Team Performance in Large-Scale Agile Software Development

Tomas Gustavsson
Karlstad Business School, Karlstad University
Karlstad, Sweden *tomas.gustavsson@kau.se*

Abstract

Agile ways of working are nowadays used in many software development departments in larger organizations. When scaling up agile ways of working, new practices for coordinating teams become necessary, and many organizations are implementing the Scaled Agile Framework. The added practices for coordinating teams could have an impact on team performance, but they have not been much studied. In this study, data were obtained by means of a survey questionnaire that was answered by 201 employees from three organizations: one from the automotive industry, one government agency, and a business bank. The study suggests that efficient inter-team coordination does not have a positive relationship to team performance, which is contrary to previous studies. However, results suggests that a high level of psychological safety has a significant positive correlation to team performance.

Keywords: Agile Software Development, Large-Scale, Team performance, Psychological safety, Scaled Agile Framework.

1. Introduction

In software development, there is an industry trend towards adopting large-scale agile ways of working [1]. Although research into the agile approach to software development has matured in the past years, agile ways of working in large-scale settings are not as much explored [18]. One of the fundamental principles in the agile way of working is to allow autonomy to the team [16]. This autonomy is a major reason for success in agile development, and research in other industries also confirms that autonomous and empowered teams are more productive and proactive [13]. The balance between the benefits of autonomous teams versus alignment towards a common goal is, therefore, an essential issue for the software industry [18]. Therefore, large-scale practices for coordinating teams have been proposed to reduce negative impacts while maintaining the positive impacts of agile ways of working in the teams. According to an annually recurring industry survey [1], the most commonly adopted framework today for large-scale agile ways of working is the Scaled Agile Framework (SAFe) which prescribes a number of inter-team coordination practices to be used by the organization. The authors of SAFe, Leffingwell et al. [15], make several claims regarding expected beneficial impacts by implementing SAFe based on case studies written by end-users. The claimed benefits include increased team performance due to better coordination and more motivated employees. No drawbacks are mentioned.

But SAFe has been criticized by agile practitioners in several ways, such as being too top-down and inflexible [20], taking away the benefits of autonomy to the team. Schwaber [20] argues that it is more critical for performance to build autonomous teams than to "suffocate" the teams with coordination practices between groups.

A study by Hoegl et al. [11] shows that inter-team coordination has a positive effect on team performance in the individual team, but only to some extent. In their study, inter-team coordination had a positive impact on schedule performance but did not show a positive impact on quality or budget, which were the three areas investigated. One could discuss if these three areas are proper to be addressed under the umbrella term “team performance”. Rather, one might call them areas of “project performance” since, for example, schedule performance could be challenged by surrounding factors such as delayed deliveries from other teams, vendors, or other parts of the organization.

In this study, the team performance construct used in Edmondson [5] will instead be applied to investigate team performance. The construct consists of four items (statements) inquiring the perceived team performance from the individual team members' perspective. This team performance construct suits this study better since the other two constructs used in this study are also based on the perceived experience from the individual team member.

Regarding efficiency in autonomous teams, Edmondson [5] shows that psychological safety is an important mechanism that influences behavioral and performance outcomes. Edmondson defines team psychological safety as a shared belief that the team is safe for interpersonal risk-taking. In a team where members trust each other, dare to seek help from each other, ask questions, and tolerate mistakes, the learning improves. This, in turn, has a positive effect on team performance [5]. Also, the trending values of Modern Agile suggests that agile organizations should stand on four guiding principles, of which one is "Make Safety a Prerequisite" [17].

The question is whether positive effects on team performance is evident in an organization working according to SAFe and if the effect has a stronger impact on team performance than inter-team coordination? The purpose of this study is, therefore, to investigate the correlation between team performance and inter-team coordination in a large-scale agile environment, where organizations have implemented the SAFe framework. This research aims at developing our understanding of the impacts of additional coordinating practices on team performance. Specifically, the following research questions are examined: 1) Is there a positive correlation between inter-team cooperation and team performance in large-scale agile software development? 2) Is there a positive correlation between psychological safety and team performance in large-scale agile software development?

First, a number of possible impacts on teamwork will be examined to understand if they are perceived as being improved by the implementation of SAFe. Secondly, the derived hypotheses regarding inter-team coordination and psychological safety constructs relating to team performance will be examined in further detail to be able to answer the research questions. The hypotheses are tested by using data from a survey study conducted at three large-scale development environments; one in the automotive industry (24 development teams), one at a government agency (7 teams), and one department in a business bank (7 development teams).

