DOCTORAL CONSORTIUM

A Quantitative Experiment: Inter-Team Retrospectives in Scaled Agile Programs

Daniel Toegl,[1] Tim Huygh,[2] Steven de Haes[3]

[1] University of Antwerp, Antwerp, Austria
daniel@toegl.at
[2] Open University of The Netherlands, Department of Information Science, Heerlen, Netherland
tim.huygh@ou.nl
[3] University of Antwerp, Department of Management Information Systems, Antwerp, Belgium
steven.dehaes@uantwerpen.be

Despite the current popularity agile project management and scaling such frameworks, there is little academic research inquiring how to scale agile frameworks to be successful at the program level. Retrospectives, are understood to play an important role in attempts to successfully adjust scaled-up project frameworks. This research studies the particular impact of such meetings as a tool to scale a known success factor from agile projects and address current gaps highlighted in recent program management literature. To better understand the impact of retrospectives, a quantitative experiment, to inquire how program value is accomplished with inter-team retrospectives, is conducted. With the proposed research design, we intend to test if there is a measurable impact on program success and efficiency by introducing inter-team retrospectives based on current approaches from the literature. This will allow to better understand and further develop the retrospective meeting design on an inter-team level within scale agile programs.

Keywords: program management, inter-team, intra-team, scaled agile, software development

University of Maribor Press

DOI https://doi.org/10.18690/um.fov.6.2023.56
ISBN 978-961-286-804-8

1 Introduction

Since the introduction of agile software development with Scrum (Schwaber and Sutherland, 2020), agile project management[1] is receiving increasing popularity. With growing numbers of applied use cases across industries, interest of scaling such frameworks is increasing to benefit from its advantages on a larger scale. However, there is currently little academic research on agile program and portfolio management to be found in the peer-reviewed literature.

In small agile projects, the development team coordinates work through frequent informal interaction among themselves and with customers. The Scrum framework has dedicated meetings for planning, review, and retrospectives. Many teams use visual boards, like in Kanban, to show who is working on what and the status of work tasks, which are designed for project size endeavors. Strode et al. (2012) explain coordination in such environments with agile teams and propose a model for coordination strategy and coordination effectiveness.

For large-scale projects, there is less support in the academic literature as described by Dingsoyr et al. (2018b). Scrum prescribes regular meetings between Scrum teams ("Scrum of Scrums") in order to manage the interfaces between teams, covering the planning of upcoming activities and dependencies, not retrospectives.

As we see from this current body of knowledge, focus is primarily on upfront planning activities at the beginning of each iteration (i.e. "Sprint Planning"). Agile however is consider as a series of cycles that aim and allow for continuous improvement for which reason the concept of a retrospective was introduced on a project level. Inter-team retrospectives are currently not part of commonly referred to scaled agile frameworks (Toegl et al., 2023), despite the original characterization from Schwaber and Sutherland (2020, page 9): "*The purpose of the Sprint Retrospective is to plan ways to increase quality and effectiveness*".

To close these identified gaps and extend the range of research methodologies for agile program management, a quantitative experiment is conducted to assess if an inter-team retrospective is a scalabale success/factor for agile program management.

[1] Several definitions of project, program, portfolio and the respective management can be found in management literature. In this review, we are referring to the denitions stated by Ireland (2002).

The result of this study would significantly support academics as well as practitioners with further developing scaled agile frameworks, their key artifacts as well as help organizations to become more successful in their agile transformation and Change initiatives.

2 Problem definition

From a business perspective, agile projects have been very successful across organizations in particular for software development. To further apply this methodology on a larger scale, agile project frameworks need to be adjusted in order to be successfully scaled-up and bank in its benefits.

While different Scaled Agile frameworks include different approaches on scheduled and unscheduled inter-team communication, the current literature investigates specific case studies as done by Paasivaara et al. (2012) or Dingsoyr et al. (2018b). Dingsoyr therefor suggest to further explore this relationship between intra-team and inter-team coordination in other approaches and settings other than his case study. From Toegl et al. (2023) we know that the academic literature covers success factors of scaled agile programs but does not investigate how to actually scale success factors known from agile projects, such as retrospectives.

