



Benefits and Challenges of Adopting the Scaled Agile Framework (SAFe): Preliminary Results from a Multivocal Literature Review

Abheeshta Putta^{1(✉)}, Maria Paasivaara^{1,2}, and Casper Lassenius^{1,3}

¹ Department of Computer Science, Aalto University, Espoo, Finland
{[abheeshta.putta](mailto:abheeshta.putta@aalto.fi), [casper.lassenius](mailto:casper.lassenius@aalto.fi)}@aalto.fi

² IT University of Copenhagen, Copenhagen, Denmark
mpaa@itu.dk

³ Simula Research Laboratory, Oslo, Norway

Abstract. Over the past few years, the Scaled Agile Framework (SAFe) has been adopted by a large number of organizations to scale agile to large enterprises. At the moment, SAFe seems to be the most predominant agile scaling framework. Despite the current popularity of SAFe in the software intensive industry, there exists surprisingly little scientific research on the benefits and challenges of SAFe adoption. To collect the existing knowledge on this topic, we conducted a multivocal literature review, which includes both peer-reviewed and non-peer reviewed case studies and experience reports on organizations that have adopted SAFe. We identified 52 unique organisations adopting SAFe, five from the scientific literature and 47 from the grey literature.

The most salient benefit categories were: transparency, alignment, productivity, predictability and time to market. The most frequently mentioned challenge categories were: change resistance, challenges with the first program increment planning and moving away from agile.

Keywords: Agile software development

Large-scale agile software development · Scaled agile framework · SAFe

1 Introduction

Agile development methods have become highly popular in software organizations since the early 2000. The principles and practices of agile development were originally designed for small and co-located teams [5]. To leverage the potential benefits also in larger enterprises, the agile practices have to be scaled [8, 33]. Large-scale agile transformations has been a burning topic [8, 26], with increased concerns for additional coordination mechanisms and integration of non-development units, such as finance and human resource management [8]. To support scaling, new frameworks such as the Scaled Agile Framework (SAFe)

[25], Large Scale Scrum (LeSS) [24], and Disciplined Agile Delivery (DAD) [3] have been proposed by agile consultants. International workshops on large-scale agile organized at XP2016 [27] and XP2017 [26] have highlighted the importance and need for research into the adoption of scaling frameworks [26].

According to the 12th State of Agile Survey, which Version One conducts yearly, SAFe continues to be the most popular scaling framework in large enterprises [29]. Moreover, a recent survey on software development approaches indicated the predominance of SAFe over LeSS and DAD [23]. As the number of organizations adopting scaling frameworks is increasing [29], this provides opportunities for researchers and software practitioners to accumulate knowledge on the usage of these frameworks through case studies, technical reports and experience reports. To our knowledge, no secondary studies on the benefits and challenges of the scaling frameworks or their adoption process have been published. This is striking, given their importance in the industry. In this paper, we start filling this gap by summarizing the benefits and challenges of adopting the SAFe framework in the form of a multi-vocal literature review.

2 Related Work

2.1 Scaled Agile Framework (SAFe)

The Scaled Agile Framework was designed by Dean Leffingwell to scale agile to large enterprises [1, 36]. It incorporates practices from Scrum, Extreme Programming, Kanban and Lean. It offers four levels: the *Team*, *Program*, *Portfolio and Value stream* levels. The team level comprises the agile teams. Agile Release Trains (ART's) are introduced to scale a large number of teams and individuals at the program level. ART's follow HIP (Hardening, Innovation, Planning) iterations to develop Potential Shippable Increments (PSI) or Program Increments (PI). PI's are planned during the release planning days. SAFe introduces additional roles such as the Agile Release Train Engineer, system teams, release management team and portfolio management team. The core values of SAFe are: *build in quality, transparency, alignment and program execution* [16]. Figure 1 gives an overview of the SAFe framework.

2.2 Secondary Studies on Large-Scale Agile

Secondary studies on large-scale agile have explored topics such as challenges and success factors of large-scale agile transformations [8], organizational, managerial and cultural aspects [32], scalability and adoptability [22], inter-team coordination [14], architectural roles [35] and quality requirements practices [2]. However, systematic literature reviews on scaling frameworks have not been found in the literature. The only review found on scaling frameworks, compares a few scaling frameworks based on team size, practices and organization type [1]. Neither that review, nor other previous reviews on large-scale agile have included the grey literature, e.g. case studies or experience reports published on the homepages of