The rest of the paper is organized as follows: Section 2 describes large-scale agile software development in more detail. In Section 3, constructs and hypotheses are explained. Section 4 shows the research methods used in the study, and Section 5 describes the results. Discussions of the results are presented in Section 6 as well as limitations to the study.

2. Large-Scale Agile Software Development and Modern Agile

Agile Software Development (ASD) is often defined by the values and principles as described in the "Manifesto for Agile Software Development" [16] which mostly explain what is important for team members, working together in one team. The same goes for Scrum, the most commonly implemented agile framework [1]. But neither the Agile Manifesto [16] nor Scrum is enough to explain how to most efficiently organize multiple teams working together towards a common goal where teams need to coordinate their work. Therefore, several new frameworks designed for large-scale agile adoptions have emerged, of which SAFe is the most commonly adopted today [1]. It was introduced in August of 2013 and has been further developed, and is now in version 5.1. SAFe describes agile ways of working on different levels, starting from the team level to the program level, portfolio level, and organizational level. SAFe prescribes several practices for coordination between teams, Product Owners and other roles in the organization. The critique against SAFe, being too top-down and inflexible [20], is that it seems like an underlying assumption that high performance in the teamwork is obtained by having detailed practices for coordination.

The trending concept Modern Agile [17] focuses on what values are important instead

of stressing specific practices or artifacts to be used. Modern Agile focuses on describing important values to be shared in the whole organization to become effective from a single team to top management. Four guiding principles are put forth: 1) Make People Awesome, 2) Make Safety a Prerequisite, 3) Experiment and Learn Rapidly, and 4) Deliver Value Continuously. Regarding the second principle, the following statement clarifies the importance: "Safety is both a basic human need and a key to unlocking high performance" [17]. For Modern Agile, an underlying assumption is that psychological safety will produce high team performance.

3. Constructs and hypotheses

The conceptualization of team performance as a multidimensional construct is widely acknowledged in the literature [11, 5] but with different views on its content. In Section 3.1, team performance and its relation to inter-team coordination, according to Hoegl et al. [11], will be discussed. The view on team performance and its relationship to psychological safety, according to Edmondson [5], will thereafter be presented in Section 3.2.

3.1. Inter-team coordination and team performance

According to Hoegl et al. [11], team performance can be defined as the extent to which a team can meet established objectives. For a development team responsible for developing specific parts of a larger system, several properties may be important, including adherence to predefined quality, a schedule where certain deliverables are expected at predefined times, and costs associated with the team's development activities [11].

In the paper by Hoegl et al. [11], earlier work in defining the latent construct teamwork quality is explained. In a study of software development teams, the within-team collaboration process was conceptualized and empirically validated as a multifaceted higher-order construct. The construct consists of six facets: communication, coordination, the balance of member contributions, mutual support, effort, and cohesion. These facets embrace elements of both task-related and social interaction within teams, and the underlying proposition is that highly collaborative teams display behaviors related to all six teamwork quality facets. The results of the study showed that teamwork quality, to some extent, has a positive influence on team performance [11].

Regarding inter-team coordination, as a team depend on input from another team for accomplishing their task, the work in one team has implications for the work and progress of other teams. There are mainly two forces that create coordination needs between teams in multi-team projects: 1) task interdependencies and 2) changes occurring during the development process [11]. While some coordination needs are possible to plan in advance, software development is always characterized by changes, which often affect the work of several teams. The complexity and uncertainty of development processes, based on interdependencies between teams and frequent changes, can only be dealt with if the information is exchanged between the teams [11].

Hoegl et al. [11] showed that although coordination with other teams might take time and resources, it has, to some degree, a positive effect on team performance. As previously described, Hoegl et al. [11] measured team performance as a combination of schedule performance (delivering on time), adherence to budget, and quality. In their study, inter-team coordination showed a significant positive relationship to schedule performance but not to quality and a negative relationship to adherence to budget.

Hypothesis 1. Coordination with other teams is positively associated with team performance.

3.2. Psychological safety and team performance

Asking for help, admitting errors, and seeking feedback are examples of behaviors that pose a threat to face [5]. Therefore, people in organizations are often reluctant to disclose their errors or are unwilling to ask for help, even when doing so would provide benefits for

the team or organization. Edmondson's [5] study shows that high levels of psychological safety have a positive effect on team performance, or rather that learning behavior mediates between team psychological safety and team performance. In her study, team performance is measured by using the scale invented by Hackman [9] to obtain self-report measures of team performance. Edmondson [5] also developed a similar scale to be used by observers who rated team performance.