Communication as well as knowledge sharing and improvement to enable learning at inter-team levels and program levels show under-researched fields (Dingsoyr et al., 2018a). Understanding the impact of retrospectives as a tool of knowledge sharing and improvement as well as for inter-team communication as well remain to be researched as described in the literature by Stettina and Schoemaker (2018) and Dingsoyr et al. (2018c). As highlighted by Dingsoyr et al. (2018c),the central challenge in coordination is identifying the right form or artifacts, arenas, and degree of formalization in large endeavors, such as scaled agile programs, with high complexity - a typical indicator to apply an agile framework as stated by Salameh (2014).

As the literature reveals, e.g. the systematic literature review by Dikert et al. (2016), agile program management is limited to experience reports and case studies of already completed programs and projects given the practicability of retrieving data.

Others applied approaches are limited to exploratory studies such as Dingsoyr et al. (2018c). Table 2.1 summarizes the questions for this study:

Table 2.1: Research questions

ID	Research question
PS1	What is the effect of inter-team retrospectives on program/business output within an agile setting?
PS2	What is the effet of inter-team retrospectives on program efficiency within an agile setting?
PS3	What are advantages and disadvantages of the different formats of inter-team retrospectives?
PS3.1	How do formats of inter-team retrospectives evolve over time and iterations?

To find answers to these questions, close the identified gaps and extend the range of research methodologies for agile program management, a quantitative experiment is conducted to assess if an inter-team retrospective is a scalabale success/factor for agile program management. The result of this study would significantly support academics as well as practitioners with further developing scaled agile frameworks, their key artifacts as well as help organizations to become more successful in their agile transformation and Change initiatives.

It is the contention that inter-team retrospectives in agile programs have a direct impact on program success and evaluated by efficiency, stakeholder satisfaction, and the success of meeting wider business goals, the relationship is also subject to other intervening, moderator impacts that must be considered as part of this study. We will follow the approach of Serrador and Pinto (2015) to achieve consistent results as shown by their work in this field.

As a result, the proposed research model is shown in Fig. 2.1, highlighting not only the tested relationship between inter-team communication in agile programs and program success, but also the potential impact of various moderators on this direct effect.

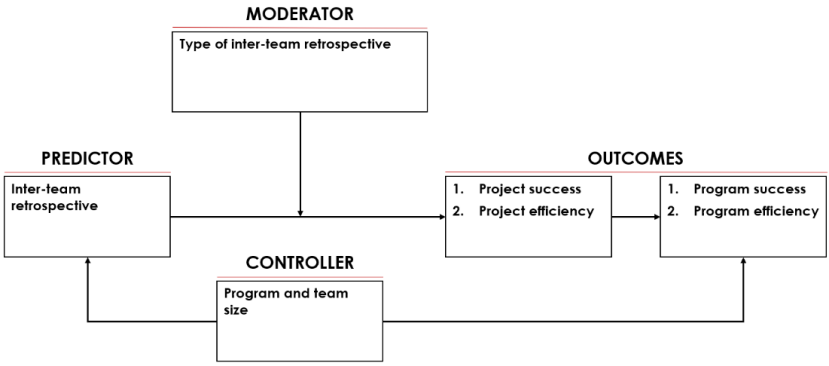

Figure 2.1: Research model

The type of inter-team retrospective type needs to be considered as a moderator because the time invested in such meetings has a significant impact on time spent and budget given the potentially high number of involved stakeholders and hence influencing project success and efficiency and subsequently program success and efficiency. However, in this research primarily focus on the impact created by the presence of such retrospectives. Table 3.3 provides an overview of retrospective formats considered in this study. The same assumption can be made on learning effects on managing defects to reduce error.

The central hypotheses (PS1, PS2) of this research are therefore about testing the effect of inter-team retrospectives as well as starting the investigation which format of inter-team retrospectives (Table 3.3) shows most effect. Table 2.2 provides an overview of the hypotheses and moderating effects that shall be investigated as part of this research.