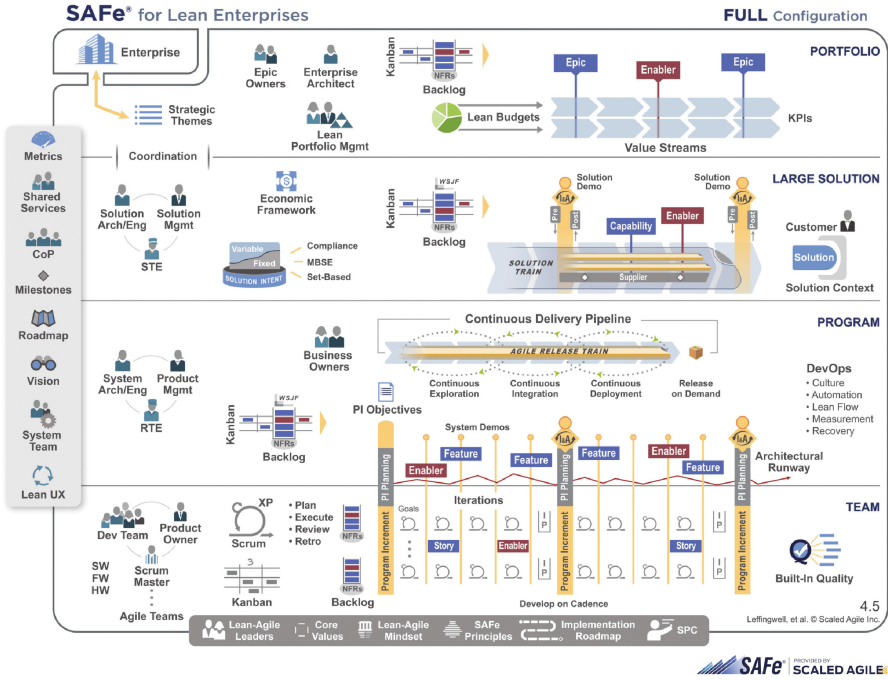


Fig. 1. Scaled agile framework, version 4.5 [19]

the frameworks. While there are inherent problems with case studies published by the proponents of a particular framework, completely eliminating such studies from literature reviews unnecessarily excludes the voice of the practitioners on the usage of scaling frameworks and implementation of agile at scale. Thus, in particular given the lack of scientific literature, we consider it important to study such cases, fully understanding the related problems, further discussed below.

3 Research Method

We conducted a multivocal literature review on the adoption of the SAFe framework to answer the following research questions:

- RQ1: What are the reported benefits of adopting SAFe?
- RQ2: What are the reported challenges of adopting SAFe?

3.1 Multivocal Literature Review

Systematic literature reviews and systematic mapping studies have been popular in the field of software engineering. They help to summarize the existing studies

reported in a specific research domain [11]. According to the widely adopted systematic literature review guidelines [21], a “*fully systematic literature review*” should include both the grey and the peer reviewed literature. Grey literature is defined as, “*(the literature), produced on all levels of government, academics, business and industry in print and electronic formats, but which is not controlled by commercial publishers, i.e., where publishing is not the primary activity of the producing body*” [11].

However, most SLRs published in the software engineering literature have not included the grey literature [12]. Including the grey literature is challenging, and the search strategy for grey literature has not been systematically addressed in the SLR guidelines [21]. This is unfortunate, as excluding this literature will eliminate the voice and opinions of the practitioners who do not publish in the academic forums [4, 12]. It has been evident that most of the software practitioners do not publish in academic fora [13].

The situation seems to be slowly changing for the better. Recent literature reviews including both peer reviewed literature, as well as grey literature from blogs, websites, and white papers, have popularized the term “*multi-vocal literature reviews*”, or MLRs [12]. Several such reviews have been conducted in software engineering to bridge the gap between the voice of practitioners and academics [4, 12, 28, 34].

The inclusion of grey literature can also be considered as a threat, as the information reported is based on the opinions and experiences of the practitioners rather than systematic data collection procedures and analysis [10]. Thus, there are severe issues with, e.g., author and publication bias that needs to be accounted for when analyzing such literature. However, several SLRs published in the SE literature have already included peer-reviewed experience reports, written by consultants and practitioners, and that rely on experiences rather than systematic data collection and analysis (e.g., [8, 9, 31]).

Table 1. Search strings

Database	Search strings	# matches
Scopus	(“safe” AND “scrum”) OR “Scaled agile framework”	33
Web of science	(“safe” AND “scrum”) OR “Scaled agile framework”	16
ACM	(+safe +(scrum)) OR “Scaled agile framework”	6
IEEEExplorer	(“safe” AND “scrum”) OR “Scaled agile framework”	8
Total		63

3.2 Study Selection

Databases: For identifying the peer-reviewed literature we searched scientific databases by formulating keywords. The number of matches and the search strings are given in Table 1.

The main source for the grey literature was the official SAFe website [19]. We included the case studies, including backlinks, published there. We used this source, as it currently is the most notable source of SAFe studies available. The case studies are based on a defined review and data collection procedure. Organizations initially answer a questionnaire [17] and thereafter, the “*scaled agile team*” reviews the answers and supplements provided by the organizations. The team contacts organizations for interviews with key members responsible during the SAFe implementation to gather background information [18]. Drafts are written with the help of case study specialists. These are reviewed and approved by the organizations before being published on the website. The aim is to publish reports of mature SAFe organizations, i.e., the reports should reflect the situation no earlier than after 18–24 months into the SAFe implementation [18].

The main benefits of this material are the standard format and questionnaire used giving some opportunity for cross-case analysis. However, the review process is likely built not only to guarantee the quality of the published case studies, but also to ensure that the SAFe framework is put in a good light. Therefore, the publication bias is extremely strong, and it can be questioned whether case studies providing negative results would make it through.

Inclusion Criteria: We used the following inclusion criteria:

1. Only articles related to the Scaled Agile Framework.
2. Only primary evidence: experience reports, case studies, action research.
3. Publication type: Conference papers, journal papers, workshop papers, white papers from the Scaled Agile Framework’s homepage.