Hypothesis 2. Psychological safety is positively associated with team performance.

Fig 1 shows the two hypotheses tested in this study.

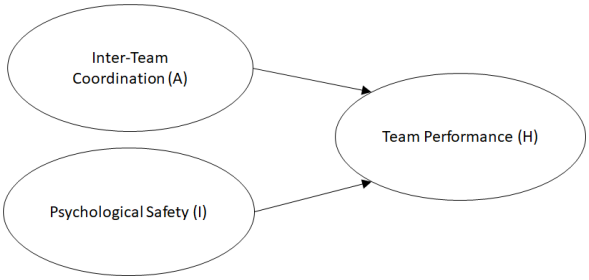

Fig. 1. A model of impact on team performance.

4. Method

The research setting is first described in this section (4.1), followed by how data was collected from the three case organizations (4.2). Measures and scales are thereafter presented (4.3), followed by descriptions of how data was prepared and analyzed (4.4).

4.1. Research setting

Two inclusion criteria were used to find cases for this study. First, the case organizations should not be new to agile ways of working, i.e. the study was not indented to compare waterfall and agile. The second criterion was that they should have implemented the same large-scale agile team coordination practices in order to make analysis relevant. Three different case organizations met the criteria and the actual names of the organizations have been anonymized but will be referred to as Motor, Agency, and Bank. The organizations have used agile ways of working for four to six years, with self-organized autonomous teams working side by side. All three organizations decided to adopt practices for team coordination from the framework SAFe. They all started implementing SAFe during the beginning of 2017. Motor was first, starting in January, while Agency began in March and Bank in April. The development organizations are divided into a number of teams with one Scrum Master per team and almost one Product Owner per team (some are responsible for two teams).

Motor is a department in an organization within the automotive industry that mainly develops software but, to some extent, hardware as well. The observed department, when the survey was conducted, was organized into 24 cross-functional teams, divided into three different set of teams or Agile Release Trains (ART) to use SAFe terminology [15]. Roughly 80 percent of the product development is software development and 20 percent hardware development. Some of the teams are rather small, so the total amount of people working in the teams is 141, which means an average of 5.9 people per team. The average age in the department is 36.9 years old, with an average of 9.5 years working at Motor.

The Agency project is a SAFe implementation that started as a pilot project in a large Swedish government agency (more than 13 000 employees). Large-scale agile processes were implemented with the aim of finding best practices to be used for the whole organization. Agency consists of seven teams working together in one ART. The total amount of employees in this software development organization is 70 people. This means that the average team consisted of 8.75 people which is the largest average team size of the three investigated organizations. The average age at Agency is 44.9 years, with an average

of 10.5 years working at Agency.

Bank is a software development department in one of the major business banks in Sweden (more than 15 000 employees) consisting of seven teams that work together in one ART. Bank decided to implement large-scale agile practices because a new software platform was being developed. This project would increase the number of dependencies between all teams in the department, hence the increased need for coordination practices. The department consists of 7 teams with 42 team members, which means that the average team consists of 6 people. The average age in the department is 38.9 years old with an average of 9.6 years working at Bank.

4.2. Data collection and sample

The data was collected using a paper-based questionnaire which was handed out and collected during a two-day planning workshop known as PI planning in SAFe [15]. After a brief introduction of the study, the questionnaire was handed to the respondent to complete by reading it himself or herself.

The survey was conducted in February 2018 at Motor and Agency and in April 2018 at Bank, which means that the organizations had used team coordination practices from SAFe for roughly one year.

For descriptive statistics, 201 survey responses could be used: 109 from Motor, 56 from Agency and 36 from Bank which represents a 78,5 percent (201/256) response rate. This involved questions on opinions on working according to the SAFe framework and perceived differences between the previous way of working and current way of working. One hundred fifty of these responses came from team members. Missing value analysis was conducted on all survey responses. No patterns emerged, and only eight observations were removed due to excessive missing data.

For structural equation modeling (i.e. questions regarding inter-team coordination, team performance, and psychological safety), only responses from employees working in teams were used. Unfortunately, most responses from Agency on these areas were left blank. There are several possible reasons for why the respondents left many statements blank: 1) Agency did not implement all SAFe practices on team-level [7], 2) out of the three cases, most negative voices were raised from employees at Agency [7], and 3) the questionnaire was handed out at the end of the first day of PI planning, and many employees wanted to leave work and go home.

