Impacts on Team Performance in Large-Scale Agile Software Development

Tomas Gustavsson[0000-0002-1512-6592]

Karlstad University, Karlstad, Sweden
tomas.gustavsson@kau.se

Abstract. Agile ways of working are nowadays used in many software development departments even in larger organizations. When scaling up agile ways of working, new practices for coordinating teams become necessary. Instead of inventing practices on their own, many organizations are implementing the Scaled Agile Framework but the impacts on team performance due to these practices are not much studied. Data was obtained by means of a survey questionnaire that was answered by 111 team members from two organizations, one from the automotive industry and one of the major Swedish banks. The study suggests that efficient inter-team coordination does not have a positive relationship to team performance. But, as shown in several other contexts, a high level of psychological safety has a significant positive impact on team performance in a large-scale agile software development setting.

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

1 Introduction

In software development there is an industry trend towards adopting agile methodologies in-the-large [1] and 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 much explored [2]. One of the fundamental principles in the agile way of working is to allow autonomy to the team. This autonomy is a major reason for success in agile development and research in other industries also confirm that autonomous and empowered teams are more productive and proactive [3]. The balance between benefits of autonomous teams versus alignment towards a common goal is, therefore, an important issue for the software industry [2]. To maintain this balance, 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 prescribe a number of inter-team coordination practices to be used by the organization. The authors of SAFe, Leffingwell et al. [4] 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 other agile practitioners in several ways such as being too top-down and inflexible [5], taking away the benefits of autonomy to the team. Schwaber [5] argues that it is more important for performance to build autonomous teams than to "suffocate" the teams with coordination practices between teams.

A study by Hoegl et al. [6] show that inter-team coordination has a positive effect on team performance in the individual team, but only in some areas. The question is if this positive effect is evident in a large-scale agile environment, where organizations have implemented the SAFe framework.

Regarding efficiency in autonomous teams, Edmondson [7] showed that psychological safety is an important mechanism that influences behavioral and performance outcomes. In a team where team members dare to seek help from each other, ask questions and tolerate mistakes improves team learning which in turn has a positive effect on team performance. The question is if this positive effect is evident in an organization working according to SAFe and whether the effect has a stronger impact on team performance than inter-team coordination?

Specifically, the following research questions are examined: (1) Is there an influence of inter-team cooperation on team performance in large-scale agile software development settings? (2) Is there an influence of psychological safety on team performance in large-scale agile software development settings? (3) Which has the most positive effect on team performance, inter-team coordination or psychological safety?

This research aims at developing our understanding of the coordinating practices and effectiveness of teams in large-scale agile software development settings. The derived hypotheses are relating inter-team coordination and psychological safety to team performance of the individual team members.

The hypotheses are tested by using data from a survey study conducted on two large-scale development projects, one in the automotive industry (20 development teams) and a major Swedish bank (7 development teams).

2 Constructs and hypotheses

The conceptualization of team performance as a multidimensional construct is widely acknowledged in the literature [6, 7]. In general, team performance can be defined as the extent to which a team is able to 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 [6].

2.1 Inter-team coordination and team performance

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 [6].

Task interdependencies refer to the direction of a workflow relationship between two teams. As a team depend on input from another team for accomplishing their own task, the work in one team has implications for the work and progress of other teams.

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.

Hoegle et al. [6] showed that although coordination with other teams might take time and resources, it has to some degree a positive effect on team performance. Hoegl et al. [6] measured team performance as a combination of schedule performance (delivering on time), adherence to budget and quality. 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.

2.2 Psychological safety and team performance

Asking for help, admitting errors, and seeking feedback are examples of behaviors that pose a threat to face [7]. 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 [7] study shows that high levels of psychological safety have a positive effect on team outcomes.

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

Fig 1 shows the two hypotheses tested in this study.