Search Procedure: During the keyword search from scientific databases 63 matches from four databases were identified. After removing the duplicates, we had a total of 41 papers. These were filtered based on the titles and abstracts by two authors resulting in eight includes and 33 excludes. After full-text filtering, six scientific papers were selected for the analysis. Finally, we selected *five papers*, eliminating one paper, as the same case was published both as a conference and a journal paper: we included only the journal paper. We backward searched by snowballing through the references of selected five papers and also forward searched by snowballing the citations. In the forward search we found one primary study meeting the inclusion criteria. Thus, in total, we included six primary studies from peer-reviewed sources.

Grey Literature: For identifying the grey literature, we manually searched the SAFe homepage [19] and identified 48 white paper reports. In addition, we used backlinks and gathered additional supplements supporting the case studies, such as downloadable presentations and external links published within each white paper report (e.g.: John Deere [G11, G12, G13, G14, G80]¹). We gathered 46

¹ G denotes non peer-reviewed sources or grey literature sources.

case study reports published on the SAFe homepage, seven downloadable reports, sixteen presentations and thirteen links (*8 internal and 5 external*). In total, we included 82 reports from the grey literature.

Search Results: In total, we selected 88² documents: 82 gathered from the grey literature and six from the scientific databases. When the same organization was described in multiple documents, we treated it as one case only if the documents described an adoption in the same organizational unit. If the adoptions were separate, e.g., coming from different organizational units of the the same company (e.g., AVL GmbH: D³6 and D7) they were treated as different *cases* coming from the same *organization*. When the same paper described multiple cases (e.g., adoption in different units at different time frames), they were separated as different cases (e.g., Comptel [P3]⁴: D3 and D4). Altogether, we got 54 unique SAFe adoption *cases* from 52 *organizations* (see Table 2).

3.3 Analysis

The qualitative data from both peer-reviewed and non-peer reviewed sources was imported into the coding tool NVivo 11 [20]. We followed the coding guidelines presented in [6]. The analysis started with open coding. The open codes were constantly compared to each other based on similarities and differences observed between them. They were grouped into higher code categories, called axial codes. Both axial and open codes formed were thoroughly discussed by the three authors constantly during the coding process that was performed by the first author. We identified 23 codes⁵ for the benefits and 15 codes for the challenges of SAFe adoption. We clustered the benefits according the core values of the SAFe framework: *alignment, build in quality, transparency*, as they are the elementary beliefs that are claimed to be of primary importance for the effective SAFe implementation [16]. We clustered the challenges into *organizational and cultural, roles, practices, as well scaled and distributed*. Regarding each benefit and challenge, we mention the number of cases to express the predominance across organizations. However, we did not make any other quantitative analysis, like ranking the benefits and challenges according to the most important and least important, as even though very interesting, that was not possible with this qualitative data.

4 Results and Discussion

We identified only six peer-reviewed primary studies on SAFe. The focus areas of these studies were: assessing the maturity of SAFe adoption [P6], SAFe self-assessment [P5], the SAFe framework in testing [P2], a real-world example on

² Due to space limitations, all the primary sources can be found using this link: <https://figshare.com/s/6be7337493b080ed70b6>.

³ D represents a peer reviewed case.

⁴ P denotes peer-reviewed sources or scientific literature sources.

⁵ More detailed analysis of the code categories will be presented in the journal version.

key elements of SAFe [P1], the adoption of SAFe in a globally distributed organization [P3] and one partially focused on the challenges [P4]. Only three studies focused on the adoption and usage of SAFe [P4, P3, P5]. We identified 47 unique cases (82 documents) from the SAFe homepage. These reports focused on the adoption reasons, transformation steps, and benefits of SAFe. Neither the peer-reviewed nor the grey literature had an explicit focus on the adoption challenges. The grey literature provided deeper insights on the SAFe adoption and usage compared to the peer-reviewed literature.

A total of 54 unique cases from 52 organizations (Table 2), were identified. Out of 54 cases, seven⁶ were identified from the peer-reviewed literature⁷ and 47 cases⁸ were identified from the grey literature. Organizations from various domains have adopted SAFe such as financial (12 cases), software (9 cases), manufacturing (6 cases), and telecommunications (6 cases). The most prominent domain was the financial services. Moreover, SAFe has been popular in globally distributed organizations.

4.1 Benefits of Adopting SAFe

The reported benefits achieved by adopting SAFe are summarized in Table 3. The most common benefits identified are: *transparency* (22 cases), *alignment* (19 cases), *quality* (17 cases), *time to market* (17 cases), *predictability* (16 cases) and *productivity* (15 cases). The benefits marked by a star (*) in Table 3, are common to both the peer-reviewed and the non-peer reviewed studies.

The core values of SAFe are: build in quality, alignment, program execution and transparency [16]. A large proportion of the cases mentioned they had gained these benefits by adopting SAFe. We compared our findings to the 12th state of agile survey. 29% of respondents of that survey had adopted SAFe [29]. This survey reported similar benefits of agile in general as our study found regarding SAFe, visibility (66%), productivity (61%), alignment (65%), morale (61%), predictability (49%), quality (47%), and time to market (62%).

