748 Garching bei München, Germany
E-mail: {shunlong.hong}@tum.de

Keywords Agile Software Development · Enterprise Agile · Large-scale Agile Development · Quantitative Study · Scaling Agile Frameworks · Survey

1 Introduction

Today's business environment is changing faster than ever before. Organizations are faced with increasing pressure from competitors, customers, and technological developments [1], [2], [3]. To survive in this environment, companies must adapt themselves and therefore choose agile methodologies [2], [4]. Agile software development methods, such as Scrum and Extreme Programming (XP), are widespread and emphasize change tolerance, collaboration, and customer involvement during software development [5], [6], [7]. Small, co-located, self-organizing teams that work in close collaboration with their respective customers in single-project environments were able to capture benefits, such as maximized customer value, increased software quality, and responsiveness to change [6], [7], [8], [9]. Rapid iterations, failing fast, and frequent feedback loops are key components for increased software quality [6]. In recent years, many large organizations started to scale agile methods onto an enterprise-level to harness the same benefits as smaller teams [10], including Nokia [11], Ericsson [12], [13], and Amazon [14]. The adoption is not without any obstacles and characterized by issues concerning communication, inter-team coordination, dependencies among existing environments, or general change resistance [15], [16], [17]. To overcome these challenges many organizations use scaling agile frameworks such as the and Disciplined Agile Delivery (DAD) [18], Large-Scale Scrum (LeSS) [19], and Scaled Agile Framework (SAFe) [20], Spotify [5], [10], [21].

Despite the fact that there is a rising interest in adopting scaling agile frameworks, scientific literature providing empirical evidence on the reasons and satisfaction with adopting scaling agile frameworks is still scarce [7], [10], [16], [22], [23], [24]. Also, there is a lack of literature comparing the adoption of different scaling agile frameworks.

We aim to fill this gap by presenting the results of an empirical study investigating the underlying reasons and satisfaction of scaling agile framework adoptions. First, we aim to shed light into the process of selecting scaling agile frameworks by understanding most significant and relevant reasons. Second, we strive to understand the satisfaction of practitioners. Third, we investigate and compare different scaling agile framework adoptions concerning their selection criterion and fulfillment of expectations. For this purpose we formulated three research questions:

- RQ1: How do organizational characteristics influence the selection process of scaling agile framework?
- RQ2: How do organizational characteristics affect the satisfaction with selected scaling agile frameworks?
- RQ3: How do different frameworks such as LeSS, internally created ones, SAFe, and Spotify affect selection process and satisfaction regarding scaling agile frameworks?

We define organizational characteristics in RQ1 and RQ2 as number of development teams, previous development model, experience with scaling agile frameworks (length of time a respective scaling agile framework was used), geographic distribution and site distribution. In addition to the organizational characteristics,

the chosen framework is considered in RQ2 and RQ3, since the expectations are directly related to the individual choice.

As elaborated before, agile development methods hold benefits that organizations strive to realize due to the growing speed of change in business. Paasivaara et al. [25] summarized the motivation factors to shift from plan-driven to agile methodologies, e.g., plan-driven causes bureaucratic overhead, excess documentation, and slow processes. The novelty of scaling agile frameworks is a problem for inexperienced practitioners and the lack of expertise requires good documentation [26]. One way to deal with this is by acquiring support from external sources and third parties, e.g., in the form of consulting projects as emphasized by Kettunen et al. [27]. As a result, we formulate two hypotheses:

– Plan-driven organizations have a stronger preference for good documentation (H1.1) and available support (H1.2) as a relevant adoption reason than agile organizations.

Large organizations with distributed teams and development sites must solve the challenge of scaling agile development [25]. At the same time, good communication becomes more relevant, due to increased number of business units, employees, and more complex structures [28]. Large organizations have more dependencies which raises the overall transformation complexity and the need for formal documentation when conducting any kind of organizational change [29]. Dikert et al. [15], Paasivaara [30], and Hossain et al. [31] found that there is a lack of guidance in literature concerning how to successfully scale agile to large projects and how to conduct agile transformations in large organizations. Conboy and Carroll [16] and Paasivaara [30] stress the importance of training for successful scaling agile adoption. We believe that increased adoption complexity leads large organizations to hire external consultants to support the framework adoption. Dikert et al. [15] and Agile One[1] [32] suggest that one of the key practices for successful scaling of agile are consultants and trainers, since some of the concepts might be misunderstood. The data provided by Kettunen et al. [27] show a similar picture. Hence, we formulate following hypotheses:

– Documentation (H1.3), support (H1.4), and scaling (H1.5) is more relevant for large development organizations than smaller ones.

Scaling agile requires long-term commitment and can take up to several years [25], [28], [30], [33]. Increased complexity, communication, and training effort are some of the factors responsible for this delay. Many organization start their agile journey with single teams, e.g., by launching pilots [15], [23]. During early stages, the amount of captured benefits is limited and cannot be fully realized until the scaling agile adoption is completed. We argue that there is a relationship between the framework experience (i.e., duration of scaling agile framework experience) and realized benefits, which act as a catalyst for respective satisfaction.

– H2.1: Organizations that used have more experience with scaling agile frameworks agree more with the statement *"The framework met the expectations of my organization"* than organizations with less experience.

[1] We are aware that the referenced survey is of non-scientific nature. Nevertheless, we decided to include this publication as it poses the largest survey on agile adoption.

Typically, agile software development methodologies are first picked up by the IT / product development department within an organization. Every agile team exists within a broader corporate environment where collaboration across business functions is inevitable. With increasing delivery speed and shorter release cycles, other units will be inevitably impacted [34]. Hence, agile teams need to communicate within, between teams, and within a broader organization [5]. Including additional areas into the scaling agile framework creates alignment and reduces frictions that can arise due to unaligned processes, practices, or lack of understanding. Subsequently, we formulate following hypothesis:

– H2.2: Organizations that included more corporate areas into scaling agile framework adoption are more satisfied with the respective adoption than organizations which included less areas in the adoption scope.

Similar for RQ1, we believe that geographic distribution of teams and having multiple development sites impacts the overall adoption satisfaction. Coordination across physical boundaries, cultural and language barriers, and time zones is cumbersome and increases required efforts. Hence we formulate following hypothesis:

– Organizations without distributed teams (H2.3) or sites (H2.4) are significantly more satisfied with their framework adoption than such with distributed teams or sites.

The literature study by Uludağ et al. [35] showed that LeSS and SAFe are extensively documented compared to other existing scaling agile frameworks (e.g., Spotify model, or internally created frameworks). Both frameworks have more than 50 official documents (contributions or cases), while Spotify model is covered by 12. LeSS and SAFe are supported by multiple companies across different countries which offer training and coaching services [10]. The Spotify model lack training courses and certifications [35]. Paasivaara [30] describes that the delivered SAFe trainings were perceived as *highly useful* by the investigated case company. Furthermore, respective curating organizations offer dedicated certification courses, e.g. LeSS offers training courses to become a *LeSS Trainer* or *LeSS-Friendly Scrum Trainer* [36]. Since internally created frameworks are highly customized to a specific organizations, the respective documentations and trainings are only available to within the company. We articulate the following hypotheses:

– Organizations that use SAFe agree significantly more with documentation (H3.1) and support (H3.2) than organizations that use Spotify or Internal frameworks.
– Organizations that use LeSS agree significantly more with documentation (H3.3) and support (H3.4) than organizations that use Spotify or Internal frameworks.

The study reported in this paper is based on a survey we conducted on a worldwide scale. We collected participant responses on expectations, adoption reasons, expected and realized benefits, and challenges. This paper is part of an international research project and additional publications will follow. We divided the data and will present findings from an analysis of a subset. We aim to identify differences in survey responses using the non-parametric Mann-Whitney U test.

Scaling more people is the one adoption reason which received the highest agreement. On the contrary, popularity as a relevant framework selection criteria received the highest disagreement.

The following paper is structured as follows: Section 2 describes the background of this paper and elaborates related research efforts. Section 3 elaborates applied research method and design. Among others, it provides an overview of the survey design process, the data preparation, and sampling process. Section 4 provides an overview of the sample and the data analysis results. Section 5 discusses key findings and our interpretation. Furthermore, it points out limitations of this research limitations. Finally section 6 concludes the paper and describes potential future work.

2 Background and Related Work

Based on our sample data, internally created frameworks, LeSS, SAFe and the Spotify model due to their number of occurrences as primary frameworks.

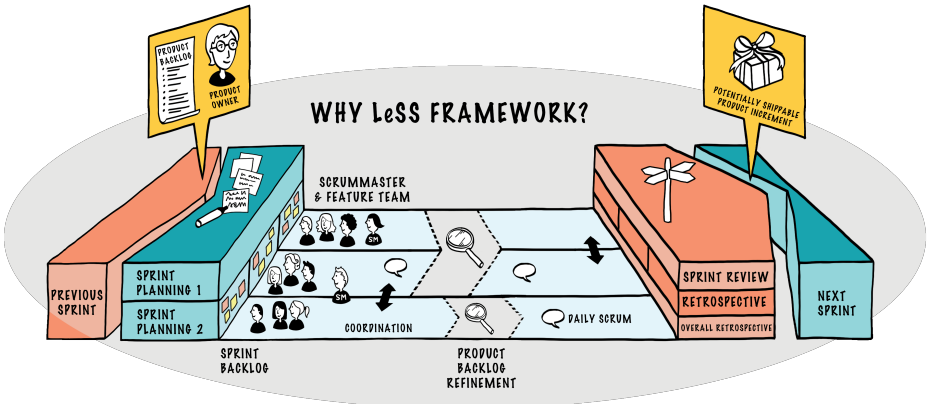
2.1 Large-Scale Scrum

Large-Scale Scrum (see Figure 1) was first formally published in 2009 by Craig Larman and Bas Vodde [37]. The main idea is to take Scrum and minimally adapt its practices to accommodate larger projects without loosing sight of the original goals [9]. Coordination is facilitated between multiple teams by having aligned sprint cycles and one Product Owner (PO) responsible for a central backlog. This leads to potentially shippable products after each sprint [38]. Furthermore, LeSS emphasizes common sprint planning and review meeting to facilitate coordination across teams. There are two versions of LeSS, normal LeSS, for up to eight Scrum teams, and LeSS Huge, which the authors claim is applicable for organizations with more than eight Scrum teams and up to a few thousand people in one product. For smaller products, all product members join the same sprint planning, and review meeting. For bigger products, a team representative should be sent to the meetings [9], [38]. LeSS consists of four key components: rules, guides, principles, and experiments [9].

- Rules: Rules provide the foundation and define key elements of LeSS.
- Guides: Guides provide support for framework adoption and a subset of experiments by providing tips and best practices.
- Experiments: The creators encourage teams to experiment and learn from experiences.
- Principles: Principles are references explaining the application of LeSS in different situations.

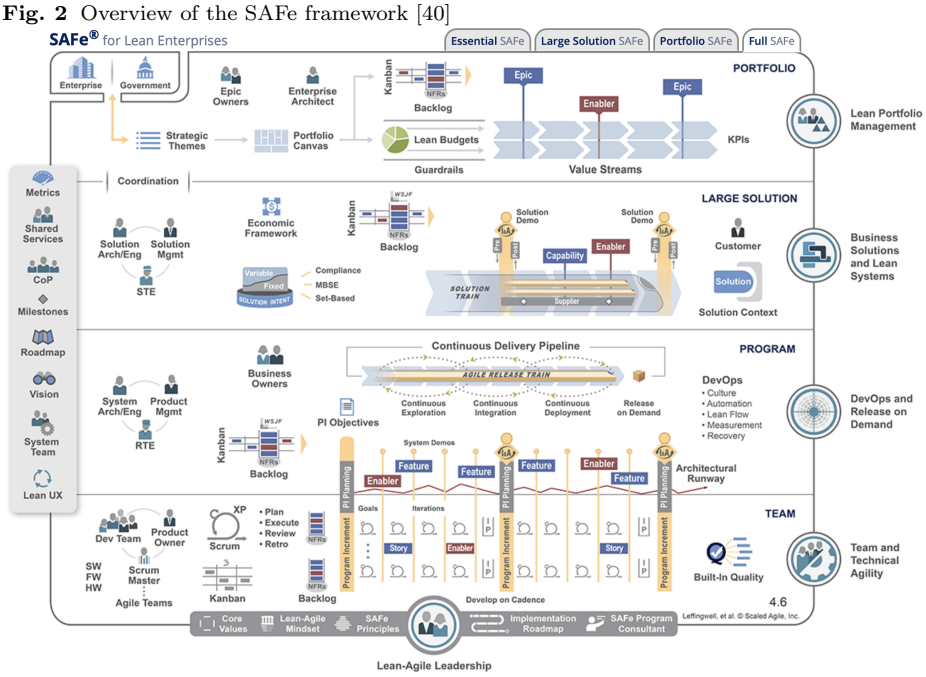
LeSS Huge is applied to team constellations with more than eight teams. To manage hundreds of people, it introduces concepts such as requirements areas (RAs), area POs (APOs) and area product backlogs (APBs). RAs are aligned to customer requirements and have their own sprint cycles, since they aim to enable continuous integration across the entire product. APOs have the same responsibility as POs, but instead of one single backlog they manage APBs and are aligned to an entire RA. Each RA performs its own sprint planning, review, and retrospective meetings [9], [38].

Fig. 1 Overview of the LeSS framework [39]



2.2 Scaled Agile Framework

Fig. 2 Overview of the SAFe framework [40]



Scaled Agile Framework (see Figure 2) was released by Dean Leffingwell in 2011 [41]. In this book, Leffingwell described *"how to apply Lean and Agile practices and principles to the Team, Program, and Portfolio Levels."* [42]. It is currently available as version 4.6 [43]. SAFe integrates aspects from lean and agile principles, and combines those into a framework for organizations and agile projects of different

sizes. For this purpose, SAFe provides four different configurations that aim to cover different organizational development environment needs. The configurations are Essential, Portfolio, Large Solution and Full. Essential SAFe build the foundation for all over configurations and consist of the core components and practices of SAFe. Portfolio SAFe consists of practices which *"helps align portfolio execution to the enterprise strategy by organizing Agile development around the flow of value through one or more value streams."* [41]. Large Solution SAFe is used in development endeavors where product requirement is large and requires multiple Agile Release Trains. According to Leffingwell [41] the most relevant industries for Large SAFe are aerospace, defense, automotive, and government. Lastly, Full SAFe represents the most comprehensive version. It is used in large organizations with hundreds of developers and utilizes all organizational levels. SAFe highlights four levels of organizations: team, program, value stream, and portfolio. Each level manages its own activities, and is aligned with the other levels [41]. In addition, each level maintains a distinct set of roles and activities [41]. For example, Product Managers (PM) exist on the program level. They supervise and direct the work of POs, which operate on the team-level. Agile Release Trains (ART) are a central concept of SAFe and a practice that exist on the program level [41]. Each ART consists of multiple teams that deliver a continuous flow of Potential Shippable Increments (PSI) and follow HIP (Hardening, Innovation, and Planning) iterations. SAFe introduces additional roles, such as the Release Train Engineer (RTE), system teams, release management team, and portfolio management team [41].

2.3 Spotify Model

The Spotify model [44] (see Figure 3 and 4) describes the principles and values of the Swedish music streaming company Spotify. It was first described in 2012 by Henrik Kniberg and Anders Ivarsson. The framework is based on Scrum and Lean practices. The initial goal was to create autonomous cross-functional development teams. The overall structure consists of Squads, Chapters, Guilds, and Tribes.

The smallest unit of development are so-called Squads. These teams consist of up to eight people which have end-to-end responsibility over developed features. Each Squad can access Agile Coaches and POs. While the Agile Coach is responsible for improving Squad work effectiveness, the PO is responsible for overall delivery management (e.g., prioritizing and managing backlog). All Squads have individual missions which align to the overall corporate strategy / goal. The way they reach those goals are decided individually, e.g., Squads are allowed to decide whether use Scrum practices or not. The Spotify model values decision autonomy over alignment: *"Autonomy is motivating. And motivated people build better stuff."* Dependencies between Squads are perceived as slowing down the development progress. Hence, the framework emphasizes reduction of dependencies.