Table 2.2: Hypotheses

	Hypotheses
HP1	Inter-team retrospectives have a positive impact on proram success.
HP2	Inter-team retrospectives have a positive impact on program efficiency.
HP3	The type/format of inter-team retrospective moderates the effect of retrospectives on program success.
HP4	The type/format of inter-team retrospective moderates the effect of retrospectives on program efficiency.

As the impact of the inter-team retrospective is expected to become more visible over time, meaning after a couple iterations, we want to assess the impact of the inter-team retrospective as a moderator. Following the approach of Treiblmaier and Putz (2020) in which a comparable multi-group field experiment was facilitated, we model inter-team retrospectives as a moderator between the program and project success as well as project efficiency and therefore program

success and efficiency and hypothesize for the statistical testing:

Table 2.3: Null hypotheses for testing program success

	Null hypotheses
H0	There is no effect on program success in an environment with inter-team retrospectives.
H1	There is an effect on program success in an environment with inter-team retrospectives (of format A or B).

Table 2.4: Null hypotheses for testing program efficiency

	Null hypotheses
H0	There is no effect on efficiency in an environment with inter-team retrospectives.
H1	There is an effect on efficiency in an environment with inter-team retrospectives (of format A or B).

3 Methodology, design and focus

We aim to research the impact of inter-team retrospectives within a simulated setting and compare it to the same setting without inter-team retrospectives. We consider program management success as the unit of analysis, following the approach of Shao et al. (2012).

A post-positivism perspective with a deductive approach is used in this study in order to operationalize the program success and program context concepts - broadly following Shao et al. (2012). An experiment is used to collect data which then allows for generalizable results within the design's boundaries. The applied research design is therefore considered to be Experimental Research.

Program business success shall be the output of the project team. Program efficiency shall be considered as reduction of error (defects). Stakeholder satisfaction to satisfy the expectations of project stakeholders is not considered to simplify the experiment.

We gather qualitative data from the participants in regards to PS3 via their results from the retrospectives to better understand how the adjusted intra-team as well as the inter-team retrospectives were perceived. To analyze data from these retrospectives, the success factors described by Hummel and Epp (2015) are applied for potential coding purposes.

3.1 Iterative research approach and Pre-studies

Following an agile approach, two pre-studies were conducted with graduate students from the University of Applied Sciences Joanneum (Graz, Austria) in autumn 2021 and 2022 to confirm the feasibility of the operating the experiment. The goal of these pre-studies was to test assumptions made and to test the operationalization of the experiment in terms of its complexity.

Findings of these pre-studies impacted the ratio of required Scrum Masters to teams (1:1), highlighted the need to play different games within a program to avoid competition (i.e. withholding information as well as mistrust between teams). In regard to the number of iterations conducted during the experiments, we raised the number of iterations..

3.2 Procedure and data collection

Data is collected during the experiment, as described in Table 3.1 and Table 3.2, by the Scrum Master typically every sprint or are calculated based on values provided from each project.

Table 3.1: Measures on program level

Dimension	Calculation of scores
Total program performance	Sum of each team's actual points delivered across all sprints.
Total program defects	Sum of each team's actual points delivered across all sprints.
Total program efficiency	Ratio of the total program performance to the total program defects.
Program sprint performance	Sum of each team's actual points delivered in a sprint.
Program sprint defects	Sum of each team's defects in a sprint.
Program sprint efficiency.	Ratio of the program sprint performance to the program sprint defects.

Table 3.2: Measures on project level

Dimension	Calculation of scores
Sprint performance	A team's actual points delivered in a sprint.
Sprint defects	Defects that occurred during a team's sprint.

Further calculations of each team's total performance/efficiency will be conducted. Inter-team retrospectives are considered as indicator variables (1 = took place; 0 = did not take place) for each program, team and sprint.

In this experiment, we are simulating programs following the Scrum of Scrum framework (Paasivaara et al., 2012), which is chosen given its lowest level of complexity once the concept of Scrum is introduced. We have multiple Scrum teams with a Scrum Master each, who will be responsible to collect the data. Other data is calculated based on their output.