According to our results, practitioners seem to think that SAFe can help to bring several business benefits, such as improved time to market, and faster and more frequent deliveries. Surprisingly, none of the business benefits were reported in the peer-reviewed studies, but the majority of non-peer reviewed studies have attributed their business success to the SAFe framework. This difference could be due to the Scaled Agile Team insisting for business benefits, “*most importantly, we look for specific business results, which may include time-to-market, productivity, quality, and employee engagement*” [18]. Moreover, non-peer reviewed studies are more inclined towards presenting the benefits of the SAFe framework compared to peer-reviewed studies, e.g., only 8 (marked by *) out of 24 benefits were reported by peer-reviewed studies.

⁶ * marked in Table 2.

⁷ marked with D in Table 2.

⁸ marked with C in Table 2.

Table 2. Domain of the case organizations

Domain	Organizations and cases
Financial Services (12)	Standard Bank: C1 [G55], Capital One: C2 [G45], Northwestern Mutual: C3 [G54], Nordea: C4 [G77, G23], SEI: C5 [G5, G6, G7], Tradestation: C6 [G28], Westpac: C7 [G43, G68], Simcorp*: D1 [P4], Vantiv: C8 [G56], Fannie Mae: C9 [G48], Edge Verve: C10 [G57, G63], Seamless Payments: C11 [G65, G64],
Electronics (4)	Thales: C12 [G44], Intel: C13 [G20, G82], Fitbit: C14 [G49], TomTom: C15 [G42, G60]
Software (9)	Sony PlayStation: C16 [G26], Amdocs: C17 [G36], HP: C18 [G39], Anonymous: C19 [G31, G76], BMC Software: C20 [G16, G78], Mitchells: C21 [G21, G66], Censhare: C22 [G47], Accenture: C23 [G32, G1], X company*: D2 [P6]
Telecommunications (6)	Comptel*: D3 and D4 [P3], Amdocs: C17 [G36], Swisscom: C24 [G41], Telstra: C25 [G27, G3, G69], Big IT: C26 [G15, G70, G71, G72, G73, G74]
Retail and Distribution (3)	Kantar Retail: C27 [G50], Travis Perkins: C28 [G79, G29, G75], DiscountTire: C29 [G18]
Medical and Pharma (4)	NHS: C30 [G53], Philips: C31 [G40], Elekta: C32 [G33, G62], AstraZeneca: C33 [G37]
Media and Marketing (2)	Valpak: C34 [G30], Sproutland: C35 [G59]
Agriculture (1)	LIC: C36 [G52]
Manufacturing (6)	Cisco: C37 [G38], TomTom: C15 [G42, G60], SK Hynix: C38 [G25, G61], Lego: C39 [G51, G10], JohnDeere: C40 [G11, G12, G13, G14, G80], Ocuco*: D5 [P5]
COT's (1)	RMIT University: C41 [G4, G34, G9, G35]
Customer Care (1)	CSG: C42 [G17]
Outsourcing (1)	Infogain: C43 [G19]
Government IT (2)	PoleEmploi: C44 [G8, G24, G2], Australian Postal Services: C45 [G67, G46]
Maritime IT (1)	NAPA: C46 [G22, G81]
Automobile (2)	AVL Gmbh* (2): D6 [P1] and D7 [P2]
Aviation (1)	Air France KLM: C47 [G58]

According to our results, practitioners clearly think that SAFe has brought benefits, however, it is also important to look into how the organizations measured these benefits. Unfortunately, not much information is given related to this. Only one study in the peer-reviewed literature focused on SAFe metrics [P5]. Most grey literature cases attributed all the mentioned benefits to the

SAFe adoption. However, it would be interesting to learn, e.g., which practices of SAFe brought the benefits. Only few cases had done that. Moreover, most of the benefits mentioned were similar to the general benefits from implementing agile. In the future it would be interesting to study what the unique benefits provided by SAFe practices, such as Agile Release Trains, PI planning meetings and value streams are.

4.2 Challenges of Adopting SAFe

The reported challenges of adopting SAFe are summarized in Table 4. The most commonly mentioned challenges are: *resistance to change* (10 cases), *moving away from agile* (7 cases), *First PI planning* (7 cases), controversies with the framework (6 cases), *Agile Release Train challenges* (6 cases), *staffing roles* (5 cases), and *GSD challenges* (4 cases). Change resistance, GSD challenges, integration of the non-development units and test automation challenges found in our review, were also mentioned in the *Systematic Literature Review on Challenges and Success Factors for Large-scale Agile Transformations* [8]. Further, *change resistance*, could be supported by the results from the 12th state of agile survey, *general resistance* [29].

11 out of 15 challenges were common both for the peer-reviewed and non-peer reviewed studies (marked by * in Table 4). It is notable that the majority of the peer-reviewed studies reported challenges during the SAFe adoption, while very few non-peer reviewed studies mentioned challenges.

Even though SAFe is a framework for scaling agile to large enterprises, several organizations felt they were *moving away from agile*. This challenge is supported by the arguments of several “agilists”, for example, Ken Schwaber (co-creator of Scrum), says that “SAFe is based on RUP, rather than Scrum” [15], Ron Jeffries (co-founder of XP) sees issues in centralized approaches and planning in the framework [15] and Stephen Denning (board of directors of Scrum Alliance), finds SAFe to enforce the horizontal ideology of agile into vertical structures by saying [7] “they run the risk that the firm will emerge back in the unproductive vertical world of hierarchical bureaucracy” [7]. Pancholi and Grover [30] argue that SAFe “murders the spirit of agile development” and claim that SAFe is sold to large organizations that fear change, but would like to increase their productivity and reduce defects. According to them the framework portrays an “agile fairy illusion” [30].