Coordination is increased by using vertical (Squads to Tribes) and horizontal (Chapters and Guilds) alignment. Squads that work in related areas are part of Tribes. People with similar skills are grouped in Chapters. Guilds are community of interests where people can share ideas and thoughts on specific topics [44].

Fig. 3 Overview of the Spotify model, part 1 [45]

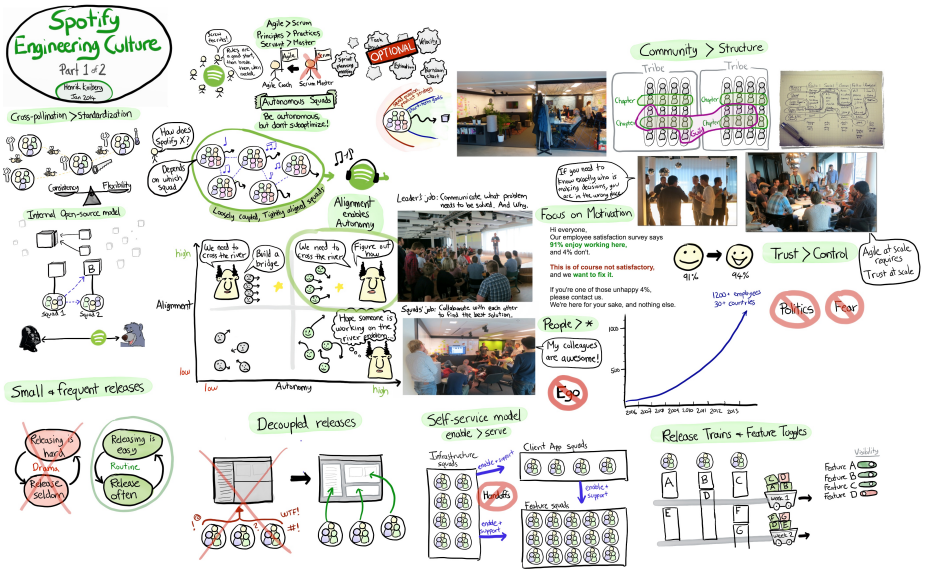


Fig. 4 Overview of the Spotify model, part 2 [46]

2.4 Related Work

Continuing the trend of 2017, Version One's *13th Annual State of Agile Report* (2018) [32] indicated that the share of companies using SAFe is growing, from 29% to 30% (n = 1319). Although overall interest is growing, there is still a lack of scientific research concerning scaling agile frameworks. Kalenda et al. [47] even argue that some base definitions like the term *"Large-Scale Agile"* are still unclear. In the following, we are going to present related research. We primarily considered peer-reviewed publications.

2.4.1 Adoption reasons

Although numerous research papers investigate the adoption of scaling agile frameworks, e.g., by presenting in-depth case studies, most of these contributions do not cover the reasons for companies to choose a particular framework. The majority of applicable publications provide evaluation frameworks to compare different scaling agile frameworks which merely document relevant framework characteristics, but do not elaborate selection criterion in actual real-world scenarios. In a nutshell, publications either provide little information on adoption reasons or implicitly point towards reasons that are relevant for choosing scaling agile frameworks.

The Agile Scaling Knowledgebase ™ Decision Matrix [48] was created in 2013 [49]. It is maintained by Richard Dolman and Steve Spearman. The current version (Version 5) provides an assessment artifact which consists of 20 criteria for consideration when deciding for a scaling agile framework, such as cost of implementation, key differentiators, inter-team coordination, and team-level frameworks (e.g., Scrum, Kanban, XP, etc.).

Alqudah and Razalo [10] provide a structured literature review comparing seven scaling agile frameworks, based on team size, training and certification, adopted methods and practices, required technical practices, and organization type. The aim is to provide organizations a decision framework to help them "pick and select" the right scaling agile framework for the individual context.

Diebold et al. [50] complain that companies select scaling agile frameworks based on criteria like popularity and recommendations by consultancies, rather than relying on structured analysis focused on fulfilling the individual organizational needs. Therefore, the researchers adapted the Agile Decision Matrix [48] and created a subway map to provide a structured approach when selecting scaling agile frameworks, based on a systematic and structured comparison of framework characteristics. The resulting approach compares 12 scaling agile frameworks from high-level (e.g., method applicability, distinguishing factors, and available information / support) and practice-level perspective (i.e., individual practices, practice groups, and categories).

Theobald et al. [26] identified common practices across 12 different scaling agile frameworks in three categories (i.e., team-level, scaled-level, team and scaled-level). Among others, the authors extended the subway map by Diebold et al. [50] by depicting identified framework practices. They suggest that instead of following the recommendations of consultants, practitioners should use their findings and compare whether a framework provides the right practices that meet the needs of the individual organization.

Conboy and Carroll [16] present a multiple case study investigating 13 agile transformations and the respective challenges and recommendations for scaling agile framework adoption. They found that the absence of any scaling agile framework assessment model is problematic, since framework selection nowadays is often based on an ad hoc basis: *"[..], wherein one or two read a book or attended a talk, and then made a decision. Sometimes the decision could not even be traced to a source."*. The researchers provide the example of a developer who wished for more due diligence in framework selection. *"Comparison and justification"* of the framework selection, incl. the communication among relevant stakeholder groups, is part of overcoming central challenges addressed by the authors.

Pries-Heje and Krohn [51] provide a case study of a software company adopting SAFe. They discuss three main challenges and present lessons learned. One of the lessons learned is that the case organization underestimated technical challenges of continuous integration and that this part was, in contradiction to other sections, not well documented. Since the adoption reasons were not in scope of the overall research paper the authors do not further investigate following statement: *"After looking at different possibilities it was decided to use SAFe."*

Paasivaara [30] provides another case study comparing challenges and success factors of two SAFe adoptions at a software company. The adoptions happened in two different business lines. Each business line decided to choose SAFe independently in the form of a top level management decision. According to Paasivaara [30], no other framework was considered.

Uludağ et al. [9] provide a case study summarizing the adoption and application of LeSS in four different product at a car manufacturer. Related challenges are discussed alongside success factors. Prior to the adoption, another department successfully introduced LeSS and served as a lighthouse project within the company. In addition, the framework offered sufficient degree of documentation to coordinate multiple teams while maintaining a moderate complexity level. As a result, the four products decided to apply the same framework.

Finally, Uludağ et al. [35] provide a primary analysis investigating the roles of architects in 20 different scaling agile frameworks. They defined maturity scores based on contribution, cases, documentation, training courses and certifications, and community, forum or blogs. Three frameworks achieved a maturity score of 1.0, namely SAFe, DAD, and LeSS.

2.4.2 Expectations

Similar to the adoption reasons (see Section 2.4.1), expectations towards scaling agile frameworks were rarely addressed in scientific literature. Existing literature, predominantly based on case studies and literature reviews, only implicitly cover expectations by investigating success factors of scaling agile framework adoptions. We found one case study which assessed the satisfaction of the respective agile adoption in one paragraph. To the best of our knowledge there is no scientific research that performs a critical, retrospective analysis of the adoption process and the overall satisfaction. We elaborate our definition of satisfaction in context of this research paper in Section 3.2.

Paasivaara [30] poses one of the few cases where expectations were assessed to a limited extend. According to an internal survey after the SAFe adoption, the satisfaction was split. One business line perceived the SAFe adoption as a success, while the others perceived it as a burden. The responses suggest that the employees failed to understand the benefits of SAFe, but instead perceived the new working methods as limitations to previously existing autonomy. *"With fixed increments they felt moving backward, towards the old waterfall."*

3 Research Method

Existing research predominantly consists of literature reviews and in-depth case studies considering selected scaling agile frameworks at a limited number of or-

ganizations. We were wondering if there is any empirical evidence proving that particular scaling agile frameworks are chosen over others due to specific reasons. Furthermore, we were interested in understanding what practitioners anticipated before they introduced scaling agile frameworks. For this purpose, we decided to conduct a survey as this is the best way to collect information from a broad population [52]. The survey aims to uncover the underlying reasons for scaling agile framework adoptions, as well as the fulfillment of expectations across different enterprise contexts.

As a general remark, we do not claim that the list of scaling agile frameworks is exhaustive. Instead, a subset of frameworks was chosen based on prior experience and of which we think are most relevant.

3.1 Sampling

Our target population are scaling agile framework practitioners who were directly involved in scaling agile framework adoptions and working environments. We dropped the requirement of having the job title related to any specific framework or role because it might lead to bias. Also, scaling agile frameworks typically involve different roles (e.g., strategic management, development staff, or system architect). We chose non-probabilistic convenient sampling for data collection [52], [53]. This allows to collect the data from people are who available and willing to answer the survey. We adopted snowball sampling, where some of the respondents recommended other subjects to collect data. Probability sampling would be favorable as it allows for generalization of derived results. Unfortunately, the required prerequisites are not given, i.e., there is no single list with a large amount of scaling agile practitioners available that we could use without risking biased responses

3.2 Survey Design

The survey design followed the principles elaborated by Linåker et al. [53]. The survey questionnaire consisted of 25 questions and six main sections. The sections are introduction, agile transformation background, reasons for framework adoption, framework evaluation, technical background, and general background / closing. Since the survey captured more data than the scope of this paper, we are only going to elaborate on the relevant sections. In the context of this survey, *"satisfaction with used frameworks"* refers to a retrospective consideration of expectations prior to the scaling agile framework adoption.

First, the introduction section consisted of a small explanation of the research goal. Second, the agile transformation background section listed 21 scaling agile frameworks based on Version One [32] and Uludağ et al. [35] in the shape of multiple choice questions. Free text field in all sections allowed participants to specify any other choice that was not listed. We also investigated participant satisfaction with their respective framework by asking them if it (a) met their expectations, (b) if they would like to move back to their old way of working, (c) if they would like to move to another framework, and (d) whether they would recommend their chosen framework to similar organizations. In the context of this survey, *"satisfaction with used frameworks"* refers to a retrospective consideration of

expectations prior to the scaling agile framework adoption. These four questions, similar to all questions following, were designed as six-point Likert scale questions. The use of an evennumbered scale aimed to better gather the participants opinions and to avoid processing neutral midpoints. *"Strongly disagree"* represents the most negative opinion (score 1) and *"Strongly agree"* the most positive one (score 5). *"I don't know"* was processed as a neutral point (score 0). Next, we asked the respondents in which area of their company they applied the framework (e.g., HR, Finance, R&D, etc.) and their previous development model (e.g., Waterfall, Scrum, etc.). We provided different length categories (e.g., <1 year, 3 - 5 years, etc.) to capture the duration since when they applied respective framework. Third, the reasons to adopt scaling agile frameworks were investigated. For this purpose, this section contained a list of six reasons based on previous practitioner surveys [32]. Then, the technical background investigated the number of teams working within the scaling agile framework, the geographic distribution of teams, and the distribution of development efforts across multiple locations. The general background section contained questions about the organization domain, the primary role, and the country the survey participant primarily works in. Lastly, the closing section contained a short thank you message

3.3 Survey Validation

Following the survey design, a respective validation aims to find potential flaws and assess the reliability and validity of a survey [53]. A preliminary validation was conducted with the help of two industry experts and one domain expert focusing in particular on understandability, acceptability, reliability, and effectiveness.

3.4 Data Collection

The data collection took place from May 22$^{\text{nd}}$ 2019 until September 27$^{\text{th}}$ 2019 using a third-party, web-based tool called *"LimeSurvey"*. To reach our target population, we used several approaches: (1) conferences, (2) Meetup groups, (3) social media groups, and (4) personal network.

We promoted the survey on three international conferences XP 2019 [54], Agile 2019 [55], and ICGSE 2019 [56]. XP 2019 took place from May 21$^{\text{st}}$ to May 25$^{\text{th}}$ 2019 in Montreal, Canada. Agile 2019 took place from August 5$^{\text{th}}$ to August 9$^{\text{th}}$ 2019 in Washington D.C., USA. Finally, the International Conference on Global Software Engineering (ICGSE 2019) took place from May 24$^{\text{th}}$ to 26$^{\text{th}}$ 2019 in Montreal, Canada.

Two researchers participated and promoted the survey at XP 2019. They also spoke to people during the breaks and mailed the link to those interested in answering the survey. On June 4$^{\text{th}}$ 2019, a link to the survey was distributed among all XP 2019 participants. At ICGSE 2019, two researchers actively promoted the survey. Later they spoke to people during the breaks and sent out the survey link by email to interested parties. At Agile 2019, one of the researchers prepared and handed out small cards with descriptions and a QR code, that would link to the online survey. A small booth was put up, which showed the survey description and link. In addition, the survey link was sent out to all conference participants during

all three conference days of Agile 2019. Finally, one week after the conference the link was sent to all participants again.

Besides active in-person promotion at Agile conferences, the survey link was also published on selected social media platforms, promoted at Agile Meetups and distributed to professionals. On June 10th 2019 the survey link was posted in the worldwide LinkedIn group *"Lean and Agile Software Development"* [57]. This is the largest online community of Lean and Agile practitioners, with more than 157.000 (157.414 on Sept 13th 2019) members from different parts of the world. The survey was also promoted at two Agile Meetups, where practitioners share their experiences, ideas and knowledge on issues regarding agile. One was held in Helsinki (Finland) on August 28th 2019 (32 participants) and one held in Copenhagen (Denmark) on June 11th 2019 (30 participants). The survey leaflets, containing the link and QR code, were sent to all participants in the respective Meetups. In parallel, professionals from different organizations worldwide were approached via email, LinkedIn, and other social media channels. The chosen individuals were asked to fill out the survey. Snowballing of contacts served the purpose of finding new contacts, i.e., the personal network of existing contacts was leveraged.

By the end of the deadline, 4037 responses were collected. However, 3836 responses were not completed and deleted due to incompletion. After data preparation 199 data points went into the subsequent data analysis

3.5 Data Preparation

For cleaning the data a two day workshop (September 5th to 6th 2019) was held in Helsinki. All authors participated in the workshop. Data preparation followed five steps: (1) deletion, (2) completion, (3) transformation, (4) creation, and (5) testing.

First, data points not required for analysis were deleted. The raw data extract from LimeSurvey contained 143 different variables. Each variable was analyzed according to its relevance for the overall research objective, analysis potential and then, if required, removed accordingly. Next, potentially biased responses and data points which can be flagged as duplicate response submissions were deleted. Biased responses were such submitted by respondents which could be clearly associated to organizations with strong relationships to specific scaling agile frameworks. Second, data points with empty fields were completed. The questions within the survey were not entirely programmed as mandatory questions. As an example, there were multiple data points with blank organization fields. Whenever possible, fields were completed by analyzing the user input from other survey fields. Third, the data was transformed. All variables were stored as nominal variables, which also includes the dependent variables (DV) under investigation. The independent variables (IV) were captured as responses from questions with Likert scale reply options. To replicate the ordering of variable options these variables were transformed into ordinal scales, by encoding the reply options as numerical categories. Next, new variables were created. Each variable holding string values was individually analyzed on a row by row basis. The content was coded into new variables. Whenever appropriate responses were clustered into categories to increase the information value of the quantitative analysis. Furthermore, certain questions with multiple choice responses

were stored as separate binary variables, e.g., one question asked for the frameworks used by respondents. These variables were manually coded and harmonized into one single variable. Finally, we tested, if all the variables were coded correctly. We took each of the IVs, e.g., primary framework category vs dependent variable.

3.6 Data Analysis

Following the advice of Linåker et al. [53], we used the statistical analysis software IBM SPSS (Version 26) to analyze collected survey data. To locate significant differences for the independent variables under investigation, following approach was applied: (1) Determine independent variable, (2) determine dependent variable, (3) formulate null hypothesis (H_0) and alternative hypothesis (H_A), and (4) accept or reject null hypothesis. All IVs in scope of the analysis are nominal, while the DVs are ordinal.

Since we cannot assume that the underlying data is normality distributed, commonly applied statistical tests, e.g., student t-test, are not applicable. Instead, the Mann-Whitney U test, a non-parametric test, was used to calculate p-values. This test is applied to determine whether the median values of two independent groups significantly differ, without having a normality distribution [53], [58]. The U value represents the test statistic, while the z value is used to determine the significance of the test statistic.