Whatever the reason, with such excessive missing data, it was decided to better exclude all responses from Agency. Rather than risking biased results due to replacing too much missing data with an algorithm method, all 56 responses from Agency were therefore excluded in this part of the analysis. Thus, a total of 80 survey responses could be used from Motor and 31 from Bank, leading up to an overall total of 111 responses, which represents a 61 percent (111/183) response rate from these two organizations. At Motor, 60 responses came from developers, 14 from Scrum Masters, and six from Product Owners. At Bank, 25 responses came from developers, four from Scrum Masters, and only two from Product Owners.

The low response rate from Product Owners is because some of them did not see themselves as part of the team. As Bass and Haxby explains [4], Product Owners in a large-scale agile context undertake a range of challenging and varied activities beyond those conventionally associated with the role. Some Product Owners work very closely with the team, while some only meet the team on an irregular basis and spend most of their time working with other stakeholders. The latter tend not to see themselves as actual team members, and they chose not to answer the questions regarding inter-team coordination, psychological safety, or team performance in this survey.

4.3. Measures and scales

The questionnaire consisted of multiple sections: (1) background (e.g., organizational unit), (2) agile role and experience, (3) opinions on working according to the SAFe

framework, (4) perceived differences between the previous way of working and current way of working, (5) inter-team collaboration in agile development, (6) teamwork quality and team performance, and (7) psychological safety.

Also, both open questions and multi-choice questions were included in the questionnaire, providing possibilities for both qualitative and quantitative data analysis. In this paper, only the quantitative data from sections 3, 4, 5, 6, and 7 are used. A mix of negatively and positively worded items was used to mitigate response set bias. The questionnaire was administered in Swedish at Agency and Bank and in English at Motor since the organization has offices in several countries and use English as the official corporate language.

To understand the overall attitudes towards SAFe practices, five questions were asked inspired by the questionnaire invented by Laanti et al. [14] presented in Table 2. For teamwork quality, one item per facet was used, as presented in Table 3. The questions were invented by the author attempting to capture the perceived difference in impacts during the time of the survey collection compared to before the implementation of SAFe. The measurement scale for inter-team coordination (Cronbach's alpha = 0.705) consists of four items. The items used are described in Hoegl et al. [11], who partly adapted them from scales used by Mott [19], who evaluated coordination, communication, and cooperation between different occupational groups in hospitals.

For psychological safety (Cronbach's alpha = 0.729), the measurement scale consists of four items used in the measurement scale invented by Edmondson [5]. The measurement scale for team performance (Cronbach's alpha = 0.712) consists of four items described in Edmondson [5], who used the scale invented by Hackman [9] to obtain self-report measures of team performance. Due to low communalities and low factor loading on one item in the construct (item H3), only three items were used in the analysis.

Exploratory factor analyses were conducted with the items of inter-team coordination, psychological safety, and team performance, respectively, to confirm the internal consistency of the three scales. Using the Kaiser criterion, the factor analyses resulted in one-factor solutions for all three constructs. Questionnaire items are displayed in table 1.

Table 1. Questionnaire items. Reversed items are shown as (R).

Construct	#	Statement
Inter-team coordination	A1	Processes and activities are well coordinated with other teams.
	A2	Duplicated and overlapping activities are avoided.
	A3	Discussions with other teams are conducted constructively.
	A4	Conflicts with other teams are settled quickly.
Team performance	H1	Recently, this team seems to be "slipping" a bit in its level of performance. (R)
	H2	Those who receive or use the work this team does often have complaints about our work. (R)
	H3	The quality of work provided by this team is improving over time.
	H4	Critical quality errors occur frequently in this team. (R)
Psychological safety	I1	If you make a mistake on this team, it is often held against you. (R)
	I2	Members of this team are able to bring up problems and tough issues.
	I3	People on this team sometimes reject others for being different. (R)
	I4	It is safe to take a risk on this team.

The first part of the analysis, with all respondents (N=201), was done in SPSS. The second part of the analysis was conducted with structural equation modeling in LISREL [12]. The analyses in LISREL are performed at the individual level, with survey responses from employees working in development teams only (N = 111). With such a small group in LISREL, the reliability of the estimates is negatively affected, and the power is low for significance tests.

Following Anderson and Gerbing's [2] two-step approach, construct validity was assessed (convergent and discriminant validity) and nomological validity in the measurement model before considering the structural model. The rationale is that this alleviates the interaction of the measurement and structural models, allowing for a more accurate assessment of validity and reliability [3].