Both organizations previously using traditional methods, as well as those having agile already in use, have shown resistance towards accepting SAFe. There is also a need to draw attention towards the specific challenges of SAFe, such as the challenges related to PI planning, value streams and agile release trains. Some faced controversies within the framework itself, like overhead, and story point normalization. Unlike the benefits, challenges have been mentioned only by 40% of the cases. Consequently, there is a need for more research into the challenges of the SAFe framework adoption and usage, as well as ways to overcome those challenges.

Table 3. Benefits of adopting SAFe

Benefit	Description	# cases
<i>Build in Quality</i>		
Quality	Improved product quality [G36, G37, G44, G56, G41, G28, G33, G39, G50, G18, G40, G17, G55, G58] ^a , higher quality releases [G78, G39], higher code quality [G5, G7, G59]	17
Defect Reduction	Drop in the defect rate [G18, G25, G61, G22, G21, G39, G27, G66, G43, G52, G48], reduction of the patches [G22], increase in the defect removal efficiency, reduction in quality assurance defects [G38] and escaped defects [G57, G63], warranty expenses down [G11], bug fixes reduced [G31, G76]	14
Continuous Improvement*	Focus on continuous improvement [G16, G39, G60, G67, G7, P4, P5] ^b	7
Waste elimination	Less duplicated work [G51], reduced rework [G51, G44], negligible waste [G42, G60], reduced waste [G79, G73]	5
<i>Alignment</i>		
Alignment*	Increased alignment teams [G32, G1, G51, G7, G23, G41, G56, G43, P1, G59], management and development teams [G47, G33, G23], solved problems of misaligned teams [G51], alignment of business [G7], customer expectations [G36], alignment between IT and business units [G30, G59], client and vendor teams [G19], alignment towards organizational goals [G56], processes and projects [G37, G2, G40], tools [G40], products [G45, G57] and priorities [G19]	19
Collaboration*	Enhanced collaboration [G9, G44, G80, P3, G6, G35], greater collaboration between team members [G57, G23], multiple teams [G23] international teams [G38, P5], diverse working groups [G44], different units (IT, Business) [G30, G19, G58, G6], cross site and cross functional collaboration [G33]	14
Dependencies*	Improved dependency management [G26, G47, G55, G52, P5], dependencies across trains are addressed [G24], less dependency problems [G10, G2]	7
Vision	Established shared vision [G50, G47, G38], shared goals [G10] broader view on company wide strategies [G49, G26]	6

(continued)

Table 3. (continued)

Benefit	Description	# cases
<i>Transparency</i>		
Transparency*	Enhanced transparency [G67, G25, G61, G22, G26, G44, G42, G31, G56, G5, G7, G35, G58, G48], process transparency [G33, G30, G50, G9, G23, G10, G56, G19], cross team dependencies are transparent [G47, P1], transparency in communication [G20]	22
Visibility*	Improved visibility [G73, G30, G49, G82, G30, G20, G50, P4, G9, G26, G35, G57, G59, G48]	11
<i>Organizational Benefits</i>		
Productivity	Improved productivity [G16, G17, G18, G67, G28, G55, G57, G58, G48], increase in productivity across teams and employees [G50, G30], increased delivery of number of products, variants and capabilities [G49, G56, G20, G26]	15
Team Autonomy*	More empowered teams [G67, G16, G47, G30, G10, G23, G5, G48, G59], self managing teams [G60, G52], self organizing teams [G30, G6, P5], improved morale [P4, G45, G30], ownership [G42, G58], control of own commitments [G60] and own code [G60]	13
Engagement	Improved employee engagement [G16, G45, G67, G49, G11, G13, G56, G43, G44, G52, G59], improved employee retention [G45, G21], decrease in attrition [G56, G50]	12
Employee Satisfaction*	Improved employee satisfaction [G67, G47, G38, P3, G58], happier teams [G15, G73, G27, G31, G76], happy employees [G31, G76]	8
Predictability	Greater predictability in the product delivery [G36, G25, G61, G82, G22, G26, G42, G50, G45, G55, G56, G7, G67, G58, G52, G59, G48]	16
Feedback	Faster feedback from customers [G36, G16, G50, G48] and greater feedback mechanisms [G56]	5

(continued)

Table 3. (continued)

Benefit	Description	# cases
<i>Business Benefits</i>		
Cost	Financial benefits [G37], reduced costs [G16, G55], controlled costs [G39], reduced cost per epic [G50, G57], reduced cost point size [G44], decreased infrastructural costs [G67], decrease in delivery costs [G15, G27], reduced quality costs [G32]	11
Frequent Deliveries	Increase in release frequency [G78, G44, G27, G34, G80, G39], frequent deliverable's [G45, G30, G36, G52, G48, G58] and value [G30], increased deployments [G55]	13
Faster Deliveries	Faster deliveries [G37, G45, G39, G44, G34], faster feature delivery [G78, G38], quick releases [G30, G80, G78], decrease in feature cycle time [G40, G54] and release cycle down [G40, G43, G38, G34]	12
On time delivery	On time delivery [G15, G50, G39, G49, G56, G48], no missing dates [G30] and schedule slips [G20]	8
Responsiveness	More responsive towards market needs [G50] and customer needs [G56] and decrease in time to respond to customer requests [G21, G50]	4
Time to market	Improved time to market [G36, G37, G45, G47, G18, G30, G31, G39, G11, G13, G50, G55, G41, G56, G58, G57, G52, G66]	17
Customer Satisfaction	Increase in the customer satisfaction [G36, G45, G31, G50, G41, G56, G6, G5, G67, G2, G58, G57] delighted [G36] and happy customers [G76]	13