To not only determine significant differences we also calculate the respective Pearson correlation coefficient, r, according to Fritz et al. [59], [60]. The correlation coefficient (effect size) describes the association or relation between two variables. In this research paper, effect size can be interpreted as the influence different groups of independent variables (e.g., framework choice, size of development organization, etc.) have on dependent variables (e.g. adoption reasons, or satisfaction). To interpret the resulting effect sizes we rely on the interpretation by Cohen [61] who defined three effect sizes: small (r = 0.1), medium (r = 0.3), and large (r = 0.5).

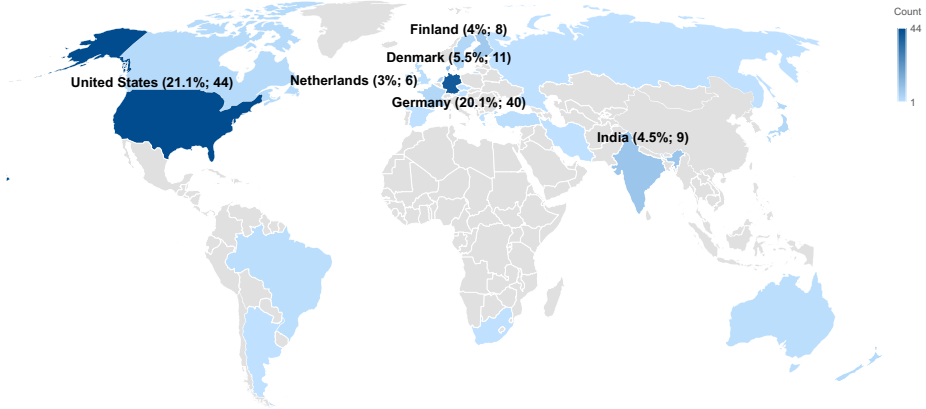
4 Results

4.1 The Sample

As mentioned before, the overall survey data and variables exceed the scope of this paper. In the following, we elaborate the variables relevant for this study. The final sample consists of 199 responses coming from 26 countries (see Figure 5). The five biggest countries are the United States (n = 44; 21.1%), followed by Germany (n = 40; 20.1%), Denmark (n = 11; 5.5%), India (n = 9; 4.5%), Finland (n = 8; 4%), and the Netherlands (n = 6; 3%) (see Table 18). The three biggest industries employ nearly 50% of the entire sample, namely 27.6% (n = 55) work in Finance / Banking / Insurance / Accounting or Property, 12.6% (n = 25) in Technology, and 9.5% (n = 19) in Automotive (see Table 19).

Concerning the role, the majority of respondents have a process related perspective, e.g., Agile Coach or Scrum Master. This role category makes up 57.4% (n = 113) of the sample. 25.4% (n = 50) have a management-related role, 10.7% (n = 21) focus on development on team-level, and 6.6% (n = 13) are developer-related roles

Fig. 5 Country distribution of survey participants



on product-level (see Table 20). Not only do responses differ in country, industry, and role, they also use different scaling agile frameworks. Out of the 21 provided frameworks, 11 (52.38%) received any responses. SAFe (n = 96; 52.2%), LeSS (n = 28; 5.2%), Internally developed models (n = 12; 6.5%), and Spotify (n = 10; 5.2%) occurred most often (see Table 21). 47.7% (n = 95) of survey participants used plan-driven developments methods before, while 25.1% (n = 50) already used agile development methods (see Table 22). 59.8% of the plan-driven teams used SAFe (n = 55) and 16.8% used LeSS (n = 16), while the residual framework share range from 1.1% to 6.5% (see Table 32). 68.2% of teams that are experienced with agile development used either SAFe (n = 14; 31.8%), LeSS (n = 5; 11.4%), Spotify (n = 6; 13.6%) model or Nexus (n = 5; 11.4%) (see Table 33). Concerning the responses which primarily use SAFe, 57.3% (n = 55) were plan-driven and 14.6% agile (n = 14) (see Table 34). The majority of LeSS organizations used to be plan-driven organizations (57.1%, n = 16), followed by 17.9% (n = 5) that already were agile (see Table 40). 15.1% (n = 30) of the respondents used the respective framework for less than one year, 41.2% (n = 82) used it between one and two years, 32.7% (n = 65) used it three to five years and 11.1% (n = 22) used it for more than five years (see Table 23). 40.6% (n = 39) of SAFe organizations have 1-2 years of experience. 38.5% (n = 37) have 3-5 years of experience and 13.5% (n = 13) of organizations that use SAFe have less than one year of experience (see Table 35). Similar, 57.1% (n = 16) of LeSS organizations had 1-2 years of experience, while those with 3-5 years of experience account for 25% (n = 7) (see Table 42). The sample of our survey consisted of 35.2% (n = 70) that indicated of having one to nine teams, 36.2% (n = 72) have 10 to 50 teams, and the remaining 28.6% (n = 57) have more than 50 teams (see Table 24). Filtering the sample for SAFe organization indicates that 38.5% (n = 37) have 10-50 teams, 35.4% (n = 34) have more than 50 teams, and 26% have 1-9 teams (n = 25) (see Table 36). 42.9% (n = 12) of LeSS organizations had 1-9 teams, while the other two categories account for 28.6% each (n = 8) (see Table 41). 17.6% (n = 35) of the teams are not geographically distributed, 76.9% (n = 153) are partially distributed and 5.5% (n = 11) are completely distributed (see Table 25). Of the SAFe organizations, the majority (80.2%; n = 77) are partially distributed. 14.6% (n = 14) are not

distributed while the remaining 5.2% (n = 5) are fully distributed (see Table 37). For LeSS organizations this share is at 75% (n = 21), while the remaining 25% (n = 7) have no distributed teams. 0% of LeSS organizations are fully distributed (see Table 43). 75.9% (n = 151) of development endeavors are geographically distributed to several sites, while the remaining 24.1% (n = 48) are not (see Table 26). 77.1% (n = 74) of the SAFe organizations have distributed development sites (see Table 38). Concerning Less organizations, the share of organizations with multiple sites is 75% (n = 21) (see Table 44). 73.9% (n = 147) of survey participants applied the scaling agile framework within the IT department, 57.8% (n = 115) in Product Development / R&D, and 13.6% (58) in Business (see Table 27).

Respondents generally agreed (*"strongly* agree" or *"agree"*) with all provided reasons (see Figure 6 and Table 28). Concerning the reasons to adopt a specific scaling agile framework, "*Scale to more people*" received 152 positive responses (76.4%). Of that share, SAFe was indicated in 53.5% (n = 76) and LeSS in 12.5% (n = 19) of the responses (see Table 29). 79.2% (n = 76) of SAFe organizations indicated that scaling was important for their framework choice (see Table 39). This goes against 64.3% (n = 19) of the LeSS organizations (see Table 45). The second and third reasons with most positive responses are documentation with 68.8% (n = 137) (i.e., how well a specific framework is defined and documented) and 68.3% (n = 136) for competitiveness (i.e., degree of increased competitiveness a single framework grants). The adoption reason that received most negative responses (either *"strongly disagree"* or *"disagree"*) was *"Widely adopted"* (n = 48; 24.2%). Spotify model (n = 6; 26.1%) and internally created scaling agile framework (n = 5; 21.7%) received the most disagreeing responses out of that proportion (see Table 30). 64.6% (n = 63) of the SAFe organizations agreed that popularity was an important adoption reason (see Table 39). On the contrary, only 14.2% (n = 4) of the LeSS organizations agreed with the same adoption reason. 42.9% (n = 13) disagreed (see Table 45).

Fig. 6 Overview of responses for framework adoption reasons

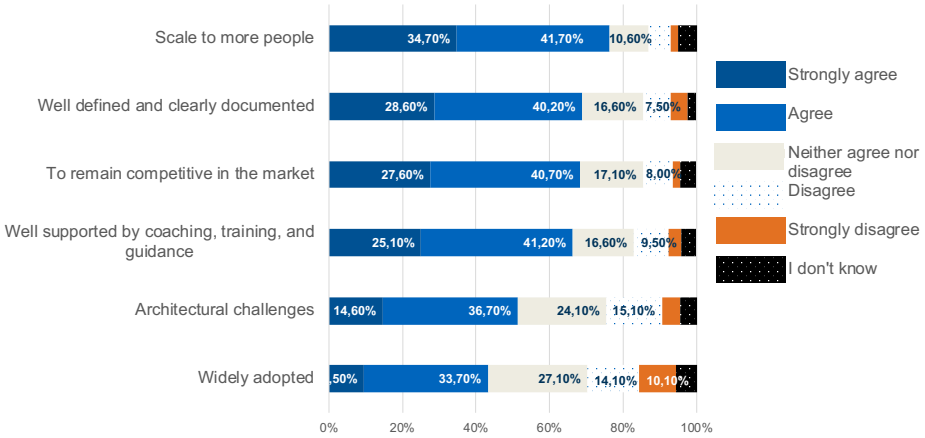
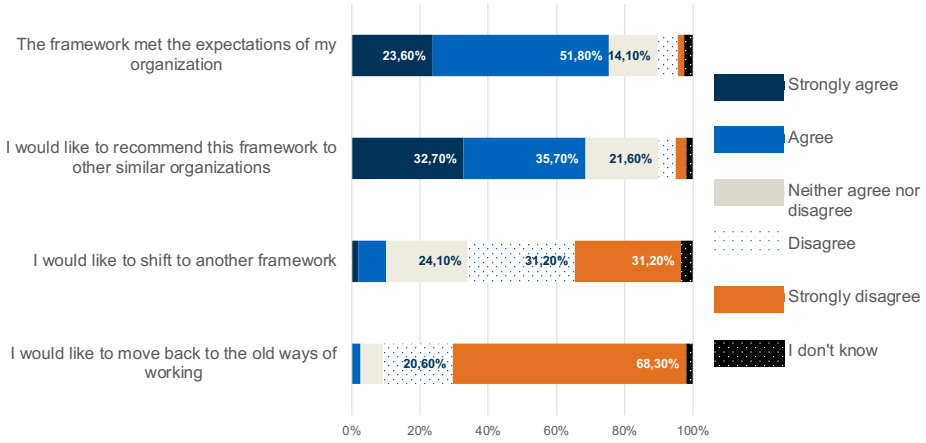


Figure 7 shows the distribution of expectations. Overall, the majority of organizations (75.4%; n = 150) felt that their expectations towards their chosen

scaling agile framework were met (see Table 46). 73.9% (n = 71) of SAFe adopters indicated that the framework fulfilled their expectations (see Table 39), while 75% (n = 21) of LeSS adopters the same (see Table 45). Less than 1/6 (14.1%, n = 28) of organizations neither agreed nor disagreed about the fulfillment of their expectations. The majority of undecided respondents used the respective framework for one to two years (42.9%; n = 12) (see Table 48). Again, the SAFe framework accounts for 54.2% (n = 13) of respondents that *"Neither agree nor disagree"* about the fulfillment of their expectations (see Table 47). 8% (n = 16) of the overall survey participants stated that the framework did not met their expectations and this share consists of 73.3% (n = 11) SAFe organizations and 13.3% (n = 2) SoS (see Table 52). 11.5% (n = 11) of SAFe adopters were not satisfied with their framework (see Table 39). There was only one case where LeSS did not fulfilled the expectations (3.6%) (see Table 45).

Fig. 7 Overview of responses for framework expectations



Also, 88.9% (n = 177) (strongly) disagreed with the statement *"I would like to move back to the old ways of working"*. Here, the share of unconvinced responses in the one or other direction is smaller (6.5%; n = 13). 2.5% (n = 5) indicated that they would like to return to their old development method and 2 come from a plan-driven with a framework experience of 1-2 years and 3-5 years (see Table 51). The number of SAFe and LeSS adopters which were willing to move back to their old development model was low. Two SAFe organizations (2.1%) (see Table 39) and one LeSS organization (3.6%) (see Table 45) wanted to move back. No survey participant indicated "strongly agree" with moving back to the initial state. 62.4% (n = 124) would choose their framework again, whereas 24.1% (n = 48) question their choice. 10% (n = 20) would like to switch to another framework and of that 55% (n = 11) use SAFe (see Table 49). Looking further into the respective SAFe organizations indicates that not a single company used the framework for more than five years (see Table 50). Nevertheless, 61% (n = 59) of SAFe organizations disagree with shifting to another framework and 11.5% (n = 11) agree (see Table 39). 82.1% of LeSS organizations disagree, while only one response (3.6%) indicated willingness to shift (see Table 45). Lastly, 68.4% would recommend their scaling

agile framework to other organizations. This is 6 % point higher than the share of agreeing responses to stay with their framework choice, but 7% less than the number of responses where expectations were met. 16 responses indicated that they would not recommend their framework, 50% (n = 8) are made up of SAFe organizations, 18.8% (n = 3) use Spotify and 12.5% (n = 2) use SoS or internally developed frameworks (see Table 53).

4.2 Inferential Analysis

In the following, we present the results of the inferential analysis. We only display statistically significant results (p <0.05) in our tables and figures, in the text we provide textual summaries. The complete test result are attached in the Appendix.

4.2.1 Previous Development Model

We divide survey participants in organizations that either used agile or plan-driven process models previously. Based on this separation, 50 agile and 95 plan-driven organizations answered the survey. To compare the answers from both groups we use the Mann-Whitney U test. We investigate the differences between both type of organizations on the answers offered to the survey questions as elaborated in Section 3.2. Table 1 presents the Mann-Whitney U test results (U, Z, and p) and calculated effect size r. Results without statistical significance are omitted (for full test results see Table 54 and 55).

Table 1 Significant results of the Mann-Whitney U test on adoption reasons are displayed on the first four columns. Median, average, mean rank values for agile (A) and plan-driven (P) organizations are displayed on the last six columns.

	U	Z	p	Median A	Median P	Avg. A	Avg. P	Mean rank A	Mean rank P	Effect size
Because it is widely adopted	1874.000	-2.266	0.023	3.00	3.00	2.56	3.19	62.44	78.56	0.188

The result show that wide adoption was more relevant for plan-driven organizations than agile ones. The respective effect size is categorized as small which means that previous development model had small effect on adoption reasons and expectations. Furthermore, the previous development model does not have a significant effect on the preference of documentation and framework support during framework selection process. Similar, agile and plan-driven methodologies as previous development models do not have significant effect on how organizations perceive the fulfillment of their expectations towards the scaling agile framework. Hence, H1.1 and H1.2 could not be supported. At this point, we also want to highlight that well support yielded a p value of 0.06, which could be an indicator that plan-driven organizations agree significantly more with this adoption reason than former agile organizations (see Table 54).

4.2.2 Number of Teams

Next, we divide the respondents into two groups depending on the size of their product development organization. The first group are small organizations with

1-9 teams (n = 70). Second group are large organizations with >50 teams (n = 57). Again, a Mann-Whitney U test was conducted, but showed no significant for neither of the investigated dependent variables (see Table 56 and 57). There is an indication that large organizations agree significantly more with popularity being an important adoption reason than small organizations, because of a p value of 0.06.

4.2.3 Framework adoption experience

Again, we divide the data into two distinct groups depending on framework adoption experience. We separate between less experienced (<1 year framework experience) (n = 30) and more experienced (>5 years framework experience) (n = 22) agile organizations and conduct a Mann-Whitney U test for adoption reasons and expectations (see Table 2 and 3 for significant results, and Table 58 and 59 for full results).

Table 2 Significant results of the Mann-Whitney U test on adoption reasons are displayed on the first four columns. Median, average, mean rank values for less experienced (L-E) and more experienced (M-E) organizations are displayed on the last six columns.

.	U	Z	p	Median L-E	Median M-E	Avg. L-E	Avg. M-E	Mean rank L-E	Mean rank M-E	Effect size
Because the framework addresses architectural challenges	185.000	-2.764	0.006	3.00	4.00	2.77	3.73	21.67	33.09	0.383
Because it is well supported by coaching, training, and guidance	215.500	-2.208	0.027	4.00	4.00	2.93	3.68	22.68	31.70	0.306

Table 3 Significant results of the Mann-Whitney U test on expectations are displayed on the first four columns. Median, average, mean rank values for less experienced (L-E) and more experienced (M-E) organizations are displayed on the last six columns.