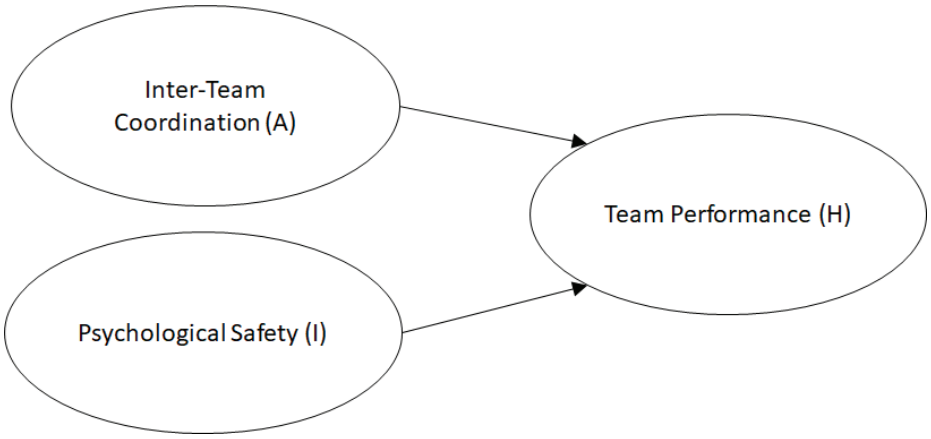

Fig. 1. A model of impact on team performance in a large-scale agile software development context.

3 Method

3.1 Research setting

This study was conducted in two different organizations. The true names of the organizations have been anonymized but will be referred to as Auto and Bank. The organizations have used agile ways of working for four to six years, with self-organized autonomous teams working side by side. But both organizations decided to adopt practices for improved coordination and started implementing SAFe during the beginning of 2017. Auto was first, starting in January while Bank started in April.

Auto is a department in an organization within the automotive industry who mainly develops software but to some extent hardware as well. The observed department is organized in 20 cross-functional teams, divided into three different value streams or Agile Release Trains (ART) to use SAFe terminology [4].

Bank is a department in one of the major business banks in Sweden consisting of seven teams that work together in one ART. They decided to implement large-scale agile practices because a new software platform was being developed which would increase the number of dependencies between all teams in the department.

3.2 Data collection

The data was collected using a paper-based questionnaire at the two organizations which was handed out and collected during a two-day planning workshop known as PI-planning in SAFe [4]. 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 Auto and in April 2018 at Bank which means that both organizations had worked according to SAFe for one year.

3.3 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.

In addition, 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 section 5, 6 and 7 is used. A mix of negatively and positively worded items was used to mitigate response set bias.

The questionnaire was administered in Swedish at Bank and in English at Auto since the organization has offices in several countries and use English as the official corporate language.

Items were based on the scales used by Hoegl et al. [6] and Edmondson [7]. The measurement scale for inter-team coordination (Cronbach's alpha = 0,69) consists of four items based on the quality of coordination and operating characteristics between the teams (e.g. constructive discussions). The items used are described in Hoegl et al. [6] who partly adapted them from scales used by Mott [8] who evaluated coordination, communication, and cooperation between different occupational groups in hospitals.

For psychological safety (Cronbach's alpha = 0,67) the measurement scale consists of four items based on the measurement scale invented by Edmondson [7].

The measurement scale for team performance (Cronbach's alpha = 0,67) consists of three items described in Edmondson [7] who used the scale invented by Hackman [9] to obtain self-report measures of team performance.

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. For both inter-team coordination construct and the psychological safety construct, four items were used in the questionnaire but due to low communalities and low factor loading on one item in the team performance construct (item H3), only three items were used in the analysis.

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)
	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.

All questionnaire items contained a five-point answer scale (1=Not true, 5 = Very true).

3.4 Data analysis

All analyses are conducted at the individual level with survey responses from employees working as team members in development teams (N = 111).

Following Anderson and Gerbing's [10] 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 [11].

4 Results

4.1 Model estimation

Table 2 shows descriptive statistics for the scales. Skewness and kurtosis statistics indicate that most of the indicators are normally distributed with a few exceptions. Therefore, raw data was used instead of polychoric correlations which is better suited for ordinal data and not dependent on a normal distribution.

Table 2. Scale statistics.

Scale	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

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.

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 [13]. The test of discriminant validity is 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 [10].