^a**G** denotes non peer-reviewed sources or grey literature sources

^b**P** denotes peer-reviewed sources or scientific literature sources

5 Limitations

This section presents the threats to validity [37] and the steps that have been taken to mitigate those threats.

Selection Bias. This occurs during the selection of primary studies based on the interpretation of inclusion and exclusion criteria. We mitigated this by involving all authors in designing the criteria and two researchers filtered the abstracts and titles of peer-reviewed articles independently. Regarding the non-peer reviewed literature, we included all the case studies published on the homepage of the Scaled Agile Framework, which mitigated the threat of selection bias.

Table 4. Challenges of adopting SAFe

Challenge	Description	# cases
<i>Organizational and cultural</i>		
Change resistance*	Resistance [G39, G45, G58, G52, G59] towards accepting change, experiencing change as negative [P3], initial hesitation from teams [G56], individuals choose to leave [G54], teams reject to take part in ART [G60], initial resistance towards ART [G45] and reject the common ways of working, strong change resistance from teams towards lack of SAFe knowledge and need to change [P3]	10
Mindset*	Difficult to implement agile mindset [P4, G72]	2
Autonomy*	Teams lacked autonomy [P3, G45]	2
Plan driven or traditional culture*	Struggles to become iterative from fixed delivery cycles [P2, G4, G34], struggle to shift from waterfall culture [G54, G34, G4]	3
<i>Roles</i>		
Resistance for new roles*	Resistance from traditional project managers roles during adoption of SAFe [G44], challenges with change of roles [P4]	2
Staffing roles	Trouble to find the Product Owners [G72, G14, G28, G6, G53] and challenging to find someone with both technical and industrial experience [G28]. Product ownership is complex across universities [G9]. Staffing scrum master was also difficult [G6]	6
<i>Practices</i>		
Value streams	Defining values streams [G62, G6, G14]	3
First Program Increment Planning*	Lack of knowledge on importance about PI [P3], chaotic event [P3, G48], people were uncomfortable during PI [P3] and considered it as unpleasant [G72], clash of time slots to fix PI planning [P3], surprises during PI planning [P3], hard to focus on PI [G10] and ambiguity about time allocation to event [G50], lack of technical knowledge (how to code) during first PI [G50], fail to implement effective PSI and find PSI cadence [G72], teams resist towards PI [P3], feature shaping to PSI was difficult [G72], logistic challenges [G53], technical dependencies in PI's [G8] and management reviews were chaotic [G10]	7
Backlog Management and feature shaping	Feature grooming [G72] and backlog prioritization challenges [G33], not finding right backlog [G9], feature prioritization did not involve every one, mostly a <i>solo effect</i> (only product manager) [G43], feature shaping clashed with deployment requirements [G72]	4

(continued)

Table 4. (continued)

Challenge	Description	# cases
Test automation*	Automated unit tests could not be applied to legacy systems [P2], automated testing was challenging [G4, P5]	3
Controversies with framework*	More complex [G10, P6] and risky [P6], confusion with the way of working [P4], framework as overhead [P3], controversies regarding story point normalization [G72], difficulties with release management in SAFe framework [G33], separation of deployment cycle from PSI cycle was challenging [G72], framework not suitable for organizations working on multiple products [G51]	6
Agile Release Train*	Failure demand of ART's due to ineffective PSI [G72], integration of teams with less dependencies into agile release trains [P1], handling cross team dependencies across the ART's [P1], rearrangement of trains for distributed teams [G33], rejection to take part into ART [G60], difficulties to define ART in organizational context [G8, G14]	6
Moving away from agile*	Moving to SAFe feels like moving back to plan driven methods (such as waterfall and RUP like) [P6, P3, G72, G59], fixed increments [P3], centralized planning [P6], not really incremental [G9], loss of incremental and iterative development [G60], too much detail [G10]	7
<i>Scaling and distribution</i>		
Large and distributed settings	Challenges of full scaling of agile to whole organization [P4] and global organization [P2], integrating non development units such as IT, HR and sales and marketing [P4]	2
GSD*	Collaborative planning meeting and critical gatherings were difficult due to distributed teams [P2], deriving global priorities [G33], different time zones [G33] [G4], scaling agile to global organization [P2], rearrangement of ART's was challenging due to geographic distribution [G33], release planning challenges due to distributed teams [P6]	4

Subjective Bias. This threat occurs during the coding of qualitative data. Coding was performed by the first author very meticulously. Coding was iterative, and all authors had several discussions during the coding process regarding the naming of the axial codes and categorization of the open codes into axial codes. The process is traceable.