	U	Z	p	Median L-E	Median M-E	Avg. L-E	Avg. M-E	Mean rank L-E	Mean rank M-E	Effect size
The framework met the expectations of my organization	186.500	-2.787	0.005	4.00	4.00	3.30	4.14	21.72	33.02	0.386
I would like to shift to another framework	212.000	-2.262	0.024	3.00	1.50	2.53	1.68	30.43	21.14	0.313
I would like to recommend this framework to other similar organizations	166.000	-3.142	0.002	3.00	4.00	2.97	4.05	21.03	33.95	0.435

Investigating the differences between less experienced and more experienced organizations concerning adoption reasons resulted in two significant results. First, a framework that addresses architectural challenges was more relevant for more experienced organizations than less experienced organizations. Similar, available coaching, training, and guidance were more relevant for the framework choice for more experienced organizations compared to less experienced organizations. The effect size for both results are medium (r >0.3) indicating that framework experience had less powerful effect on both variables.

Testing the expectations returned three significant results. First, the group that used their respective framework for more than five years is significantly more satisfied than the group that used it for less than a year. Second, the group of more experienced organizations disagreed significantly more with shifting framework than less experienced ones. Lastly, our test results show that there is evidence for

more experienced teams agreeing more with recommending their framework than teams with less than one year of experience.

4.2.4 Geographical Team and Site Distribution

We group the respondents into two groups. The first group are fully distributed teams (n = 11), meaning their team members are split across different geographic sites. The second group does not have any distributed teams (n = 35). The test results are captured in Table 4.

Table 4 Significant results of the Mann-Whitney U test on expectations are displayed on the first five columns. Median, average, mean rank values for not distributed (N-D) and fully distributed (F-D) organizations are displayed on the last six columns.

	U	Z	p	p*	Median N-D	Median F-D	Avg. N-D	Avg. F-D	Mean rank N-D	Mean rank F-D	Effect size
I would like to move back to the old ways of working Exact Sig. [2*(1-tailed Sig.)]	80.000	-3.361	0.001	0.003	1.00	2.00	1.23	2.27	20.29	33.73	0.495

Completely distributed teams show significantly higher agreement with the statement *"I would like to shift to another framework"* than teams that are not distributed, using the exact sampling distribution of U (Dinneen and Blakesley [62]). The respective effect size has a tendency to be large ($r = 0.5$). This means that the different groups of geographical team distribution have a strong effect on the dependent variable. For full test results please see Table 60 and 61.

Similar, we investigate the effect of site distribution. Survey participants could answer with yes and no. There was no statistical significant difference between teams that use multiple sites for development concerning the adoption reasons and expectations (see Table 62 and 63).

4.2.5 Organizational Areas

Lastly, we group responses into two groups depending on the number of areas covered by the scaling agile framework. In total respondents were able to select up to 12 areas. Responses with six or more areas are the broad adoption group (n = 34). Replies with six or less areas are the narrow adoption group (n = 165). Mann-Whitney U test results are listed in Table 5 and 6. For full results please refer to Table 64 and 65.

Table 5 Significant results of the Mann-Whitney U test on adoption reasons are displayed on the first four columns. Median, average, mean rank values for narrow (N) and broad area (B) organizations are displayed on the last six columns.

.	U	Z	p	Median N	Median B	Avg. N	Avg. B	Mean rank N	Mean rank B	Effect size
To remain competitive in the market	1784.500	-3.507	0.000	4.00	4.50	3.58	4.29	93.82	130.01	0.248

For respondents that apply scaling agile frameworks in more than six areas, remaining competitive is significantly more important than respondents from the narrow adoption group. The first group also agrees significantly more with the

Table 6 Significant results of the Mann-Whitney U test on expectations are displayed on the first four columns. Median, average, mean rank values for narrow (N) and broad area (B) organizations are displayed on the last six columns.

	U	Z	p	Median N	Median B	Avg. N	Avg. B	Mean rank N	Mean rank B	Effect size
The framework met the expectations of my organization	2133.500	-2.389	0.017	4.00	4.00	3.76	4.06	95.93	119.75	0.169
I would like to shift to another framework	2187.500	-2.100	0.036	2.00	1.00	2.14	1.79	103.74	81.84	0.148

statement *"The framework met the expectations of my organization"* than the group with narrow framework adoption. Finally, the group that adopted scaling agile in more areas also disagrees more significantly with the statement *"I would like to shift to another framework"* than the other group. The respective effect sizes are small. Scale to more people returned a p value of 0.972, which indicates that both groups have a very similar distribution concerning this adoption reason.

4.2.6 Scaling Agile Frameworks

For the next set of statistical tests we divide the survey participants into groups corresponding to their framework choice. We decided to choose the four frameworks with the most replies. Thus, we have 96 organizations that use SAFe, 28 LeSS, 12 Internal, and 10 Spotify.

4.2.6.1 SAFe

Table 7 Significant results of the Mann-Whitney U test on adoption reasons are displayed on the first four columns. Median, average, mean rank values for SAFe and LeSS organizations are displayed on the last six columns.

	U	Z	p	Median SAFe	Median LeSS	Avg. SAFe	Avg. LeSS	Mean rank SAFe	Mean rank LeSS	Effect size
Because it is widely adopted	579.500	-4.760	0.000	4.00	2.00	3.56	2.18	70.46	35.20	0.427

Table 8 Significant results of the Mann-Whitney U test on expectations are displayed on the first four columns. Median, average, mean rank values for SAFe and LeSS organizations are displayed on the last six columns.

	U	Z	p	Median SAFe	Median LeSS	Avg. SAFe	Avg. LeSS	Mean rank SAFe	Mean rank LeSS	Effect size
I would like to shift to another framework	874.000	-2.983	0.003	2.00	1.00	2.23	1.61	67.40	45.71	0.267

First, comparing SAFe with LeSS returned two significant results (see Table 7 and 8). Respondents that use SAFe agree significantly more with wide adoption as an adoption reason than LeSS respondents. Furthermore, they (SAFe organizations) also agree significantly more with wanting to shift framework. The p value for documentation in favor of SAFe points ($p = 0.058$) towards significance (see Table 66 and 67).

Second, the comparison between SAFe and Spotify returned three significant results (see Table 9 and 10 for significant results, and Table 68 and 69 for full results).

Documentation and support were more relevant for SAFe organizations than Spotify organizations. SAFe organizations also showed more significant agreement with the statement *"I would like to recommend this framework to other similar organizations"*.

Table 9 Significant results of the Mann-Whitney U test on adoption reasons are displayed on the first four columns. Median, average, mean rank values for SAFe and Spotify organizations are displayed on the last six columns.

	U	Z	p	Median SAFe	Median Spotify	Avg. SAFe	Avg. Spotify	Mean rank SAFe	Mean rank Spotify	Effect size
Because the framework is well defined and clearly documented	112.000	-4.262	0.000	4.00	2.00	4.14	2.30	57.33	16.70	0.413
Because it is well supported by coaching, training, and guidance	207.000	-3.154	0.002	4.00	3.00	3.99	2.70	56.34	26.20	0.306

Table 10 Significant results of the Mann-Whitney U test on expectations are displayed on the first four columns. Median, average, mean rank values for SAFe and Spotify organizations are displayed on the last six columns.

	U	Z	p	Median SAFe	Median Spotify	Avg. SAFe	Avg. Spotify	Mean rank SAFe	Mean rank Spotify	Effect size
I would like to recommend this framework to other similar organizations	194.500	-3.222	0.001	4.00	3.00	3.90	2.60	56.47	24.95	0.312

Third, contrasting SAFe against internally created frameworks returned four significant results (see Table 11 and 12 for significant results, and Table 70 and 71 for full results). Again, SAFe received significant higher agreement for documentation and support than internal frameworks. In addition, SAFe organizations state that popularity was a relevant adoption reason compared to Internal framework organizations. Finally, SAFe organizations also showed significant higher willingness to recommend the framework. All effect sizes were medium.

Table 11 Results of the Mann-Whitney U test on adoption reasons are displayed on the first four columns. Median, average, mean rank values for SAFe and Internal organizations are displayed on the last six columns.

	U	Z	p	Median SAFe	Median Internal	Avg. SAFe	Avg. Internal	Mean rank SAFe	Mean rank Internal	Effect size
Because it is widely adopted	154.000	-4.348	0.000	4.00	1.50	3.56	1.83	58.90	19.33	0.418
Because the framework is well defined and clearly documented	101.000	-4.945	0.000	4.00	2.50	4.14	2.08	59.45	14.92	0.475
Because it is well supported by coaching, training, and guidance	206.500	-3.846	0.000	4.00	2.00	3.99	2.33	58.35	23.71	0.370

Table 12 Significant results of the Mann-Whitney U test on expectations are displayed on the first four columns. Median, average, mean rank values for SAFe and Internal organizations are displayed on the last six columns.

	U	Z	p	Median SAFe	Median Internal	Avg. SAFe	Avg. Internal	Mean rank SAFe	Mean rank Internal	Effect size
I would like to recommend this framework to other similar organizations	353.000	-2.280	0.023	4.00	3.00	3.90	3.08	56.82	35.92	0.219

4.2.6.2 LeSS

Testing the group of LeSS and Spotify users returned two significant results, using the exact sampling distribution of U (Dinneen and Blakesley [62]) (see Table 13 and 14 for significant results, and Table 72 and 73 for full results). Documentation is significantly more important for the LeSS group than Spotify. Similar, LeSS organizations also agree significantly more with the willingness to recommend the framework than Spotify organizations. Both results had a (rather) large effect size.

Table 13 Significant results of the Mann-Whitney U test on adoption reasons are displayed on the first four columns. Median, average, mean rank values for LeSS and Spotify organizations are displayed on the last six columns.

.	U	Z	p	p*	Median LeSS	Median Spotify	Avg. LeSS	Avg. Spotify	Mean rank LeSS	Mean rank Spotify	Effect size
Because the framework is well defined and clearly documented Exact Sig. [2*(1-tailed Sig.)]	50.000	-3.071	0.002	0.002	4.00	2.00	3.79	2.30	22.71	10.50	0.498

Table 14 Significant results of the Mann-Whitney U test on expectations are displayed on the first four columns. Median, average, mean rank values for LeSS and Spotify organizations are displayed on the last six columns.

.	U	Z	p	p*	Median LeSS	Median Spotify	Avg. LeSS	Avg. Spotify	Mean rank LeSS	Mean rank Spotify	Effect size
I would like to recommend this framework to other similar organizations Exact Sig. [2*(1-tailed Sig.)]	41.000	-3.435	0.001	0.001	4.50	3.00	4.21	2.60	23.04	9.60	0.557

Comparing LeSS to internally created framework revealed three significant results (see Table 15 and 16 for significant results, and 74 and 75 for full results). Documentation and support are significantly more important for LeSS practitioners. Similar LeSS organizations show a significant higher agreement towards recommending the framework to similar organizations.

Table 15 Results of the Mann-Whitney U test on adoption reasons are displayed on the first four columns. Median, average, mean rank values for LeSS and Internal organizations are displayed on the last six columns.

	U	Z	p	p*	Median LeSS	Median Internal	Avg. LeSS	Avg. Internal	Mean rank LeSS	Mean rank Internal	Effect size
Because the framework is well defined and clearly documented	52.000	-3.156	0.000	0.000	4.00	2.50	3.79	2.08	24.64	10.83	0.499
Because it is well supported by coaching, training, and guidance Exact Sig. [2*(1-tailed Sig.)]	80.000	-2.674	0.008	0.008	4.00	2.00	3.68	2.33	23.46	13.17	0.422

Table 16 Significant results of the Mann-Whitney U test on expectations are displayed on the first four columns. Median, average, mean rank values for LeSS and Internal organizations are displayed on the last six columns.

	U	Z	p	p*	Median LeSS	Median Internal	Avg. LeSS	Avg. Internal	Mean rank LeSS	Mean rank Internal	Effect size
The framework met the expectations of my organization	150.500	-0.554	0.580	0.610	4.00	4.00	3.82	3.83	21.13	19.04	
I would like to move back to the old ways of working	152.000	-0.652	0.532	0.652	1.00	1.00	1.32	1.33	19.93	21.83	
I would like to shift to another framework	150.500	-0.554	0.580	0.610	1.00	1.50	1.61	1.92	19.88	21.96	
I would like to recommend this framework to other similar organizations Exact Sig. [2*(1-tailed Sig.)]	80.000	-2.721	0.007	0.008	4.50	3.00	4.21	3.08	23.64	13.17	0.430

4.2.6.3 Spotify
Comparing Spotify with internal frameworks showed that Spotify organizations agree significantly more with wide adoption as a framework selection criteria than organizations using internally created ones, using the exact sampling distribution of U (Dinneen and Blakesley [62]) (see Table 17 for significant results, and Table 76 and 77 for full results). The effect size equals a large one (r >0.5).

Table 17 Significant results of the Mann-Whitney U test on adoption reasons are displayed on the first four columns. Median, average, mean rank values for Spotify and Internal organizations are displayed on the last six columns.

	U	Z	p	p*	Median Spotify	Median Internal	Avg. Spotify	Avg. Internal	Mean rank Spotify	Mean rank Internal	Effect size
Because it is widely adopted Exact Sig. [2*(1-tailed Sig.)]	25.000	-2.395	0.017	0.021	4.00	1.50	3.10	1.83	15.00	8.58	0.510

5 Discussion

5.1 Key Findings

Little is known in the literature about adoption reasons and satisfaction concerning scaling agile frameworks. In the following we discuss and interpret our findings based on the conducted survey.

Our results are in line with Agile One's survey [32] regarding the popularity of the SAFe framework and respondent's role demographics. We were surprised about the gap concerning *Scrum of Scrums* (SOS) framework. Only 4.5% (n = 9) of our survey respondents used SoS, while 16% used it according to Agile One. Also, 34% or their survey participants were Scrum Masters or Internal Coaches, 15% Development Leadership, and 11% Project / Program Manager. Our survey data consisted of 58% (n = 109) Process Perspective, 25% (n = 47) Management, 10% (n = 18) Development on team-level, and 7% (n = 13) Development on product-level. Concerning the scaling agile framework experience demographics our sample data is comparable with Agile One's data. In Agile One's survey, 34% of survey respondents had 3-5 years of experience, 27% had more than five years of experience, and 23% had 1-2 years of experience. Our survey data consists of 41% (n = 82) with 1-2 years, 33% (n = 64) with 3-5 years, 15% (n = 30) with less than one year, and 11% (n = 22) with more than five years of experience. Similar, geographic distribution of teams and development sites are comparable to the statistics provided in Agile One's survey [32] where 78% of respondents used distributed teams and 68% collaborated across different sites. Our survey data consisted of 77% (n = 153) with partially distributed teams and 5.5% (n = 11) that are completely distributed. 76.9% (n = 153) have completely distributed development sites (see Section 4.1).

The most commonly perceived adoption reason is scaling to more people. Agile development has become an inevitable step of structural development of organizations and connected with that - scaling more teams [63]. This is inline with Paasivaara et al. [25] arguing that *"Large organizations often have big projects executed by large and distributed development organizations, requiring agile methods to*

be scaled.". *Because it is widely adopted* received the most disagreement concerning its relevance as an adoption reason. Hence, we could not support the expectation that the primary reason to adopt a specific scaling agile framework is based either on sole popularity or consultant recommendation [26], [50].

Related to that, Spotify model and *"Internal"* frameworks represent the majority of frameworks, where popularity is a secondary adoption reason. This indicates that those frameworks come along with a framework assessment based on factors other than popularity. Regarding the internally developed frameworks we have no information about the underlying framework practices. It would be interesting to compare the satisfaction of SAFe and LeSS users against Spotify and Internal framework users based on a large sample (e.g., n for each framework larger 30).

According to Uludağ et al. [35], the SAFe framework has the largest amount of case studies and documentation. Following on that, SAFe accounts for 77.5% (n = 62) of the responses that agree with *widely adopted* being an important decision making characteristic (see Table 31). We explain this with two arguments. First, SAFe provides a large amount of documentations and case studies, which also serve as marketing material. Decision makers that had little exposure to agile development in the past rely on these materials to learn and make decisions. Second, the number of SAFe-specialized consultants and coaches is very large. According to Laanti and Kettunen [22], SAFe is now supported by over 300.000 certified practitioners in 110 countries. 70% of all Fortune 100 companies use SAFe or employ certified professionals [43]. Similar to Theobald et al. [26] and Diebold et al. [50], we believe that these points lead decision makers to follow the majority of peer organizations and the recommendation by consultants without conducting an assessment driven by organizational needs. Paasivaara reports a case study where SAFe was the only framework under consideration [30]. The fact that SAFe also poses the biggest share of frameworks which did not fulfilled the expectations supports our suspicion that SAFe adoptions are mainly build upon following popular choice.