5. Results

In this section, research data is presented and analyzed to evaluate findings in the descriptive data (Section 5.1), and thereafter, a model estimation using LISREL is presented (Section 5.2).

5.1. Measures and scales

The perceptions of the respondents regarding large-scale agile development work according to SAFe were captured in the "opinions" section of the survey. Firstly, three specific statements were presented to evaluate the respondents' view of three inter-team coordination practices. The fourth statement read: "There are many problems with working in a large-scale setting like ours, using SAFe" and the fifth statement asked whether the respondent wanted to go back to the old way of working before SAFe. These questions/statements were answered by all employees (N=201).

A Likert scale from 1 to 5 was used to collect the responses: 1 = totally disagree, 3 = neutral, and 5 = totally agree. The data distribution to each of the questions was symmetric and normal. In Table 2, descriptive statistics regarding the five statements are presented in three ways. First, the whole population (N=201), second, employees working in teams (N=150), and third, the stakeholders who are all respondents not working in a team (N=51).

Table 2. Opinions statements and descriptive statistics.

Statements (Section 3 in the questionnaire)	*N*	Mean	Variance	Std. Dev.
1) The PI planning gives me a good overview of our work	201	3.91	0.758	0.871
- Team members	150	3.78	0.710	0.843
- Stakeholders	51	4.25	0.872	0.934
2) The program board is very helpful for coordination between teams	201	3.20	0.966	0.983
- Team members	150	3.18	0.945	0.972
- Stakeholders	51	3.30	1.087	1.043
3) The Scrum of Scrums (SoS) meetings solve inter-team coordination problems	201	3.33	0.962	0.981
- Team members	150	3.28	1.037	1.018
- Stakeholders	51	3.54	0.520	0.721
4) There are many problems with working in a large-scale setting like ours, using SAFe	201	3.48	1.020	1.010
- Team members	150	3.48	0.989	0.994
- Stakeholders	51	3.57	0.995	0.998
5) I would like to go back to the old way of working (instead of working according to SAFe)	201	2.13	1.404	1.185
- Team members	150	2.27	1.409	1.187
- Stakeholders	51	1.74	1.147	1.071

As can be seen, responses to the three statements regarding the benefits of coordination practices (PI planning, program board, and Scrum of Scrums) were all above average. In all three statements, stakeholders are more positive than team members. According to the fifth statement ("I would like to go back to the old way of working (instead of working according to SAFe)"), the average is closer to "agree" than "disagree" for both groups. These numbers show that most respondents do not want to go back to the former way of working, before SAFe, especially not the stakeholders.

After having conducted an analysis of variance test between the two groups, using ANOVA, with a p-value threshold of 0.05, the null hypothesis could only be rejected for the first and fifth statements. Hence, there is a significant difference between the two groups regarding the value of PI planning and whether to go back to the old way of working

before SAFe.

The next section of survey questions was aimed at team members only. Nine attitude questions were presented to evaluate the respondents' overall attitude toward the perceived impacts on teamwork, based on the areas suggested by Hoegl et al. [11] and Edmondson [5]. A Likert-like scale from 1 to 5 was used to collect the responses: 1 = worse, 3 = same, and 5 = better. The data distribution to each of the questions was symmetric and normal.

Table 3. Attitudes towards perceived impacts on teamwork.

Statements (Section 4 in the questionnaire)	*N*	Mean	Variance	Std. Dev.
How does coordination *with other teams* work now, compared to one year ago (before implementing SAFe)?	150	3.50	1.041	1.020
How does communication *within the team* work now, compared to one year ago (before implementing SAFe)?	150	2.92	1.497	1.224
How is the *balance* between member contributions within the team now, compared to one year ago (before implementing SAFe)?	150	2.93	1.206	1.098
How does coordination *within the team* work now, compared to one year ago (before implementing SAFe)?	150	3.09	1.196	1.094
How does *support* within the team (helping each other) work now, compared to one year ago (before implementing SAFe)?	150	3.01	1.437	1.199
How is the *cohesion* within the team now, compared to one year ago (before implementing SAFe)?	150	3.12	1.395	1.181
How does the personal *effort* to achieve goals differ now, compared to one year ago (before implementing SAFe), within the team?	150	3.11	1.270	1.127
How does your *team performance* differ now, compared to one year ago (before implementing SAFe)?	150	3.04	1.144	1.070
In what way has *trust* and sense of personal security changed within the team compared to one year ago (before implementing SAFe)?	150	2.75	1.408	1.187

As can be seen, the highest positive response was inter-team coordination with an average of 3.50, and the lowest was the statement regarding psychological safety with an average of 2.75, which is below neutral. In other words, the employees perceive an impact of better inter-team coordination but a lowered perceived psychological safety on average. For the seven other statements, they were all close to neutral, with two statements below neutral and five slightly above.