Restricted Time Span. In the database search, we included primary studies published in the selected databases until November 2017. For the non-peer reviewed literature, we included all case studies published by May 4th, 2018.

Publication Bias. Including grey, non-peer reviewed, literature can be seen as a limitation. Non-peer reviewed articles usually present positive results [37]. This was also evident from our study, as majority of these cases gave attention to the benefits of the framework. In addition, the Scaled Agile team reviewed all case studies reported by the organizations. There might be a possibility for them to influence the organizations to present only the positive elements of the SAFe adoption process. However, out of 82 documents from grey literature 26 documents came directly from the organizations and other online websites, as additional supplements. These supplements (e.g.: presentations) reported the same information as was published under the case studies on the website. This threat of bias was partially mitigated by comparing the benefits from peer-reviewed primary studies, identified in this MLR (6 studies) and the State of Agile survey [29]. The challenges of adopting SAFe were compared to the findings of the SLR on challenges of large-scale agile transformations [8] as well as to the State of Agile survey [29]. However, to establish scientific evidence there is a strong need for more empirical research on benefits and challenges of SAFe.

Information Loss. The codes with only a few quotes and cases (3 cases or less for benefits, 1 case for challenges) were not reported. The keyword search could have missed some studies. We mitigated this by going through the references and citations of all 5 selected studies. We found one additional case study [P5] from the citations of already selected papers [P3].

6 Conclusions and Future Work

The number of organizations adopting scaling frameworks has increased tremendously during the recent years. A few studies have given insights on agile usage in large organizations, however, the literature on the adoption and usage of scaling frameworks has not been systematically reviewed. Moreover, systematic literature reviews on large-scale agile, have not included the grey literature. This means that most published information about the scaling frameworks has been excluded, giving an incomplete picture, as current research literature on them is very limited. Therefore, we included also grey literature in this multivocal literature review.

We analyzed 54 peer and non-peer reviewed cases on the adoption of the Scaled Agile Framework. The most salient benefit categories were: *transparency, alignment, productivity, predictability and time to market*. The most frequently mentioned challenge categories were: *change resistance, challenges with the first program increment planning and moving away from agile*. The most important difference between the peer-reviewed and grey literature was the bias in reporting benefits, especially with respect to business benefits received from the usage of the framework. These benefits were mentioned only in the grey literature.

This emphasizes the need for validation of the claimed benefits reported in the grey literature. The majority of the challenges were common for both the peer-reviewed and grey literature.

Apart from the challenges related to scaling agile, SAFe has brought in new challenges with respect to practices such as PI planning, value streams, and Agile Release Trains. Empirical research on how the SAFe framework is addressing the existing challenges, that have been reported in the *agile in the large* literature, could be interesting for practitioners. Moreover, finding solutions for the challenges reported in this MLR, would help organizations to address these challenges.

We identified only six peer reviewed primary studies on SAFe since the introduction of the framework (year 2011). Literature lacks in-depth primary studies on the usage and adoption of SAFe. Some of the non-peer reviewed cases published at the SAFe home page had deep insights on the rationale behind the SAFe adoption, transformation steps, implementation of practices, as well as the benefits of the adoption. Unfortunately, both peer-reviewed and grey literature lack extensive information on challenges and the negative traits of SAFe in large enterprises, as there likely is an inherent positive bias in the cases published at the SAFe home page. Hence, it is crucial to conduct more in-depth primary studies on SAFe adoptions to establish scientific evidence on the SAFe framework usage in large scale.

References

1. Alqudah, M., Razali, R.: A review of scaling agile methods in large software development. *Int. J. Adv. Sci. Eng. Inf. Technol.* **6**(6), 828–837 (2016)
2. Alsaqaf, W., Daneva, M., Wieringa, R.: Quality requirements in large-scale distributed agile projects – a systematic literature review. In: Grünbacher, P., Perini, A. (eds.) REFSQ 2017. LNCS, vol. 10153, pp. 219–234. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-54045-0_17
3. Ambler, S.W., Lines, M.: *Disciplined Agile Delivery: A Practitioner’s Guide to Agile Software Delivery in the Enterprise*. IBM Press, Indianapolis (2012)
4. Ampatzoglou, A., Ampatzoglou, A., Chatzigeorgiou, A., Avgeriou, P.: The financial aspect of managing technical debt: a systematic literature review. *Inf. Softw. Technol.* **64**, 52–73 (2015)
5. Boehm, B., Turner, R.: Management challenges to implementing agile processes in traditional development organizations. *IEEE Softw.* **22**(5), 30–39 (2005)
6. Corbin, J.M., Strauss, A.L.: *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, 3rd edn. Sage Publications Inc., Los Angeles (2008). <http://www.loc.gov/catdir/toc/ecip0725/2007034189.html>
7. Denning, S.: Agile: its time to put it to use to manage business complexity. *Strat. Leadersh.* **43**(5), 10–17 (2015)
8. Dikert, K., Paasivaara, M., Lassenius, C.: Challenges and success factors for large-scale agile transformations: a systematic literature review. *J. Syst. Softw.* **119**, 87–108 (2016)
9. Dybå, T., Dingsøy, T.: Empirical studies of agile software development: a systematic review. *Inf. Softw. Technol.* **50**(9–10), 833–859 (2008)