Table 3 presents the measurement model statistics. The model fits the data well and All t-statistics for indicator loadings are significant, so nomological and convergent validity can be concluded. However, variance extracted for the constructs are not above 0.5 and composite reliability is not above 0.7, so the measured constructs are not fully reliable [13].

Table 3. 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: x2 = 59.98, df = 41, p = 0.028, RMSEA = 0.065						

In Table 4 displaying correlations between latent constructs, none of the confidence intervals around the standardized correlations between latent variables include one, so the model shows discriminant validity.

Table 4. 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 5 shows the result of the hypotheses tested in the study.

Table 5. Structural model statistics.

Hypothesis	Standardized loading	*t*-value	Error	Outcome
H1: Inter-team coordination -> Team performance	0.14	1.17	0.14	Not supported
H2: Psychological safety -> Team performance	0.60	4.17	0.21	Supported

H1 is not supported: in the large-scale agile settings where SAFe has been implemented, better inter-team coordination skills does not improve team performance.

H2, however, is supported: the stronger the individual psychological safety is, the greater the sense of team performance. The statistics of the structural model are also displayed in figure 2.

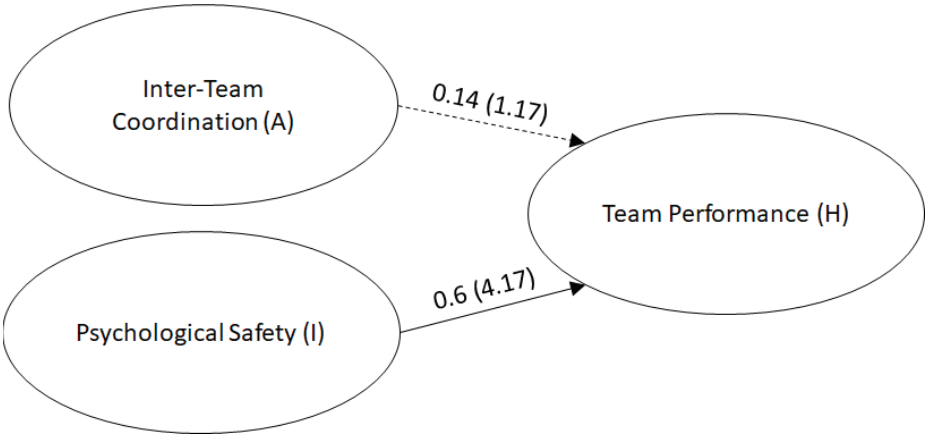

Fig. 2. Structural model.

5 Discussion

The results of this survey indicate that inter-team coordination in large-scale agile software development, using SAFe, does not improve team performance. Although this result is contrary to Hoegl et al. [6] 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. the individuals perceived performance of the team. However, although research shows a positive relationship between inter-team coordination and the overall project success (e. g. [14]), maybe inter-team coordination is not as important for the performance of the individual team. Maybe it should rather be seen as a teamwork skill on its own, not specifically important for team performance but only for project performance.

The managerial implication of this result is that inter-team coordination does not seem specifically important for team performance. Although originators of SAFe [6] claim increased productivity and more motivated employees, this study does not support that the claimed benefits stem from the added inter-team coordination practices.

The results do, however, support the second hypothesis, that psychological safety improves team performance in a large-scale agile software development setting, where the organization has implemented SAFe. Since hypothesis 1 is rejected and hypothesis 2 is supported, it also gives an answer to the third research question that psychological safety seems to have a stronger impact on team performance than inter-team coordination.

Regarding managerial implications, this means that large-scale agile software development organizations will benefit from helping teams to increase their level of psychological safety in order to raise the team performance. This confirms previous studies of the impact of high levels of psychological safety [7].

These results need to be further confirmed since the reliability of the constructs are questionable with low variance extracted as well as composite reliability. Further work needs to be done e.g. by investigating more organizations using SAFe since the number of participants in the study was rather low (N=111).

6 Overview of the Thesis Project

Although research on agile software development (ASD) projects has matured in the past years, a number of open questions remains. Agile in large scale settings in general [15] and inter-team coordination in particular [2] is seen as important issues for areas of further ASD project research.