5.2. Model estimation

Table 4 shows descriptive statistics for the three constructs used for hypothesis testing. The skewness and kurtosis statistics indicate that most of the indicators are normally distributed, with some exceptions. Because of these exceptions, raw data was used in LISREL since it is not dependent on a normal distribution.

Table 4. Scale statistics.

Scales (Sections 5, 6, and 7 in the questionnaire)	Item	Mean	Std. Dev.	Skewness	Kurtosis
Inter-team coordination	A1	3.54	0.850	-0.580	-0.030
	A2	3.67	0.800	-0.322	-0.234
	A3	4.04	0.713	-0.513	0.433
	A4	4.01	0.803	-0.874	1.359
Team performance	H1	3.64	1.034	-0.483	-0.333
	H2	4.14	0.720	-0.655	0.602
	H4	4.13	0.832	-0.916	1.064
Psychological safety	I1	4.61	0.649	-1.648	2.289
	I2	4.45	0.599	-0.581	-0.573
	I3	4.58	0.793	-2.209	5.137
	I4	3.91	0.880	-0.557	0.110

Convergent validity is achieved if the model fits the data well, t-values associated with the individual items are significant, and if the measures are reliable [6].

Table 5 presents the measurement model statistics. The model fits the data well, and all t-statistics for indicator loadings are significant, which indicates convergent validity [6]. However, the variance extracted for the constructs is not above 0.5, and composite reliability is not above 0.7, so the measured constructs are not entirely reliable [6]. Nomological validity was assessed through the normed Satorra-Bentler scaled x^2 and degrees of freedom, which measures the distance between data and model. The ratio of the x^2 divided by the degrees of freedom should be below 2 [12]. Unfortunately, the p-value which provides an additional measure should be above 0.05 for significance at the 5 percent level but is only 0.028. However, the root mean square error of approximation of 0.065 is well below 0.08 [12].

Table 5. Measurement model statistics.

Scale	Item	Standardized loading	*t*-value	Error	Composite reliability	Variance extracted
Inter-team coordination	A1	0.66	Fixed	0.57	0.67	0.34
	A2	0.60	4.14	0.64		
	A3	0.53	3.89	0.72		
	A4	0.53	3.90	0.72		
Team performance	H1	0.46	3.56	0.79	0.63	0.40
	H2	0.92	Fixed	0.15		
	H4	0.38	3.15	0.86		
Psychological safety	I1	0.70	Fixed	0.51	0.65	0.33
	I2	0.46	3.82	0.79		
	I3	0.60	4.64	0.64		
	I4	0.49	4.03	0.76		

Notes: $x^2 = 59.98$, df = 41, p = 0.028, RMSEA = 0.065

The test of discriminant validity was conducted to estimate a confidence interval (+/- 2 standard errors) around the standardized correlations between latent constructs (off-diagonal of the Φ matrix in LISREL). The interval should not include 1 [18]. In Table 6 displaying correlations between latent constructs, none of the confidence intervals around the standardized correlations between latent variables include one. The three constructs all passed this test.

Table 6. Correlations between latent constructs.

Correlations			**Confidence interval**
	Psychological safety (PS)	Inter-team coordination	PS-ITC: 0.31 + 2(0.04) = 0.31
Inter-team coordination (ITC)	0.31 (0.04)		PS-TP: 0.65 + 2(0.05) = 0.75
Team performance (TP)	0.65 (0.05)	0.33 (0.05)	ITC-TP: 0.33 + 2(0.05) = 0.43

In the structural model, there is the added component of causal relationships between constructs. Table 7 shows the result of the hypotheses tested in the study.

Table 7. Structural model statistics.

Hypothesis	**Standardized loading**	***t*-value**	**Error**	**Outcome**
H1: Inter-team coordination -> Team performance	0.14	1.17	0.14	Fails to reject H_0
H2: Psychological safety -> Team performance	0.60	4.17	0.21	Reject H_0

With a *t*-value below two (1.17), H1 is not supported. This means that inter-team

coordination is not significantly correlated to team performance. H2, with a t-value of 4.17, is supported. Psychological safety is significantly correlated to team performance. The statistics of the structural model are also displayed in Fig. 2.