10. Garousi, V., Felderer, M., Mäntylä, M.V.: The need for multivocal literature reviews in software engineering: complementing systematic literature reviews with grey literature. In: Proceedings of the 20th International Conference on Evaluation and Assessment in Software Engineering, p. 26. ACM (2016)
11. Garousi, V., Felderer, M., Mäntylä, M.V.: Guidelines for including the grey literature and conducting multivocal literature reviews in software engineering. arXiv preprint [arXiv:1707.02553](https://arxiv.org/abs/1707.02553) (2017)
12. Garousi, V., Mäntylä, M.V.: When and what to automate in software testing? A multi-vocal literature review. *Inf. Softw. Technol.* **76**, 92–117 (2016)
13. Glass, R.L.: *Software Creativity 2.0*. developer.* Books (2006)
14. Gustavsson, T.: Assigned roles for inter-team coordination in large-scale agile development: a literature review. In: Proceedings of the XP2017 Scientific Workshops, p. 15. ACM (2017)
15. van Haaster, K.: Agile in-the-large: getting from paradox to paradigm
16. Scaled Agile Inc.: Core Values. <http://www.scaledagileframework.com/safe-core-values/>
17. Scaled Agile Inc.: Questionnaire for safe adopters. <https://www.scaledagile.com/resources/submit-a-case-study/>
18. Scaled Agile Inc.: Review process. <https://www.scaledagile.com/case-study-faqs/>
19. Scaled Agile Inc.: Scaled Agile Framework. <http://www.scaledagileframework.com/case-studies/>
20. QSR International: Coding Tool for Qualitative Analysis. <http://www.qsrinternational.com/nvivo/support-overview/downloads>
21. Keele, S., et al.: Guidelines for performing systematic literature reviews in software engineering. In: Technical report, Ver. 2.3 EBSE Technical Report. EBSE. sn (2007)
22. Khalid, H., Ahmed, M., Sameer, A., Arif, F.: Systematic literature review of agile scalability for large scale projects. *Int. J. Adv. Comput. Sci. Appl. (IJACSA)* **6**(9), 63–75 (2015)
23. Kuhrmann, M., et al.: Hybrid software and system development in practice: waterfall, scrum, and beyond. In: Proceedings of the 2017 International Conference on Software and System Process, pp. 30–39. ACM (2017)
24. Larman, C., Vodde, B.: *Practices for Scaling Lean & Agile Development: Large, Multisite, and Offshore Product Development with Large-scale Scrum*. Pearson Education, London (2010)
25. Leffingwell, D.: *Scaling Software Agility: Best Practices for Large Enterprises*. Pearson Education, Boston (2007)
26. Moe, N.B., Dingsøy, T.: Emerging research themes and updated research agenda for large-scale agile development: a summary of the 5th international workshop at xp2017. In: Proceedings of the XP2017 Scientific Workshops, p. 14. ACM (2017)
27. Moe, N.B., Olsson, H.H., Dingsøy, T.: Trends in large-scale agile development: a summary of the 4th workshop at xp2016. In: Proceedings of the Scientific Workshop Proceedings of XP2016, p. 1. ACM (2016)
28. Myrbakken, H., Colomo-Palacios, R.: DevSecOps: a multivocal literature review. In: Mas, A., Mesquida, A., O'Connor, R.V., Rout, T., Dorling, A. (eds.) *SPICE 2017*. CCIS, vol. 770, pp. 17–29. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67383-7_2
29. One, V.: State of Agile Survey. <https://explore.versionone.com/state-of-agile/versionone-12th-annual-state-of-agile-report>
30. Pancholi, A., Grover, S.: Scaled agile framework: a blight. *Int. J. Innov. Res. Dev.* **3**(5) (2014)

31. Rafi, D.M., Moses, K.R.K., Petersen, K., Mäntylä, M.V.: Benefits and limitations of automated software testing: systematic literature review and practitioner survey. In: Proceedings of the 7th International Workshop on Automation of Software Test, pp. 36–42. IEEE Press (2012)
32. Razavi, A.M., Ahmad, R.: Agile development in large and distributed environments: a systematic literature review on organizational, managerial and cultural aspects. In: 2014 8th Malaysian Software Engineering Conference (MySEC), pp. 216–221. IEEE (2014)
33. Rolland, K.H., Fitzgerald, B., Dingsøy, T., Stol, K.J.: Problematizing agile in the large: alternative assumptions for large-scale agile development (2016)
34. Tom, E., Aurum, A., Vidgen, R.: An exploration of technical debt. *J. Syst. Softw.* **86**(6), 1498–1516 (2013)
35. Uludağ, Ö., Kleehaus, M., Xu, X., Matthes, F.: Investigating the role of architects in scaling agile frameworks. In: 2017 IEEE 21st International Enterprise Distributed Object Computing Conference (EDOC), pp. 123–132. IEEE (2017)
36. Vaidya, A.: Does dad know best, is it better to do less or just be safe? Adapting scaling agile practices into the enterprise. In: PNSQC. ORG, pp. 1–18 (2014)
37. Zhou, X., Jin, Y., Zhang, H., Li, S., Huang, X.: A map of threats to validity of systematic literature reviews in software engineering. In: 2016 23rd Asia-Pacific Software Engineering Conference (APSEC), pp. 153–160. IEEE (2016)