There is no significant preference of plan-driven organizations for documentation and support as adoption reasons compared to agile organizations. As a result, we could not confirm H1.1 and H1.2. This opposes our initial assumption, because plan-driven organizations need to rely on external or third party knowledge to adopt agile methodologies as they lack the necessary experience. Furthermore, formerly plan-driven organization agreed significantly more with wide adoption as an important consideration for framework choice than agile organizations. This and the fact that documentation and support (i.e., coaching, training, and guidance) were not significantly affected by the previous development model supports the view suggested by Diebold et al. [50] concerning the missing individual framework assessment.

Geographical distribution of development teams and sites had no significant influence on documentation, support, scaling people, and solving architectural as adoption reasons. Thus, we could not confirm the claims made by existing scientific literature. When scaling agile practices in large organizations, distribution and size of the development organizations are one of the first aspects to consider [64]. Having distributed teams or several development sites in itself is difficult to deal with in the context of scaling agile frameworks, as explained in the case study provided by Paasivaara et al. [25]. These type of organizations have more dependencies, which slow down organizational change and increase the need for clear and extensive documentation [29].

The results of inferential analysis showed that the size of a development organization had no significant influence on the perception of documentation, support, scaling people, and solving architectural challenges as relevant adoption reasons. Thus, we could not confirm our initial hypotheses H1.3 - H1.5.

The test results comparing less experienced and more experienced organizations (related to their respective scaling agile framework experience) confirmed our initial hypothesis H2.1. Connected to more significant satisfaction, the tendency of organizations with more experience towards recommending the framework to similar organizations makes sense. The significant disagreement of more experienced organizations against shifting framework can be explained by reflecting that these organizations used the respective framework for more than five years. This is a significant investment and by this time abandoning the framework will be difficult. One the one side, investments made are relevant. On the other side, significant organizational and cultural change that start in short intervals decreases employee satisfaction. Nevertheless, it would be interesting to further investigate cases in which companies decided to switch to another framework after having used it for more than five years.

In line with our expectations, the number of areas covered by scaling agile framework had significant influence on adoption satisfaction. We explain this by referring to the fact that agile development teams exist within a broader corporate environment where collaboration across business functions is required. Agile teams need to communicate within, between teams, and within a broader organization [5]. Thus, including different areas into the scaling agile framework creates alignment and reduces frictions that can arise due to unaligned processes, practices, or lack of understanding. Hence, we could confirm H2.2.

It is surprising that there is no significant difference between the framework satisfaction of distributed and non-distributed teams. We expected non-distributed teams to be more satisfied, since adopting scaling agile framework across physical boundaries is arguably more complicated than at one single site. This also refers to the missing significance when comparing organizations with and without distributed development sites. As a result, we could not confirm H2.3 and H2.4

Hypotheses H3.1 - H3.3 could be confirmed, whereas H3.4 could not be entirely confirmed, since the comparison between LeSS and Spotify was insignificant. We anticipated the preference in favor of SAFe and LeSS because both are supported by many publications and training offerings [35], [15]. The lack of significance concerning documentation and support in either direction between SAFe and LeSS suggests that the distribution of responses was very similar.

The absence of any significant preference for the Spotify model because of the documentation is in line with existing research. Salameh and Bass [65] identified a gap in research regarding the Spotify model. In addition, this framework lacks of structured documentation. Available information on the underlying values and practices is based on two YouTube videos and one paper published in 2012 [44]. There is no official documentation provided by Spotify. A literature research also uncovered a lack of formal training offerings [10].

All effect sizes in the conducted inferential analysis ranged from small ($d = 0.2$) to large ($d = 0.5$). The smallest effect size was found for testing the relevance of popularity for adopting a scaling agile framework of agile organizations against plan-driven organizations. A small effect implicates that the previous development model had little effect on the overall perception of popularity as an adoption reason.

The highest effect size (0.557) was found when testing organizations that use LeSS against such that use Spotify on their willingness to recommend the framework. A large effect size means that both groups had a big effect on the respective expectation category. Overall the effect sizes of the significant Mann-Whitney U tests were small to large.

5.2 Threats to Validity

In the following, we present possible threats to validity. Construct, internal, and external validity are important and must be addressed here [52]. First, to address **internal validity** we based our survey on extensive research of relevant literature in the field of large-scale agile. We also invited experts to review the survey and refined it multiple times to ensure that survey participants can understand it. To avoid misunderstanding we also provided explanations and examples for each question. Second, **external validity** is addressed by designing a sampling plan and recruiting survey participants from multiple venues. Our participants belong have different roles, domains and organizational sizes and locations. This also increases reliability and replicability. Third, as **construct validity** refers to *"the issue of whether a survey is measuring what it intends to measure."* [52]. To address this, we based our questionnaire on scientific literature and we discussed our conceptualization of adoption reasons and adoption satisfaction with experienced scaling agile experts, who helped us to refine the survey.

The fact that the majority of respondents agreed or strongly agreed with all provided adoption reasons can be an indicator for acquiescence bias, meaning that respondents have a tendency to agree with every response option provided to them. Nevertheless, by providing the option *"I don't know"* to reduced this risk.

As mentioned before the survey data consisted of little to no responses from South America, Africa and the Asia-Pacific area. This is an indicator for limited generalizability in relation to location.

6 Conclusion and Future Work

Changing business environments and the success of agile methods urge big organizations to apply them on a large scale using scaling agile frameworks [1], [2], [3]. Although there is growing attention towards these frameworks, scientific literature providing in-depth analysis is still scarce [7], [10], [16], [22], [23], [24]. We aim to fill this gap by providing an empirical study investigating the adoption reasons and satisfaction of scaling agile frameworks based on comparisons between different organizational characteristics and scaling agile frameworks.

We conducted a global survey to investigate the reasons to choose specific scaling agile frameworks and to assess the satisfaction with the framework adoption. Most survey participants agreed that scaling agile development to more people is a relevant reason to choose a framework. Against the assumption of existing research papers, choosing scaling agile framework based on popularity received the most negative answers. We formed different sample groups, among others, according to development organization size, geographical distribution of teams, and development sites. Against our initial assumption, these groups had no significant

influence on preferences for certain framework characteristics (i.e., documentation, and support). Organizations that are more experienced with using scaling agile framework are significantly more satisfied than organizations with less experience. Same applies to the willingness to recommend the framework to other organizations and the aversion of shifting to another framework. Organizations that included more areas into the scaling agile framework were significantly more satisfied with the framework adoption than organizations that included less areas. Organizations that use SAFe and LeSS chose these frameworks because they care about two things more significantly compared to organizations that use Spotify or internally created frameworks. They prioritize well definition and clear documentation, and available coaching, training and guidance in their framework selection process.

Existing research lacks in-depth investigations of the selection process concerning scaling agile frameworks. Case studies across different organizations are required to analyze the underlying selection reasons. Many existing case studies barely mention the selection process and fail to deliver more detailed information. Same phenomenon applies to the satisfaction of organizations with their chosen framework. There is no dedicated scientific research analyzing the satisfaction with the scaling agile framework adoption. Our findings emphasize that there are still many open questions concerning scaling agile frameworks. Hence, we encourage researchers to conduct more detailed research to investigate adoption reasons and satisfaction. As mentioned before, this research paper is part of a series. We plan to publish additional papers by presenting the results of the residual data collected in our survey. Among others we will present the findings of analyzing expected and realized benefits, and challenges of scaling agile adoptions.

References

1. P. Kettunen, "Adopting key lessons from agile manufacturing to agile software product developmenta comparative study," *Technovation*, vol. 29, no. 6, pp. 408 – 422, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0166497208001302
2. P. Kettunen and M. Laanti, "Future software organizations–agile goals and roles," *European Journal of Futures Research*, vol. 5, no. 1, p. 16, 2017.
3. B. Sherehiy, W. Karwowski, and J. K. Layer, "A review of enterprise agility: Concepts, frameworks, and attributes," *International Journal of Industrial Ergonomics*, vol. 37, no. 5, pp. 445 – 460, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169814107000236
4. K. Petersen and C. Wohlin, "The effect of moving from a plan-driven to an incremental software development approach with agile practices," *Empirical Software Engineering*, vol. 15, no. 6, pp. 654–693, Dec 2010. [Online]. Available: https://doi.org/10.1007/s10664-010-9136-6
5. T. Dingsøyr and N. B. Moe, "Towards principles of large-scale agile development," in *Agile Methods. Large-Scale Development, Refactoring, Testing, and Estimation*, T. Dingsøyr, N. B. Moe, R. Tonelli, S. Counsell, C. Gencel, and K. Petersen, Eds. Cham: Springer International Publishing, 2014, pp. 1–8.
6. P. Kettunen, "Extending software project agility with new product development enterprise agility," *Software Process: Improvement and Practice*, vol. 12, no. 6, pp. 541–548, 2007. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/spip.342
7. T. Dingsoeyr, D. Falessi, and K. Power, "Agile development at scale: The next frontier," *IEEE Software*, vol. 36, no. 2, pp. 30–38, March 2019.
8. B. Boehm and R. Turner, "Management challenges to implementing agile processes in traditional development organizations," *IEEE Software*, vol. 22, no. 5, pp. 30–39, Sep. 2005.

9. Ö. Uludağ, M. Kleehaus, N. Dreymann, C. Kabelin, and F. Matthes, "Investigating the adoption and application of large-scale scrum at a german automobile manufacturer," in *Proceedings of the 14th International Conference on Global Software Engineering*, ser. ICGSE '19. Piscataway, NJ, USA: IEEE Press, 2019, pp. 22–29. [Online]. Available: https://doi.org/10.1109/ICGSE.2019.00-11
10. M. Alqudah and R. Razali, "A review of scaling agile methods in large software development," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 6, no. 6, pp. 828–837, 2016.
11. R. Nyman, I. Aro, and R. Wagner, "Automated acceptance testing of high capacity network gateway," in *Agile Processes in Software Engineering and Extreme Programming*, A. Sillitti, A. Martin, X. Wang, and E. Whitworth, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 307–314.
12. M. Paasivaara, C. Lassenius, V. T. Heikkil, K. Dikert, and C. Engblom, "Integrating global sites into the lean and agile transformation at ericsson," in *2013 IEEE 8th International Conference on Global Software Engineering*, Aug 2013, pp. 134–143.
13. M. Paasivaara, B. Behm, C. Lassenius, and M. Hallikainen, "Towards rapid releases in large-scale xaas development at ericsson: A case study," in *2014 IEEE 9th International Conference on Global Software Engineering*, Aug 2014, pp. 16–25.
14. A. Atlas, "Accidental adoption: The story of scrum at amazon.com," in *2009 Agile Conference*, Aug 2009, pp. 135–140.
15. 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, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0164121216300826
16. K. Conboy and N. Carroll, "Implementing large-scale agile frameworks: Challenges and recommendations," *IEEE Software*, vol. 36, no. 2, pp. 44–50, March 2019.
17. Ö. Uludağ, M. Kleehaus, C. Caprano, and F. Matthes, "Identifying and structuring challenges in large-scale agile development based on a structured literature review," in *2018 IEEE 22nd International Enterprise Distributed Object Computing Conference (EDOC)*, Oct 2018, pp. 191–197.
18. S. W. Ambler and M. Lines, *Disciplined agile delivery: A practitioner's guide to agile software delivery in the enterprise*. IBM press, 2012.
19. The Less Company B. V., "Large scale scrum," 2019 (accessed September 19, 2019). [Online]. Available: https://less.works/case-studies/index.html
20. Scaled Agile Inc., "Scaled agile framework," 2019 (accessed September 19, 2019). [Online]. Available: http://www.scaledagileframework.com
21. H. Kniberg and A. Ivarsson, "Scaling agile @ spotify with tribes, squads, chapters & guilds," 2012 (accessed October 19, 2019). [Online]. Available: https://blog.crisp.se/wp-content/uploads/2012/11/SpotifyScaling.pdf
22. M. Laanti and P. Kettunen, "Safe adoptions in finland: A survey research," in *Agile Processes in Software Engineering and Extreme Programming – Workshops*, R. Hoda, Ed. Cham: Springer International Publishing, 2019, pp. 81–87.
23. O. Turetken, I. Stojanov, and J. J. M. Trienekens, "Assessing the adoption level of scaled agile development: a maturity model for scaled agile framework," *Journal of Software: Evolution and Process*, vol. 29, no. 6, p. e1796, 2017, e1796 JSME-15-0085.R2. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/smr.1796
24. A. Putta, M. Paasivaara, and C. Lassenius, "Benefits and challenges of adopting the scaled agile framework (safe): Preliminary results from a multivocal literature review," in *Product-Focused Software Process Improvement*, M. Kuhrmann, K. Schneider, D. Pfahl, S. Amasaki, M. Ciolkowski, R. Hebig, P. Tell, J. Klünder, and S. Küpper, Eds. Cham: Springer International Publishing, 2018, pp. 334–351.
25. M. Paasivaara, B. Behm, C. Lassenius, and M. Hallikainen, "Large-scale agile transformation at ericsson: a case study," *Empirical Software Engineering*, vol. 23, no. 5, pp. 2550–2596, Oct 2018. [Online]. Available: https://doi.org/10.1007/s10664-017-9555-8
26. S. Theobald, A. Schmitt, and P. Diebold, "Comparing scaling agile frameworks based on underlying practices," in *Agile Processes in Software Engineering and Extreme Programming – Workshops*, R. Hoda, Ed. Cham: Springer International Publishing, 2019, pp. 88–96.
27. P. Kettunen, M. Laanti, F. Fagerholm, T. Mikkonen, and T. Männistö, "Finnish enterprise agile transformations: A survey study," in *Agile Processes in Software Engineering and Extreme Programming – Workshops*, R. Hoda, Ed. Cham: Springer International Publishing, 2019, pp. 97–104.