The overall research question that the thesis project aims to answer is the following:

How is inter-team coordination conducted in large-scale agile software development projects and what are the impacts on teamwork?

The research is based on the project-as-practice approach to research which was originally promoted by the Scandinavian School [16, 17, 18]. The approach implies a focus on the micro-level practices and routines, attempting to understand how

practitioners (team members, scrum masters and project managers) act and make sense of their situation. Practice research focus on the actual activities performed by individuals and teams in projects, and how these align with or deviate from established norms, routines, and behavioral expectations [19]. Attention is focused on the ways in which activities are enacted and modified and negotiated by and among the actors involved. Further, it could also put light on how these enacted practices have an impact on individuals and teams. The perspective encourages researchers to conceptualize norms and ways of doing things as potentially dynamic, contextual and subject to change, rather than finding optimizations for "one model that fits all".

The project-as-practice approach has been an important foundation for the current debate on temporary organizing [20], focusing on how temporary structures and processes affect the way individuals are coordinated within and across organizations [21].

Three organizations working in large-scale agile settings are studied in the thesis project using both qualitative and quantitative methods for the past 1,5 years. During observations and interviews, the performance of inter-team coordination and psychological safety were addressed as important topics for further studies. Therefore, this particular study, presented in this paper, was conducted using a quantitative survey analysis to study the impacts on team performance in two of the organizations. The third organization was delimited from the study due to poor data quality in the survey responses.

References

1. 11th Annual State of Agile™ Report, http://stateofagile.versionone.com, last accessed 2018/06/11
2. Moe, N. B., Olsson, H. H., and 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)
3. Kirkman, B. L. & Rosen, B.: Beyond self-management: Antecedents and consequences of team empowerment. Academy of Management Journal, 42(1): 58-74, (1999)
4. 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)
5. Schwaber, K.: UnSAFe at any speed, http://kenschwaber.wordpress.com/2013/08/06/unsafe-at-any-speed/, last accessed 2018/06/02
6. Hoegl, M., Weinkauf, K., & Gemuenden, H. G.: Interteam coordination, project commitment, and teamwork in multiteam R&D projects: A longitudinal study. Organization science, 15(1), 38-55 (2004)
7. Edmondson, A.: Psychological safety and learning behavior in work teams. Administrative science quarterly, 44(2), 350-383 (1999)
8. Mott, P.E.: The Characteristics of Effective Organizations. Harper & Row, New York (1972)
9. Hackman, J.R.: Groups that Work (and Those That Don't). Jossey-Bass. San Fransisco (1990)
10. 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)

11. 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)
12. Jöreskog, K., Sörbom, D.: Lisrel 8: Structural Equation Modeling with the Simplis Command Language, Lawrence Erlbaum Associates, Hillsdale (1993)
13. 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)
14. 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).
15. Dingsøyr, T., Nerur, S., Balijepally, V., Moe, N. B.: A decade of agile methodologies: Towards explaining agile software development. Journal of Systems and Software, 85(6), 1213–1221 (2012)
16. Engwall, M.: No project is an island: linking projects to history and context. Res. Policy 32, 789–808 (2003)
17. Söderlund, J.: What project management really is about: alternative perspectives on the role and practice of project management. Int. J. Tech. Manag. 32 (3/4), 371–387 (2005)
18. Blomquist, T., Hällgren, M., Nilsson, A., Söderholm, A.: Project-as-practice: in search of project management research that matters. Project Manag. J. 41 (1), 5–16 (2010)
19. Manning, S.: Embedding projects in multiple contexts − a structuration perspective. Int. J. Project Manage. 26 (1), 30–37 (2008)
20. Bakker, R.M., DeFillippi, R.J., Schwab, A., Sydow, J.: Temporary organizing: promises, processes, problems. Organ. Stud. 37 (12), 1703–1719 (2016)
21. Hällgren, M., Söderholm, A.: Projects-as-practice: new approach, new insights. In: Morris, P., Pinto, J., Söderlund, J. (Eds.), The Oxford Handbook of Project Management. Oxford University Press, Oxford (2011)