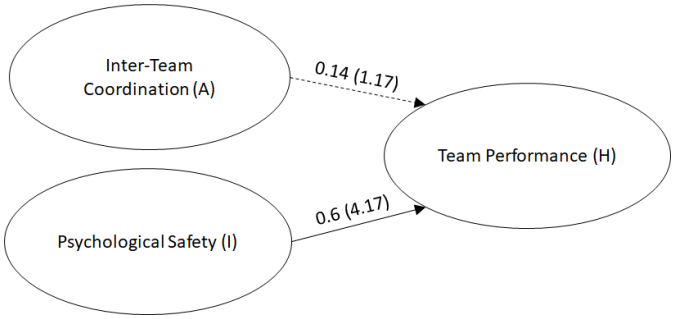

Fig. 2. Structural model.

In order to clarify the results, we need to elaborate and discuss the correlations as well as the descriptive statistics on attitudes towards working according to SAFe. This is done in the next section (Section 6).

6. Discussion

The descriptive statistics show that most respondents do not want to go back to the old way of working as it was before implementing SAFe. That was especially evident for the stakeholders. The answers to the following statement showed a statistically significant difference between stakeholders and team members: "The PI planning gives me a good overview of our work". Although the average was above 3 for both groups (4.25 for stakeholders and 3.78 for team members), this suggests that coordination practices are more appreciated by stakeholders than team members.

The result of investigating the perceived impacts of different aspects of teamwork was somewhat surprising. The six facets of teamwork quality [11] (communication, coordination, balance of member contributions, mutual support, effort, and cohesion) were all close to neutral, with four averages above three and two below three. A high average for increased inter-team coordination was expected (3.5) since the added practices are intended for coordination. Surprisingly, psychological safety was on average perceived as lower (2.75) than before the implementation of SAFe. A possible reason to the decreased psychological safety could be that some people are shy or afraid of speaking up in front of an audience [8]. Several practices, such as PI planning and system demos, requires people to speak up in front of a large number of people. Mistakes could then be highlighted for the whole ART, not just in the single team, which might induce a sense of insecurity for the individual team member.

The results of the structured equation modeling indicate that successful inter-team coordination in large-scale agile software development does not correlate to high team performance. Although this result is contrary to Hoegl et al. [11], one must remember that their study only confirmed a positive relationship with schedule performance (delivering on time), not with quality or adherence to budget. Also, team performance in this study is based on self-reflecting answers, i.e., how the individuals perceived the performance of the team. However, although research shows a positive relationship between inter-team coordination and the overall project success (e. g. [10]), maybe inter-team coordination is not as important for the performance of the individual team. Perhaps it should rather be seen as a teamwork skill on its own, not specifically important for team performance but rather for the performance of a whole project, consisting of several teams.

The study confirms a positive correlation between psychological safety and perceived team performance for teams using SAFe practices for team coordination. This correlation has been proven in other areas as well [5] and does not come as a surprise. The result somewhat confirms the assumption of Modern Agile, which declares psychological safety as a "key to unlocking high performance" [17].

7. Conclusion

When larger organizations scale up their agile ways of working, new practices for coordinating teams becomes necessary. The implemented practices for coordinating teams could have an impact on team performance, but they have so far not been much studied. This study suggests a positive correlation between psychological safety and perceived team performance for teams working in a large-scale setting. However, successful inter-team coordination does not correlate to high team performance, which is contrary to previous studies [11].

The managerial implication of this result is that inter-team coordination does not seem important for team performance, per se. Although originators of SAFe [20] claim increased productivity and more motivated employees, this study does not support that the claimed benefits stem from the added inter-team coordination practices. This does not mean that inter-team coordination is of low importance. It only suggests that improving inter-team coordination will not increase team performance for the individual team. The results do, however, support the second hypothesis, that psychological safety correlates with team performance in a large-scale agile software development setting, where the organization has implemented SAFe team coordination practices.

Regarding managerial implications, this means that large-scale agile software development organizations will benefit from helping teams to increase their level of psychological safety to raise team performance. This confirms previous studies on the impact of high levels of psychological safety [5]. Also, it is important for managers to be aware of the potential decrease of psychological safety. This study shows the average of perceived psychological safety now, compared to before implementing SAFe, was only 2.75 (where three is neutral).