28. G. Papadopoulos, "Moving from traditional to agile software development methodologies also on large, distributed projects." *Procedia - Social and Behavioral Sciences*, vol. 175, pp. 455 – 463, 2015, proceedings of the 3rd International Conference on Strategic Innovative Marketing (IC-SIM 2014). [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877042815012835
29. M. Lindvall, D. Muthig, A. Dagnino, C. Wallin, M. Stupperich, D. Kiefer, J. May, and T. Kahkonen, "Agile software development in large organizations," *Computer*, vol. 37, no. 12, pp. 26–34, Dec 2004.
30. M. Paasivaara, "Adopting safe to scale agile in a globally distributed organization," in *2017 IEEE 12th International Conference on Global Software Engineering (ICGSE)*, May 2017, pp. 36–40.
31. E. Hossain, M. A. Babar, and H. Paik, "Using scrum in global software development: A systematic literature review," in *2009 Fourth IEEE International Conference on Global Software Engineering*, July 2009, pp. 175–184.
32. Agile Version One, "13th Annual State of Agile Survey," 2019 (accessed September 19, 2019). [Online]. Available: https://www.stateofagile.com/#ufh-i-521251909-13th-annual-state-of-agile-report/473508
33. C. Ebert and M. Paasivaara, "Scaling agile," *IEEE Software*, vol. 34, no. 6, pp. 98–103, November 2017.
34. I. Gat, "How bmc is scaling agile development," in *AGILE 2006 (AGILE'06)*, July 2006, pp. 314–320.
35. Ö. Uludağ, M. Kleehaus, X. Xu, and F. Matthes, "Investigating the role of architects in scaling agile frameworks," in *2017 IEEE 21st International Enterprise Distributed Object Computing Conference (EDOC)*, Oct 2017, pp. 123–132.
36. The LeSS Company B.V., "Certified less practitioner - principles to practices," 2019 (accessed October 17, 2019). [Online]. Available: https://less.works/courses/less-practitioner.html
37. C. Larman, *Scaling lean & agile development: thinking and organizational tools for large-scale Scrum*. Pearson Education India, 2008.
38. M. Paasivaara and C. Lassenius, "Scaling scrum in a large globally distributed organization: A case study," in *2016 IEEE 11th International Conference on Global Software Engineering (ICGSE)*, Aug 2016, pp. 74–83.
39. Emendare GmbH Co. KG, "Why LeSS Framework?" 2018 (accessed September 25, 2019). [Online]. Available: https://www.emendare.de/less-large-scale-scrum/
40. Scaled Agile, Inc., "Safe for lean enterprises - full safe," 2019 (accessed September 25, 2019). [Online]. Available: https://www.scaledagileframework.com/
41. D. Leffingwell, *SAFe 4.5 Reference Guide: Scaled Agile Framework for Lean Enterprises*. Addison-Wesley Professional, 2018.
42. Scaled Agile, Inc., "About us," 2019 (accessed October 18, 2019). [Online]. Available: https://www.scaledagile.com/about/about-us/
43. Scaled Agile Inc., "Welcome to scaled agile framework 4.6!" 2019 (accessed October 18, 2019). [Online]. Available: https://www.scaledagileframework.com/about/
44. H. Kniberg, "Scaling agile at spotify with tribes, squads, chapters and guilds," 2012 (accessed October 17, 2019). [Online]. Available: https://blog.crisp.se/2012/11/14/henrikkniberg/scaling-agile-at-spotify
45. H. Kniberg, "Spotify engineering culture (part 1)," 2014 (accessed October 19, 2019). [Online]. Available: https://labs.spotify.com/2014/03/27/spotify-engineering-culture-part-1/
46. ——, "Spotify engineering culture (part 2)," 2014 (accessed October 19, 2019). [Online]. Available: https://labs.spotify.com/2014/09/20/spotify-engineering-culture-part-2/
47. M. Kalenda, P. Hyna, and B. Rossi, "Scaling agile in large organizations: Practices, challenges, and success factors," *Journal of Software: Evolution and Process*, vol. 30, no. 10, 2018, e1954 smr.1954. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/smr.1954
48. Agile Scaling, "ASK: Agile Scaling Knowledge The Matrix," n.d. (accessed September 25, 2019). [Online]. Available: http://www.agilescaling.org/ask-matrix.html
49. ——, "Agile scaling - history & credits," n.d. (accessed September 30, 2019). [Online]. Available: http://www.agilescaling.org/history-credits.html
50. P. Diebold, A. Schmitt, and S. Theobald, "Scaling agile: How to select the most appropriate framework," in *Proceedings of the 19th International Conference on Agile Software Development: Companion*, ser. XP '18. New York, NY, USA: ACM, 2018, pp. 7:1–7:4. [Online]. Available: http://doi.acm.org/10.1145/3234152.3234177

51. J. Pries-Heje and M. M. Krohn, "The safe way to the agile organization," in *Proceedings of the XP2017 Scientific Workshops*, ser. XP '17. New York, NY, USA: ACM, 2017, pp. 18:1–18:3. [Online]. Available: http://doi.acm.org/10.1145/3120459.3120478
52. M. Kasunic, "Designing an effective survey," Carnegie-Mellon Univ Pittsburgh PA Software Engineering Inst, Tech. Rep., 2005.
53. J. Linåker, S. M. Sulaman, R. Maiani de Mello, and M. Höst, "Guidelines for conducting surveys in software engineering," [Publisher information missing], Tech. Rep., 2015.
54. Agile Alliance, "XP 2019," 2019 (accessed September 13, 2019). [Online]. Available: https://agilealliance.org/XP2019
55. ——, "Agile 2019," 2019 (accessed September 13, 2019). [Online]. Available: https://agilealliance.org/agile2019
56. ICGSE 2019, "ICGSE 2019," 2019 (accessed September 13, 2019). [Online]. Available: https://conf.researchr.org/track/icgse-2019/icgse-2019?track=ICGSE%20%20Research%20Papers
57. LinkedIn Corporation, "Lean and Agile Development Group," 2019 (accessed September 13, 2019). [Online]. Available: https://www.linkedin.com/groups/37631
58. A. Hart, "Mann-whitney test is not just a test of medians: differences in spread can be important," *BMJ*, vol. 323, no. 7309, pp. 391–393, 2001. [Online]. Available: https://www.bmj.com/content/323/7309/391
59. C. O. Fritz, P. E. Morris, and J. J. Richler, "Effect size estimates: current use, calculations, and interpretation." *Journal of experimental psychology: General*, vol. 141, no. 1, pp. 2–18, 2012.
60. K. Pearson, "Note on regression and inheritance in the case of two parents," *Proceedings of the Royal Society of London*, vol. 58, pp. 240–242, 1895. [Online]. Available: http://www.jstor.org/stable/115794
61. J. Cohen, "Statistical power analysis," *Current Directions in Psychological Science*, vol. 1, no. 3, pp. 98–101, 1992. [Online]. Available: https://doi.org/10.1111/1467-8721.ep10768783
62. L. C. Dinneen and B. C. Blakesley, "Algorithm as 62: A generator for the sampling distribution of the mann- whitney u statistic," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 22, no. 2, pp. 269–273, 1973. [Online]. Available: http://www.jstor.org/stable/2346934
63. U. Eklund, H. Holmström Olsson, and N. J. Strøm, "Industrial challenges of scaling agile in mass-produced embedded systems," in *Agile Methods. Large-Scale Development, Refactoring, Testing, and Estimation*, T. Dingsøyr, N. B. Moe, R. Tonelli, S. Counsell, C. Gencel, and K. Petersen, Eds. Cham: Springer International Publishing, 2014, pp. 30–42.
64. J. Eckstein, *Agile Software Development in the Large: Diving Into the Deep*. New York, NY, USA: Dorset House Publishing Co., Inc., 2004.
65. A. Salameh and J. M. Bass, "Spotify tailoring for promoting effectiveness in cross-functional autonomous squads," in *Agile Processes in Software Engineering and Extreme Programming – Workshops*, R. Hoda, Ed. Cham: Springer International Publishing, 2019, pp. 20–28.

Table 18 Complete descriptive - country distribution

		Frequency	Percent	Valid Percent	Cumulative Percent
	Missing	44	22.1	22.1	22.1
	Argentina	1	0.5	0.5	22.6
	Australia	2	1.0	1.0	23.6
	Austria	1	0.5	0.5	24.1
	Bosnia and Herzegovina	1	0.5	0.5	24.6
	Brazil	2	1.0	1.0	25.6
	Canada	3	1.5	1.5	27.1
	Czech Republic	1	0.5	0.5	27.6
	Denmark	11	5.5	5.5	33.2
	Finland	8	4.0	4.0	37.2
	France	4	2.0	2.0	39.2
	Germany	40	20.1	20.1	59.3
	Greece	1	0.5	0.5	59.8
Valid	India	9	4.5	4.5	64.3
	Iran	1	0.5	0.5	64.8
	Ireland	1	0.5	0.5	65.3
	Japan	4	2.0	2.0	67.3
	Netherlands	6	3.0	3.0	70.4
	New Zealand	3	1.5	1.5	71.9
	Russia	1	0.5	0.5	72.4
	Singapore	1	0.5	0.5	72.9
	South Africa	1	0.5	0.5	73.4
	Spain	1	0.5	0.5	73.9
	Sweden	4	2.0	2.0	75.9
	Turkey	2	1.0	1.0	76.9
	United Kingdom	2	1.0	1.0	77.9
	United States	44	22.1	22.1	100.0
	Total	199	100.0	100.0	

Table 19 Complete descriptive - industry distribution

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Aerospace / Aviation	5	2.5	2.5	2.5
	Automotive	19	9.5	9.5	12.1
	Construction / Home Improvement	1	0.5	0.5	12.6
	Consulting	14	7.0	7.0	19.6
	Education / University	3	1.5	1.5	21.1
	Engineering / Architecture	5	2.5	2.5	23.6
	Entertainment / Recreation	3	1.5	1.5	25.1
	Finance / Banking / Insurance / Accounting / Property	55	27.6	27.6	52.8
	Government / Public Sector	20	10.1	10.1	62.8
	Healthcare / Medical	11	5.5	5.5	68.3
	Logistics / Shipping / Transportation	8	4.0	4.0	72.4
	Manufacturing / Production	7	3.5	3.5	75.9
	Non-Profit	1	0.5	0.5	76.4
	Retail / Wholesale	7	3.5	3.5	79.9
	Technology	25	12.6	12.6	92.5
	Telecommunication	10	5.0	5.0	97.5
	Utilities	3	1.5	1.5	99.0
	Media	1	0.5	0.5	99.5
	Other	1	0.5	0.5	100.0
	Total	199	100.0	100.0	

Table 20 Complete descriptive - role distribution

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Management	50	25.1	25.4	25.4
	Development Team-Level	21	10.6	10.7	36.0
	Development Product-View	13	6.5	6.6	42.6
	Process Perspective	113	56.8	57.4	100.0
	Total	197	99.0	100.0	
Missing	System	2	1.0		
Total		199	100.0		

Table 21 Complete descriptive - framework distribution

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	SAFe	96	48.2	52.2	52.2
	LeSS	28	14.1	15.2	67.4
	Spotify	10	5.0	5.4	72.8
	Nexus	9	4.5	4.9	77.7
	SoS	9	4.5	4.9	82.6
	DAD	4	2.0	2.2	84.8
	S@S	9	4.5	4.9	89.7
	CAF	2	1.0	1.1	90.8
	FAST	2	1.0	1.1	91.8
	Gill	1	0.5	0.5	92.4
	Internal	12	6.0	6.5	98.9
	Other	2	1.0	1.1	100.0
	Total	184	92.5	100.0	
Missing	System	15	7.5		
Total		199	100.0		

Table 22 Complete descriptive - previous development model distribution

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Agile	50	25.1	25.1	25.1
	Iterative	9	4,5	4,5	29,6
	Hybrid	26	13,1	13,1	42,7
	Plan-Driven	95	47,7	47,7	90,5
	No Process Model	10	5,0	5,0	95,5
	Unclear	9	4,5	4,5	100,0
	Total	199	100,0	100,0	

Table 23 Complete descriptive - framework experience distribution

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	<1 year	30	15.1	15.1	15.1
	1-2 years	82	41.2	41.2	56.3
	3-5 years	65	32.7	32.7	88.9
	>5 years	22	11.1	11.1	100.0
	Total	199	100.0	100.0	

Table 24 Complete descriptive - team size distribution

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1-9 Teams	70	35.2	35.2	35.2
	10-50 Teams	72	36.2	36.2	71.4
	>50 Teams	57	28.6	28.6	100.0
	Total	199	100.0	100.0	

Table 25 Complete descriptive - geographical team distribution

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No distribution	35	17.6	17.6	17.6
	Partial distribution	153	76.9	76.9	94.5
	Complete distribution	11	5.5	5.5	100.0
	Total	199	100.0	100.0	

Table 26 Complete descriptive - development site distribution

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	151	75.9	75.9	75.9
	No	48	24.1	24.1	100.0
	Total	199	100.0	100.0	

Table 27 Complete descriptive - organization areas

Organization areas	
Accounting / Finance	27 13.6%
Administrative / Management	27 13.6%
Business	58 29.1%
Customer Service / Support	32 16.1%
Distribution / Logistics	11 5.5%
Entire Enterprise	29 14.6%
Human Resources	26 13.1%
IT	147 73.9%
Marketing and Sales	30 15.1%
Operations	54 27.1%
Product development & R&D	115 57.8%
Production	34 17.1%

Table 28 Complete descriptive - adoption reasons distribution

	I don't know	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Because it is widely adopted	11	20	28	54	67	19
	5.5%	10.1%	14.1%	27.1%	33.7%	9.5%
Because the framework is well defined and clearly documented	5	9	15	33	80	57
	2.5%	4.5%	7.5%	16.6%	40.2%	28.6%
Because the framework addresses architectural challenges	9	10	30	48	73	29
	4.5%	5.0%	15.1%	24.1%	36.7%	14.6%
Because it is well supported by coaching, training, and guidance	8	7	19	33	82	50
	4.0%	3.5%	9.5%	16.6%	41.2%	25.1%
To remain competitive in the market	9	4	16	34	81	55
	4.5%	2.0%	8.0%	17.1%	40.7%	27.6%
Scale to more people	10	4	12	21	83	69
	5.0%	2.0%	6.0%	10.6%	41.7%	34.7%

Table 29 Complete descriptive - framework distribution with applied filtering for "strongly agree" and "agree" for the adoption reason "Scale to more people"

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	SAFe	76	50.0	53.5	53.5
	LeSS	19	12.5	13.4	66.9
	Spotify	8	5.3	5.6	72.5
	Nexus	6	3.9	4.2	76.8
	SoS	7	4.6	4.9	81.7
	DAD	3	2.0	2.1	83.8
	S@S	8	5.3	5.6	89.4
	CAF	2	1.3	1.4	90.8
	FAST	2	1.3	1.4	92.3
	Internal	9	5.9	6.3	98.6
	Other	2	1.3	1.4	100.0
	Total	142	93.4	100.0	
Missing	System	10	6.6		
Total		152	100.0		

Table 30 Complete descriptive - framework distribution with applied filtering for "strongly disagree" and "disagree" for the adoption reason "Widely adopted"

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	SAFe	9	18,8	19,1	19,1
	LeSS	12	25,0	25,5	44,7
	Spotify	3	6,3	6,4	51,1
	Nexus	4	8,3	8,5	59,6
	SoS	4	8,3	8,5	68,1
	DAD	1	2,1	2,1	70,2
	S@S	4	8,3	8,5	78,7
	CAF	1	2,1	2,1	80,9
	FAST	2	4,2	4,3	85,1
	Internal	6	12,5	12,8	97,9
	Other	1	2,1	2,1	100,0
	Total	47	97,9	100,0	
Missing	System	1	2,1		
Total		48	100,0		

Table 31 Complete descriptive - framework distribution with applied filtering for "strongly agree" and "agree" for the adoption reason "Widely adopted"

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	SAFe	62	72.1	77.5	77.5
	LeSS	4	4.7	5.0	82.5
	Spotify	6	7.0	7.5	90.0
	Nexus	1	1.2	1.3	91.3
	SoS	3	3.5	3.8	95.0
	DAD	1	1.2	1.3	96.3
	S@S	3	3.5	3.8	100.0
	Total	80	93.0	100.0	
Missing	System	6	7.0		
Total		86	100.0		

Table 32 Complete descriptive - framework distribution with applied filtering for previous development model "plan-driven"

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	SAFe	55	57,9	59,8	59,8
	LeSS	16	16,8	17,4	77,2
	Spotify	2	2,1	2,2	79,3
	Nexus	1	1,1	1,1	80,4
	SoS	4	4,2	4,3	84,8
	DAD	2	2,1	2,2	87,0
	S@S	3	3,2	3,3	90,2
	CAF	1	1,1	1,1	91,3
	Gill	1	1,1	1,1	92,4
	Internal	6	6,3	6,5	98,9
	Other	1	1,1	1,1	100,0
	Total	92	96,8	100,0	
Missing	System	3	3,2		
Total		95	100,0		

Table 33 Complete descriptive - framework distribution with applied filtering for previous development model "agile"

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	SAFe	14	28,0	31,8	31,8
	LeSS	5	10,0	11,4	43,2
	Spotify	6	12,0	13,6	56,8
	Nexus	5	10,0	11,4	68,2
	SoS	2	4,0	4,5	72,7
	DAD	1	2,0	2,3	75,0
	S@S	3	6,0	6,8	81,8
	CAF	1	2,0	2,3	84,1
	FAST	2	4,0	4,5	88,6
	Internal	4	8,0	9,1	97,7
	Other	1	2,0	2,3	100,0
	Total	44	88,0	100,0	
Missing	System	6	12,0		
Total		50	100,0		

Table 34 Complete descriptive - previous development framework distribution with applied filtering for primary framework "SAFe"

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Agile	14	14.6	14.6	14.6
	Iterative	4	4.2	4.2	18.8
	Hybrid	16	16.7	16.7	35.4
	Plan-Driven	55	57.3	57.3	92.7
	No Process Model	2	2.1	2.1	94.8
	Unclear	5	5.2	5.2	100.0
	Total	96	100.0	100.0	

Table 35 Complete descriptive - duration distribution with applied filtering for primary framework "SAFe"

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	<1 year	13	13.5	13.5	13.5
	1-2 years	39	40.6	40.6	54.2
	3-5 years	37	38.5	38.5	92.7
	>5 years	7	7.3	7.3	100.0
	Total	96	100.0	100.0	

Table 36 Complete descriptive - team size distribution with applied filtering for primary framework "SAFe"