3.3 Structure of the experiment

The experiment aims to simulate scale agile programs by playing agile games. To ensure comparable results, every program follows the same strategic goals.

The Ball Flow Game was chosen as a feasible game for the Scrum teams (Fowler et al., 2001) in the experiment given the comparison by Przybylek and Olszewski (2016) in which it scored high in the considered factors, such as easy-to-understand. In addition, its scoring mechanism in every iteration of the game made it a suitable candidate for this research.

The class-room activities described by Hurbungs and Nagowah (2019) Paper Planes (Heintz) and Pizza Game (Agile42) were chosen for the Scrum teams due to their suitability in a tertiary educational context as well as their objective to agile learning and teaching. In addition, we can apply the same scoring mechanism in every iteration of the games.

For this experiment, we time-box a complete sprint/iteration to 15 minutes. During these 15 minutes, every team has to complete all Scrum ceremonies. The actual play time of each game is defined as 3 minutes.

The Scrum Master is responsible to facilitate the game, stop the time and collect the results.

Adoption of Scrum-of-Scrum questions We can adapt questions used by Paasivaara et al. (2012) in the Daily Scrum to the needs of the retrospective, along with (Dingsoyr et al., 2018a) on the category "other teams to facilitate "double loop" learning from the framework of Argyris (1976).

Participants Participants are university students with no to little prior knowledge about agile project management. To further understand their knowledge, we perform a survey at the beginning of the experiment and dismiss students that have prior experience with the selected agile games.

As part of the experiment, a set of roles is required that follow the Scrum (Fowler et al., 2001) and Scrum of Scrum frameworks (Paasivaara et al., 2012). In this set-up, Release Manager and Product Manager can be neglected. The role of the Product Owner is replaced with a rule set and guidelines for each sprint. This information is available to the Scrum Masters from the beginning and during each iteration will be revealed to the project team. As part of this the Scrum Masters will receive separate

information and is instructed to perform additional tasks, such as disrupting the process with adding three balls during a sprint as part of the Ball Point Game.

3.4 Types/formats of inter-team retrospectives

The different types or formats of inter-team retrospectives considered in this research are described in Table 3.3. Inter-team retrospectives are planned only for teams or representatives from the same program and not across multiple programs.

Table 3.3: Considered formats of inter-team retrospectives

Type	Short description
A	Inter-team retrospectives with by the team selected members of each Scrum team.
B	Inter-team retrospectives between the Scrum Masters of each Scrum team.
C	Inter-team retrospectives with all Scrum teams and all their members.
D	No inter-team retrospective.

We plan for team retrospectives after every sprint/iteration. Team retrospectives are planned once every team completed its team retrospective.

Type C is only mentioned for completeness and are not considered suitable for any real-world program. It is therefore not further considered in this experiment.

Sequence of sprints/iterations For this experiment, we plan to have nine sprints/iterations for each team in every program executed in parallel. Data is collected before, during and after each sprint/iteration.

3.5 Data collection approach

Data is collected to collect each team's output and defects and therefore its contribution to the program. We expect, this allows to understand the performance on a program as well as project level.

In general, the simplest relationships are examined firsthand then analysis continued using progressively more involved techniques. This includes a path-analysis of each scrum team over time.

Other potential moderators such as Product Vision/Goals, Project Complexity and Team Experience are not further considered as these factors are harmonized in all our programs within the given experimental design.

3.6 Statistical evaluation

The experiment provides a purposeful sample size with 30 inter-team retrospectives (10 for each program) with a total of 90 sprints (30 for each program), which is deemed sufficient for the statistical analysis with a regular t-test, as described by De Winter (2013). If required, the experiment can be scaled up include more teams in each program, however saturation is expected by 10 sprints. If we cannot confirm normal distribution with the gained data, the Mann-Whitney-U test can be used.

The Jonckheere's trend test could additionally be used to detect differences in the teams across the multiple sprints. In particular, we would expect to see a trend when comparing the project teams with inter-team retrospectives to the comparison project teams without inter-team retrospective over time.