Some limitations needs to be discussed. The results of the descriptive statistics on attitudes towards team coordination practices and perceived impacts could be seen as indicators and are of low explanatory value. For each teamwork quality aspect, only one item has been used per construct, which was not the case in Hoegl et al. [11], where several items per construct were used. But in this study, the constructs were not used for hypothesis testing, only as a way of understanding overall attitudes and perceived impacts. This understanding was important to decide on further analysis in the study. For the structural equation modeling, these results need to be further confirmed since the reliability of the constructs is questionable with low variance extracted as well as composite reliability.

For each investigated factor, only four items were used, which is rather low, and due to low communalities and low factor loading on one item in the team performance construct (item H3), only three items were used for that factor. The argument for not using more than four items per factor was that this was only a part of the survey, which in total consisted of 53 questions. With too many questions, there might be an increased risk of getting few filled out or only partially filled out survey responses. Fortunately, the number of questions did not seem to discourage participation since the response rate reached as high as 78.5 percent in total (201/256) and 61 percent (111/183) useful for hypotheses testing.

Future research needs to be done to confirm these findings on team performance, e.g. by investigating more organizations using SAFe or with similar team coordination practices since the number of participant responses used for hypotheses testing was somewhat low (N=111).

References

1. 14th Annual State of Agile™ Report (2020), http://stateofagile.versionone.com. Accessed December 18, 2020
2. Anderson, J.C., Gerbing, D.W.: Structural equation modeling in practice: a review and recommended two-step approach, Psychological Bulletin, 103(3), pp. 411-423 (1988)
3. Anderson, J.C., Gerbing, D.W.: Assumptions and comparative strengths of the two-step approach: comment on Fornell and Yi, Sociological Methods and Research, 20(3), pp. 321-333 (1992)

4. Bass, J.M., Haxby, A.: Tailoring Product Ownership in Large-Scale Agile. IEEE Software, December, (2018)
5. Edmondson, A.: Psychological safety and learning behavior in work teams. Administrative science quarterly, 44(2), 350-383 (1999)
6. Fornell, C., Larcker, D.F.: Evaluating structural equation models with unobservable variables and measurement error, Journal of Marketing Research, 18(1), pp. 39-50 (1981)
7. Gustavsson, T.: Changes Over Time in a Planned Inter-Team Coordination Routine. In Proceedings of the XP2019 Scientific Workshops, May 21-25, Montreal, Canada (2019)
8. Gustavsson, T: Voices From the Teams - Impacts on Autonomy in Large-Scale Agile Software Development Settings. In Proceedings of the XP2019 Scientific Workshops, May 21-25, Montreal, Canada (2019)
9. Hackman, JR: Groups that Work (and Those That Don't). Jossey-Bass, San Fransisco (1990)
10. Hise, R.T., O'Neal, L. Parasuraman, A., McNeal, J.U.: Marketing/R&D interaction in new product development: Implications for new product success rates. J. Product Innovation Management 7 142–155 (1990)
11. Hoegl, M., Weinkauf, K., Gemuenden, H. G.: Interteam coordination, project commitment, and teamwork in multi-team R&D projects: A longitudinal study. Organization Science, 15(1), 38-55 (2004)
12. Joreskog, K., Sorbom, D.: Lisrel 8: Structural Equation Modeling with the Simplis Command Language, Lawrence Erlbaum Associates, Hillsdale (1993)
13. Kirkman, B. L., Rosen, B.: Beyond self-management: Antecedents and consequences of team empowerment. Academy of Management Journal, 42(1): 58-74, (1999)
14. Laanti, M., Salo, O., Abrahamsson, P.: Agile methods rapidly replacing traditional methods at Nokia: A survey of opinions on agile transformation. Information and Software Technology 53 (3), 276-290 (2011)
15. Leffingwell, D., Yakyama, A., Knaster, R., Jemilo, D., Oren, I.: SAFe reference guide: Scaled Agile Framework for Lean Software and Systems Engineering: Addison Wesley, USA (2017)
16. Manifesto for Agile Software Development (2001), http://agilemanifesto.org. Accessed March 15, 2021
17. Modern Agile (2014), http://modernagile.org. Accessed January 11, 2021
18. Moe, N. B., Olsson, H. H., Dingsøyr, T.: Trends in Large-Scale Agile Development: A Summary of the 4th Workshop at XP2016. In: Proceedings of the Scientific Workshop Proceedings of XP2016, (2016)
19. Mott, P.E.: The Characteristics of Effective Organizations. Harper & Row, New York (1972)
20. UnSAFe at any speed (2013), http://kenschwaber.wordpress.com/2013/08/06/unsafe-at-any-speed/. Accessed January 12, 2019