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1-9 Teams	25	26.0	26.0	26.0
	10-50 Teams	37	38.5	38.5	64.6
	>50 Teams	34	35.4	35.4	100.0
	Total	96	100.0	100.0	

Table 37 Complete descriptive - geographical team distribution with applied filtering for primary framework "SAFe"

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No distribution	14	14.6	14.6	14.6
	Partial distribution	77	80.2	80.2	94.8
	Complete distribution	5	5.2	5.2	100.0
	Total	96	100.0	100.0	

Table 38 Complete descriptive - geographic distributed sites distribution with applied filtering for primary framework "SAFe"

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	74	77.1	77.1	77.1
	No	22	22.9	22.9	100.0
	Total	96	100.0	100.0	

Table 39 Complete descriptive - adoption reason (first six rows) and expectation (last four rows) distribution with applied filtering for primary framework "SAFe"

	I don't know	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Because it is	4	2	7	21	47	15
widely adopted	4.2%	2.1%	7.3%	21.9%	49.0%	15.6%
Because the framework is	3	0	2	9	44	38
well defined and clearly documented	3.1%	0.0%	2.1%	9.4%	45.8%	39.6%
Because the framework	6	3	12	23	39	13
addresses architectural challenges	6.3%	3.1%	12.5%	24.0%	40.6%	13.5%
Because it is well supported	3	0	6	10	44	33
by coaching, training, and guidance	3.1%	0.0%	6.3%	10.4%	45.8%	34.4%
To remain competitive	6	0	7	14	46	23
in the market	6.3%	0.0%	7.3%	14.6%	47.9%	24.0%
Scale to more people	5	1	6	8	42	34
	5.2%	1.0%	6.3%	8.3%	43.8%	35.4%
The framework met the	1	4	7	13	51	20
expectations of my organization	1.0%	4.2%	7.3%	13.5%	53.1%	20.8%
I would like to move back	1	69	16	8	2	0
to the old ways of working	1.0%	71.9%	16.7%	8.3%	2.1%	0.0%
I would like to shift to	2	22	37	24	9	2
another framework	2.1%	22.9%	38.5%	25.0%	9.4%	2.1%
I would like to recommend this	1	3	5	20	34	33
framework to other similar organizations	1.0%	3.1%	5.2%	20.8%	35.4%	34.4%

Table 40 Complete descriptive - previous development model distribution with applied filtering for primary framework LeSS

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Agile	5	17.9	17.9	17.9
	Iterative	2	7.1	7.1	25.0
	Hybrid	1	3.6	3.6	28.6
	Plan-Driven	16	57.1	57.1	85.7
	No Process Model	3	10.7	10.7	96.4
	Unclear	1	3.6	3.6	100.0
	Total	28	100.0	100.0	

Table 41 Complete descriptive - team size distribution with applied filtering for primary framework LeSS

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1-9 Teams	12	42.9	42.9	42.9
	10-50 Teams	8	28.6	28.6	71.4
	>50 Teams	8	28.6	28.6	100.0
	Total	28	100.0	100.0	

Table 42 Complete descriptive - duration distribution with applied filtering for primary framework LeSS

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	<1 year	2	7.1	7.1	7.1
	1-2 years	16	57.1	57.1	64.3
	3-5 years	7	25.0	25.0	89.3
	>5 years	3	10.7	10.7	100.0
	Total	28	100.0	100.0	

Table 43 Complete descriptive - geographical team distribution with applied filtering for primary framework LeSS

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No distribution	7	25.0	25.0	25.0
	Partial distribution	21	75.0	75.0	100.0
	Total	28	100.0	100.0	

Table 44 Complete descriptive - geographic distributed sites distribution with applied filtering for primary framework LeSS

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	21	75.0	75.0	75.0
	No	7	25.0	25.0	100.0
	Total	28	100.0	100.0	

Table 45 Complete descriptive - adoption reason (first six rows) and expectation (last four rows) distribution with applied filtering for primary framework "LeSS"

	I don't know	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Because it is	4	5	7	8	2	2
widely adopted	14.3%	17.9%	25.0%	28.6%	7.1%	7.1%
Because the framework	0	0	4	6	10	8
is well defined and clearly documented	0.0%	0.0%	14.3%	21.4%	35.7%	28.6%
Because the framework	0	1	7	10	4	6
addresses architectural challenges	0.0%	3.6%	25.0%	35.7%	14.3%	21.4%
Because it is well supported	0	0	4	8	9	7
by coaching, training, and guidance	0.0%	0.0%	14.3%	28.6%	32.1%	25.0%
To remain competitive	0	1	0	9	8	10
in the market	0.0%	3.6%	0.0%	32.1%	28.6%	35.7%
Scale to more people	2	1	2	4	8	11
	7.1%	3.6%	7.1%	14.3%	28.6%	39.3%
The framework met the	2	0	1	4	12	9
expectations of my organization	7.1%	0.0%	3.6%	14.3%	42.9%	32.1%
I would like to move back	0	22	4	1	1	0
to the old ways of working	0.0%	78.6%	14.3%	3.6%	3.6%	0.0%
I would like to shift to	1	14	9	3	1	0
another framework	3.6%	50.0%	32.1%	10.7%	3.6%	0.0%
I would like to recommend this	0	0	1	6	7	14
framework to other similar organizations	0.0%	0.0%	3.6%	21.4%	25.0%	50.0%

Table 46 Complete descriptive - framework expectations

	I don't know	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
The framework met the	5	4	12	28	103	47
expectations of my organization	2.5%	2.0%	6.0%	14.1%	51.8%	23.6%
I would like to move back	4	136	41	13	5	0
to the old ways of working	2.0%	68.3%	20.6%	6.5%	2.5%	0.0%
I would like to shift	7	62	62	48	16	4
to another framework	3.5%	31.2%	31.2%	24.1%	8.0%	2.0%
I would like to recommend this	4	6	10	43	71	65
framework to other similar organizations	2.0%	3.0%	5.0%	21.6%	35.7%	32.7%

Table 47 Complete descriptive - framework distribution with applied filtering for expectation met "Neither agree nor disagree"

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	SAFe	13	46.4	54.2	54.2
	LeSS	4	14.3	16.7	70.8
	Spotify	1	3.6	4.2	75.0
	SoS	1	3.6	4.2	79.2
	S@S	1	3.6	4.2	83.3
	CAF	1	3.6	4.2	87.5
	FAST	1	3.6	4.2	91.7
	Internal	2	7.1	8.3	100.0
	Total	24	85.7	100.0	
Missing	System	4	14.3		
Total		28	100.0		

Table 48 Complete descriptive - duration distribution with applied filtering for expectation met "Neither agree nor disagree"

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	<1 year	9	32.1	32.1	32.1
	1-2 years	12	42.9	42.9	75.0
	3-5 years	4	14.3	14.3	89.3
	>5 years	3	10.7	10.7	100.0
	Total	28	100.0	100.0	

Table 49 Complete descriptive - framework distribution with applied filtering for shifting frameworks "Strongly agree" or "Agree"

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	SAFe	11	55.0	55.0	55.0
	LeSS	1	5.0	5.0	60.0
	Spotify	3	15.0	15.0	75.0
	SoS	3	15.0	15.0	90.0
	Internal	2	10.0	10.0	100.0
	Total	20	100.0	100.0	

Table 50 Complete descriptive - duration distribution with applied filtering for shifting frameworks "Strongly agree" and "Agree" and primary framework "SAFe"

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	<1 year	3	27.3	27.3	27.3
	1-2 years	4	36.4	36.4	63.6
	3-5 years	4	36.4	36.4	100.0
	Total	11	100.0	100.0	

Table 51 Complete descriptive - previous development model distribution with applied filtering for moving back "Strongly agree" or "Agree"

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Agile	2	40.0	40.0	40.0
	Iterative	1	20.0	20.0	60.0
	Plan-Driven	2	40.0	40.0	100.0
	Total	5	100.0	100.0	

Table 52 Complete descriptive - primary framework distribution with applied filtering for expectations met "Strongly disagree" or "Disagree"

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	SAFe	11	68.8	73.3	73.3
	LeSS	1	6.3	6.7	80.0
	SoS	2	12.5	13.3	93.3
	Internal	1	6.3	6.7	100.0
	Total	15	93.8	100.0	
Missing	System	1	6.3		
Total		16	100.0		

Table 53 Complete descriptive - primary framework with applied filtering for recommending "Strongly disagree" or "Disagree"

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	SAFe	8	50.0	50.0	50.0
	LeSS	1	6.3	6.3	56.3
	Spotify	3	18.8	18.8	75.0
	SoS	2	12.5	12.5	87.5
	Internal	2	12.5	12.5	100.0
	Total	16	100.0	100.0	

Table 54 Results of the Mann-Whitney U test on adoption reasons are displayed on the first four columns. Median, average, mean rank values for agile (A) and plan-driven (P) organizations are displayed on the last six columns.

	U	Z	p	Median A	Median P	Avg. A	Avg. P	Mean rank A	Mean rank P	Effect size
Because it is widely adopted	1874.000	-2.266	0.023	3.00	3.00	2.56	3.19	62.44	78.56	0.188
Because the framework is well defined and clearly documented	2145.500	-1.002	0.316	4.00	4.00	3.48	3.81	68.41	75.42	
Because the framework addresses architectural challenges	2071.000	-1.310	0.190	4.00	3.00	3.40	3.17	79.08	69.80	
Because it is well supported by coaching, training, and guidance	1943.000	-1.882	0.060	4.00	4.00	3.30	3.77	64.36	77.55	
To remain competitive in the market	2245.000	-0.565	0.572	4.00	4.00	3.48	3.73	70.40	74.37	
Scale to more people	2299.500	-0.332	0.740	4.00	4.00	3.70	3.82	71.49	73.79	

Table 55 Results of the Mann-Whitney U test on expectations are displayed on the first four columns. Median, average, mean rank values for agile (A) and plan-driven (P) organizations are displayed on the last six columns.

	U	Z	p	Median A	Median P	Avg. A	Avg. P	Mean rank A	Mean rank P	Effect size
The framework met the expectations of my organization	2116.500	-1.179	0.238	4.00	4.00	3.92	3.69	78.17	70.28	
I would like to move back to the old ways of working	2105.000	-1.336	0.182	1.00	1.00	1.52	1.35	78.40	70.16	
I would like to shift to another framework	2145.000	-0.993	0.321	2.00	2.00	2.00	2.18	68.40	75.42	
I would like to recommend this framework to other similar organizations	2252.500	-0.533	0.594	4.00	4.00	3.60	3.80	70.55	74.29	

Table 56 Results of the Mann-Whitney U test on adoption reasons are displayed on the first four columns. Median, average, mean rank values for small (S) and large (L) product development organizations are displayed on the last six columns.

	U	Z	p	Median S	Median L	Avg. S	Avg. L	Mean rank S	Mean rank L	Effect size
Because it is widely adopted	1620.000	-1.878	0.060	3.00	4.00	2.79	3.19	58.64	70.58	
Because the framework is well defined and clearly documented	1899.500	-0.484	0.628	4.00	4.00	3.70	3.72	62.64	65.68	
Because the framework addresses architectural challenges	1965.000	-0.150	0.881	4.00	3.00	3.29	3.23	64.43	63.47	
Because it is well supported by coaching, training, and guidance	1921.000	-0.374	0.708	4.00	4.00	3.54	3.51	65.06	62.70	
To remain competitive in the market	1746.000	-1.285	0.199	4.00	4.00	3.64	3.91	60.44	68.73	
Scale to more people	1887.500	-0.559	0.576	4.00	4.00	3.77	3.93	62.46	65.89	

Table 57 Results of the Mann-Whitney U test on expectations are displayed on the first four columns. Median, average, mean rank values for small (S) and large (L) product development organizations are displayed on the last six columns.

	U	Z	p	Median S	Median L	Avg. S	Avg. L	Mean rank S	Mean rank L	Effect size
The framework met the expectations of my organization	1869.000	-0.663	0.507	4.00	4.00	3.79	3.88	62.20	66.21	
I would like to move back to the old ways of working	1911.000	-0.498	0.618	1.00	1.00	1.33	1.46	62.80	65.47	
I would like to shift to another framework	1716.500	-1.405	0.160	2.00	2.00	1.89	2.19	60.02	68.89	
I would like to recommend this framework to other similar organizations	1886.000	-0.556	0.579	4.00	4.00	3.94	3.72	65.56	62.09	

Table 58 Results of the Mann-Whitney U test on adoption reasons are displayed on the first four columns. Median, average, mean rank values for less experienced (L-E) and more experienced (M-E) organizations are displayed on the last six columns.

.	U	Z	p	Median L-E	Median M-E	Avg. L-E	Avg. M-E	Mean rank L-E	Mean rank M-E	Effect size
Because it is widely adopted	313.000	-0.322	0.747	3.00	3.00	2.93	2.77	27.07	25.73	
Because the framework is well defined and clearly documented	276.000	-1.030	0.303	4.00	4.00	3.27	3.59	24.70	28.95	
Because the framework addresses architectural challenges	185.000	-2.764	0.006	3.00	4.00	2.77	3.73	21.67	33.09	0.383
Because it is well supported by coaching, training, and guidance	215.500	-2.208	0.027	4.00	4.00	2.93	3.68	22.68	31.70	0.306
To remain competitive in the market	241.500	-1.703	0.089	4.00	4.00	3.40	3.82	23.55	30.52	
Scale to more people	284.000	-0.891	0.373	4.00	4.00	3.43	3.95	24.97	28.59	

Table 59 Results of the Mann-Whitney U test on expectations are displayed on the first four columns. Median, average, mean rank values for less experienced (L-E) and more experienced (M-E) organizations are displayed on the last six columns.

	U	Z	p	Median L-E	Median M-E	Avg. L-E	Avg. M-E	Mean rank L-E	Mean rank M-E	Effect size
The framework met the expectations of my organization	186.500	-2.787	0.005	4.00	4.00	3.30	4.14	21.72	33.02	0.386
I would like to move back to the old ways of working	254.500	-1.592	0.111	1.00	1.00	1.60	1.27	29.02	23.07	
I would like to shift to another framework	212.000	-2.262	0.024	3.00	1.50	2.53	1.68	30.43	21.14	0.313
I would like to recommend this framework to other similar organizations	166.000	-3.142	0.002	3.00	4.00	2.97	4.05	21.03	33.95	0.435

Table 60 Results of the Mann-Whitney U test on adoption reasons are displayed on the first five columns. Median, average, mean rank values for not distributed (N-D) and fully distributed (F-D) organizations are displayed on the last six columns.

	U	Z	p	p*	Median N-D	Median F-D	Avg. N-D	Avg. F-D	Mean rank N-D	Mean rank F-D	Effect size
Because it is widely adopted	140.500	-1.400	0.161	0.183	3.00	3.00	2.87	3.27	22.01	28.23	
Because the framework is well defined and clearly documented	122.000	-1.973	0.048	0.071	4.00	4.00	4.03	3.73	25.51	17.09	0.290
Because the framework addresses architectural challenges	164.000	-0.774	0.439	0.476	4.00	3.00	3.90	3.18	24.31	20.91	
Because it is well supported by coaching, training, and guidance	162.500	-0.814	0.416	0.445	4.00	3.00	3.60	3.27	24.36	20.77	
To remain competitive in the market	163.500	-0.781	0.435	0.461	4.00	4.00	3.90	3.55	24.33	20.86	
Scale to more people	169.500	-0.632	0.528	0.559	4.00	4.00	4.30	3.82	24.16	21.41	
Exact Sig. [2*(1-tailed Sig.)]											

Table 61 Results of the Mann-Whitney U test on expectations are displayed on the first five columns. Median, average, mean rank values for not distributed (N-D) and fully distributed (F-D) organizations are displayed on the last six columns.

	U	Z	p	p*	Median N-D	Median F-D	Avg. N-D	Avg. F-D	Mean rank N-D	Mean rank F-D	Effect size
The framework met the expectations of my organization	149.000	-1.218	0.223	0.272	4.00	4.00	3.83	3.45	24.74	19.55	
I would like to move back to the old ways of working	80.000	-3.361	0.001	0.003	1.00	2.00	1.23	2.27	20.29	33.73	0.495
I would like to shift to another framework	129.500	-1.689	0.091	0.105	2.00	3.00	1.86	2.55	21.70	29.23	
I would like to recommend this framework to other similar organizations	130.500	-1.679	0.093	0.111	4.00	4.00	4.00	3.27	25.27	17.86	
Exact Sig. [2*(1-tailed Sig.)]											

Table 62 Results of the Mann-Whitney U test on adoption reasons are displayed on the first four columns. Median, average, mean rank values for distributed sites (DS) and non distributed sites (NDS) are displayed on the last six columns.

	U	Z	p	Median DS	Median NDS	Avg. DS	Avg. NDS	Mean rank DS	Mean rank NDS	Effect size
Because it is widely adopted	3389.000	-0.698	0.485	3.00	3.00	3.03	2.98	101.56	95.10	
Because the framework is well defined and clearly documented	3168.500	-1.377	0.169	4.00	4.00	3.66	3.96	96.98	109.49	
Because the framework addresses architectural challenges	3568.000	-0.167	0.867	4.00	4.00	3.25	3.33	99.63	101.17	
Because it is well supported by coaching, training, and guidance	3118.500	-1.526	0.127	4.00	4.00	3.55	3.88	96.65	110.53	
To remain competitive in the market	3358.000	-0.804	0.421	4.00	4.00	3.66	3.85	98.24	105.54	
Scale to more people	3443.500	-0.522	0.581	4.00	4.00	3.89	3.77	101.20	96.24	

Table 63 Results of the Mann-Whitney U test on expectations are displayed on the first four columns. Median, average, mean rank values for distributed sites (DS) and non distributed sites (NDS) are displayed on the last six columns.

	U	Z	p	Median DS	Median NDS	Avg. DS	Avg. NDS	Mean rank DS	Mean rank NDS	Effect size
The framework met the expectations of my organization	3251.500	-1.166	0.244	4.00	4.00	3.85	3.69	102.47	92.24	
I would like to move back to the old ways of working	3542.000	-0.288	0.773	1.00	1.00	1.38	1.42	99.46	101.71	
I would like to shift to another framework	3162.000	-1.382	0.167	2.00	2.00	2.13	1.92	103.06	90.38	
I would like to recommend this framework to other similar organizations	3541.500	-0.249	0.803	4.00	4.00	3.85	3.79	99.45	101.72	

Table 64 Results of the Mann-Whitney U test on adoption reasons are displayed on the first four columns. Median, average, mean rank values for narrow (N) and broad area (B) organizations are displayed on the last six columns.

.	U	Z	p	Median N	Median B	Avg. N	Avg. B	Mean rank N	Mean rank B	Effect size
Because it is widely adopted	2308.500	-1.667	0.093	3.00	3.00	3.08	2.71	103.01	85.40	
Because the framework is well defined and clearly documented	2730.500	-0.265	0.798	4.00	4.00	3.75	3.68	99.55	102.19	
Because the framework addresses architectural challenges	2265.500	-1.830	0.067	3.00	4.00	3.21	3.59	96.73	115.87	
Because it is well supported by coaching, training, and guidance	2658.500	-0.503	0.615	4.00	4.00	3.62	3.65	99.11	104.31	
To remain competitive in the market	1784.500	-3.507	0.000	4.00	4.50	3.58	4.29	93.82	130.01	0.248
Scale to more people	2795.500	-0.035	0.972	4.00	4.00	3.87	3.82	100.06	99.71	

Table 65 Results of the Mann-Whitney U test on expectations are displayed on the first four columns. Median, average, mean rank values for narrow (N) and broad area (B) organizations are displayed on the last six columns.

	U	Z	p	Median N	Median B	Avg. N	Avg. B	Mean rank N	Mean rank B	Effect size
The framework met the expectations of my organization	2133.500	-2.389	0.017	4.00	4.00	3.76	4.06	95.93	119.75	0.169
I would like to move back to the old ways of working	2491.000	-1.253	0.210	1.00	1.00	1.41	1.29	101.90	90.76	
I would like to shift to another framework	2187.500	-2.100	0.036	2.00	1.00	2.14	1.79	103.74	81.84	0.148
I would like to recommend this framework to other similar organizations	2302.500	-1.723	0.085	4.00	4.00	3.79	4.09	96.65	114.78	

Table 66 Results of the Mann-Whitney U test on adoption reasons are displayed on the first four columns. Median, average, mean rank values for SAFe and LeSS organizations are displayed on the last six columns.

	U	Z	p	Median SAFe	Median LeSS	Avg. SAFe	Avg. LeSS	Mean rank SAFe	Mean rank LeSS	Effect size
Because it is widely adopted	579.500	-4.760	0.000	4.00	2.00	3.56	2.18	70.46	35.20	0.427
Because the framework is well defined and clearly documented	1049.000	-1.896	0.058	4.00	4.00	4.14	3.79	65.75	51.96	
Because the framework addresses architectural challenges	1228.500	-0.715	0.475	4.00	3.00	3.30	3.25	63.70	58.38	
Because it is well supported by coaching, training, and guidance	1061.500	-1.795	0.073	4.00	4.00	3.99	3.68	65.44	52.41	
To remain competitive in the market	1257.000	-0.551	0.582	4.00	4.00	3.70	3.93	61.59	65.61	
Scale to more people	1289.500	-0.346	0.729	4.00	4.00	3.91	3.71	63.07	60.55	

Table 67 Results of the Mann-Whitney U test on expectations are displayed on the first four columns. Median, average, mean rank values for SAFe and LeSS organizations are displayed on the last six columns.

	U	Z	p	Median SAFe	Median LeSS	Avg. SAFe	Avg. LeSS	Mean rank SAFe	Mean rank LeSS	Effect size
The framework met the expectations of my organization	1224.500	-0.733	0.439	4.00	4.00	3.76	3.82	61.26	66.77	
I would like to move back to the old ways of working	1278.000	-0.509	0.611	1.00	1.00	1.39	1.32	63.19	60.14	
I would like to shift to another framework	874.000	-2.983	0.003	2.00	1.00	2.23	1.61	67.40	45.71	0.267
I would like to recommend this framework to other similar organizations	1132.500	-1.332	0.183	4.00	4.50	3.90	4.21	60.30	70.05	

Table 68 Results of the Mann-Whitney U test on adoption reasons are displayed on the first four columns. Median, average, mean rank values for SAFe and Spotify organizations are displayed on the last six columns.

	U	Z	p	Median SAFe	Median Spotify	Avg. SAFe	Avg. Spotify	Mean rank SAFe	Mean rank Spotify	Effect size
Because it is widely adopted	388.000	-1.071	0.284	4.00	4.00	3.56	3.10	54.46	44.30	
Because the framework is well defined and clearly documented	112.000	-4.262	0.000	4.00	2.00	4.14	2.30	57.33	16.70	0.413
Because the framework addresses architectural challenges	421.000	-0.666	0.505	4.00	3.50	3.30	3.00	54.11	47.60	
Because it is well supported by coaching, training, and guidance	207.000	-3.154	0.002	4.00	3.00	3.99	2.70	56.34	26.20	0.306
To remain competitive in the market	445.000	-0.403	0.687	4.00	4.00	3.70	3.40	53.86	50.00	
Scale to more people	387.500	-1.075	0.282	4.00	4.00	3.91	3.70	54.46	44.25	

Table 69 Results of the Mann-Whitney U test on expectations are displayed on the first four columns. Median, average, mean rank values for SAFe and Spotify organizations are displayed on the last six columns.

	U	Z	p	Median SAFe	Median Spotify	Avg. SAFe	Avg. Spotify	Mean rank SAFe	Mean rank Spotify	Effect size
The framework met the expectations of my organization	458.500	-0.256	0.798	4.00	4.00	3.76	3.60	53.72	51.35	
I would like to move back to the old ways of working	441.500	-0.541	0.607	1.00	1.00	1.39	1.40	53.90	49.65	
I would like to shift to another framework	470.500	-0.107	0.915	2.00	2.00	2.23	2.50	53.40	54.45	
I would like to recommend this framework to other similar organizations	194.500	-3.222	0.001	4.00	3.00	3.90	2.60	56.47	24.95	0.312

Table 70 Results of the Mann-Whitney U test on adoption reasons are displayed on the first four columns. Median, average, mean rank values for SAFe and Internal organizations are displayed on the last six columns.

	U	Z	p	Median SAFe	Median Internal	Avg. SAFe	Avg. Internal	Mean rank SAFe	Mean rank Internal	Effect size
Because it is widely adopted	154.000	-4.348	0.000	4.00	1.50	3.56	1.83	58.90	19.33	0.418
Because the framework is well defined and clearly documented	101.000	-4.945	0.000	4.00	2.50	4.14	2.08	59.45	14.92	0.475
Because the framework addresses architectural challenges	512.500	-0.649	0.516	4.00	3.50	3.30	3.00	55.16	49.21	
Because it is well supported by coaching, training, and guidance	206.500	-3.846	0.000	4.00	2.00	3.99	2.33	58.35	23.71	0.370
To remain competitive in the market	471.500	-1.084	0.278	4.00	3.50	3.70	3.17	55.59	45.79	
Scale to more people	555.500	-0.214	0.830	4.00	4.00	3.91	3.83	54.71	52.79	

Table 71 Results of the Mann-Whitney U test on expectations are displayed on the first four columns. Median, average, mean rank values for SAFe and Internal organizations are displayed on the last six columns.

	U	Z	p	Median SAFe	Median Internal	Avg. SAFe	Avg. Internal	Mean rank SAFe	Mean rank Internal	Effect size
The framework met the expectations of my organization	571.000	-0.054	0.957	4.00	4.00	3.76	3.83	54.55	54.08	
I would like to move back to the old ways of working	556.000	-0.246	0.806	1.00	1.00	1.39	1.33	54.29	56.17	
I would like to shift to another framework	476.000	-1.018	0.309	2.00	1.50	2.23	1.92	55.54	46.17	
I would like to recommend this framework to other similar organizations	353.000	-2.280	0.023	4.00	3.00	3.90	3.08	56.82	35.92	0.219

Table 72 Results of the Mann-Whitney U test on adoption reasons are displayed on the first four columns. Median, average, mean rank values for LeSS and Spotify organizations are displayed on the last six columns.

	U	Z	p	p*	Median LeSS	Median Spotify	Avg. LeSS	Avg. Spotify	Mean rank LeSS	Mean rank Spotify	Effect size
Because it is widely adopted	84.500	-1.877	0.061	0.065	2.00	4.00	2.18	3.10	17.52	25.05	
Because the framework is well defined and clearly documented	50.000	-3.071	0.002	0.002	4.00	2.00	3.79	2.30	22.71	10.50	0.498
Because the framework addresses architectural challenges	128.000	-0.409	0.683	0.708	3.00	3.50	3.25	3.00	19.93	18.30	
Because it is well supported by coaching, training, and guidance	84.000	-1.922	0.055	0.065	4.00	3.00	3.68	2.70	21.50	13.90	
To remain competitive in the market	118.500	-0.744	0.457	0.482	4.00	4.00	3.93	3.40	20.27	17.35	
Scale to more people	123.000	-0.592	0.554	0.590	4.00	4.00	3.71	3.70	20.11	17.80	
Exact Sig. [2*(1-tailed Sig.)]											

Table 73 Results of the Mann-Whitney U test on expectations are displayed on the first four columns. Median, average, mean rank values for LeSS and Spotify organizations are displayed on the last six columns.

	U	Z	p	p*	Median LeSS	Median Spotify	Avg. LeSS	Avg. Spotify	Mean rank LeSS	Mean rank Spotify	Effect size
The framework met the expectations of my organization	120.500	-0.699	0.484	0.524	4.00	4.00	3.82	3.60	20.20	17.55	
I would like to move back to the old ways of working	133.000	-0.290	0.772	0.832	1.00	1.00	1.32	1.40	19.75	18.80	
I would like to shift to another framework	100.000	-1.423	0.155	0.194	1.00	2.00	1.61	2.50	18.07	23.50	
I would like to recommend this framework to other similar organizations	41.000	-3.435	0.001	0.001	4.50	3.00	4.21	2.60	23.04	9.60	0.557
Exact Sig. [2*(1-tailed Sig.)]											

Table 74 Results of the Mann-Whitney U test on adoption reasons are displayed on the first four columns. Median, average, mean rank values for LeSS and Internal organizations are displayed on the last six columns.

	U	Z	p	p*	Median LeSS	Median Internal	Avg. LeSS	Avg. Internal	Mean rank LeSS	Mean rank Internal	Effect size
Because it is widely adopted	147.000	-0.639	0.523	0.550	2.00	1.50	2.18	1.83	21.25	18.75	
Because the framework is well defined and clearly documented	52.000	-3.156	0.000	0.000	4.00	2.50	3.79	2.08	24.64	10.83	0.499
Because the framework addresses architectural challenges	160.500	-0.228	0.820	0.827	3.00	3.50	3.25	3.00	20.77	19.88	
Because it is well supported by coaching, training, and guidance	80.000	-2.674	0.008	0.008	4.00	2.00	3.68	2.33	23.46	13.17	0.422
To remain competitive in the market	124.500	-1.335	0.182	0.202	4.00	3.50	3.93	3.17	22.05	16.88	
Scale to more people	166.000	-0.062	0.951	0.965	4.00	4.00	3.71	3.83	20.43	20.67	
Exact Sig. [2*(1-tailed Sig.)]											

Table 75 Results of the Mann-Whitney U test on expectations are displayed on the first four columns. Median, average, mean rank values for LeSS and Internal organizations are displayed on the last six columns.

	U	Z	p	p*	Median LeSS	Median Internal	Avg. LeSS	Avg. Internal	Mean rank LeSS	Mean rank Internal	Effect size
The framework met the expectations of my organization	150.500	-0.554	0.580	0.610	4.00	4.00	3.82	3.83	21.13	19.04	
I would like to move back to the old ways of working	152.000	-0.652	0.532	0.652	1.00	1.00	1.32	1.33	19.93	21.83	
I would like to shift to another framework	150.500	-0.554	0.580	0.610	1.00	1.50	1.61	1.92	19.88	21.96	
I would like to recommend this framework to other similar organizations	80.000	-2.721	0.007	0.008	4.50	3.00	4.21	3.08	23.64	13.17	0.430
Exact Sig. [2*(1-tailed Sig.)]											

Table 76 Results of the Mann-Whitney U test are displayed on the first four columns. Median, average, mean rank values for Spotify and Internal organizations are displayed on the last six columns.

	U	Z	p	p*	Median Spotify	Median Internal	Avg. Spotify	Avg. Internal	Mean rank Spotify	Mean rank Internal	Effect size
Because it is widely adopted	25.000	-2.395	0.017	0.021	4.00	1.50	3.10	1.83	15.00	8.58	0.510
Because the framework is well defined and clearly documented	54.500	-0.376	0.707	0.722	2.00	2.50	2.30	2.08	12.05	11.04	
Because the framework addresses architectural challenges	59.500	-0.034	0.973	0.974	3.50	3.50	3.00	3.00	11.45	11.54	
Because it is well supported by coaching, training, and guidance	52.000	-0.549	0.583	0.628	3.00	2.00	2.70	2.33	12.30	10.83	
To remain competitive in the market	55.000	-0.338	0.736	0.771	4.00	3.50	3.40	3.17	12.00	11.08	
Scale to more people	51.500	-0.617	0.537	0.582	4.00	4.00	3.70	3.83	10.65	12.21	
Exact Sig. [2*(1-tailed Sig.)]											

Table 77 Results of the Mann-Whitney U test are displayed on the first four columns. Median, average, mean rank values for Spotify and Internal organizations are displayed on the last six columns.

	U	Z	p	p*	Median Spotify	Median Internal	Avg. Spotify	Avg. Internal	Mean rank Spotify	Mean rank Internal	Effect size
The framework met the expectations of my organization	113.500	-0.115	0.908	0.923	4.00	4.00	3.60	3.83	11.35	11.63	
I would like to move back to the old ways of working	109.000	-0.447	0.655	0.722	1.00	1.00	1.40	1.33	10.90	12.00	
I would like to shift to another framework	126.000	-0.824	0.410	0.456	2.00	1.50	2.50	1.92	12.70	10.50	
I would like to recommend this framework to other similar organizations	101.000	-0.968	0.333	0.381	3.00	3.00	2.60	3.08	10.10	12.67	
Exact Sig. [2*(1-tailed Sig.)]											