3.7 Data quality

The data will be collected every sprint from the Scrum Masters and is limited to only very few input points. Photos of the retrospective boards with their sticky notes are taken in every sprint to collect the required data, which will be particularly relevant regarding PS3. Photos allow to track and validate not only the raised points, but additionally allow to count how often certain topics are raised within a retrospective. A retrospective board also allows to visually see where the majority of points are raised which allows to easily identify if things went well or not. As described by Schulze (2007), taking photos forces the participants to reect, while gaining insights on what is most significant. This simplicity is expected to provide sufficient meaningful data for every Scrum team in every iteration, compared to other media such as, e.g. video.

To ensure data is properly collected, the Scrum Masters will receive an introductory training in advance. All Scrum Masters will be PSM1 certified.

3.8 Validity and reliability

Following the structure and results of Brink (1993), the four threats (the researcher, the subjects participating, the situation and the methods of data collection and analysis) are addressed.

To reduce bias of the researcher by participating as well as the risk of reducing bias with data collection, the Scrum Masters will be responsible for data collection. They will receive an introductory training in advance, together with a retrospective guide and a fieldnote template that covers the required structure to cover and collect data of the required key success factors. Such a preparation is considered especially necessary to ensure the researcher is trusted by the Scrum Masters (Leininger and Reynolds, 1991).

Data will be collected every sprint and is limited to only very few input points, to ensure the amount of work is minimized.

Given the simplicity of the game and time constraints, the depth and variety of findings in retrospective meetings are expected to diverge little among teams when comparing to real-world large scale programs.

3.9 Ethics

Approval of the Ethical board of the University of Antwerp is expected before conducting the experiment.

The idea is to perform this primarily with adults, that are monetarily incentivized (e.g. EUR 1 / point) for each successfully delivered point.

4 Preliminary/Expected results

In terms of reserach methodology, researching inter-team retrospective with an experiment expands the toolbox for conducting research in the eld of agile program management.

Regarding the results of the experiment, the initial contention is that the simulated programs with inter-team retrospectives are more successful than those neglecting this ceremony. We expect that the null hypothesis described in Table 2.4 and in Table 2.3 are rejected. This means, a possible outcome could be, that we can show a statistically significant impact of inter-team retrospectives in terms of the business outcomes (program success and program efficiency), while identifying a better performing retrospective format. We additionally expect to gain knowledge how intra-team and inter-team retrospectives evolve over time. We expect to further gain knowledge on who become ambassadors to represent teams at inter-team retrospectives and how the team members utilize inter-team retrospectives to their project's benefit.

With this result, we believe found a success factor known from agile project management, that can be scaled to agile program management.

These results will allow us to undergo real-world tests with this format with lower managerial risks as the proposed concept resulted from thorough academic research. This knowledge could impact scaled agile frameworks in a way, that inter-team retrospectives become widely adopted.

Expected outcomes Type A and B retrospectives are expected to be the types with the most positive effect among the researched types. Additionally, as the iterations continue and changes in the set-up might be suggested by the intra-team as well as as the inter-team retrospectives, type A retrospectives might evolve into a type B retrospective, or alternatively, type B retrospectives might evolve into a type A retrospective.

5 Future development

This research is part of a broader research question part of the author's PhD. Tying three research studies together, a possible combined outcome could be, that we contributed to the body of knowledge in multiple ways and lay out a way on how to improve success of scaled agile programs.

Firstly, we identified an under-researched academic eld with high interest for practitioners and prepared an overview that supports further academic research more easily. The results of a systematic literature revealed that retrospectives are a known success factor in agile projects but academia barely researched on how to scale up this success factor to scaled agile programs.

The here described experiment is planned as the second paper to solve the "managerial puzzle" described in the author's PhD project.

Building on the results from this research paper, we want to gain in-depth knowledge on the perceived value of inter-team retrospectives, their connection to intra-team retrospectives and their evolution over time at a software service providing company, Parkside Interactive, in a third study. The best performing format of inter-team retrospective from the experiment will be taken as a starting point for this case study.

With these steps, we additionally close a number of previously identified research gaps and for the fist time analyze agile programs with a quantitative experiment.

References

Agile42. Pizza game. URL https://www.agile42.com/en/agile-teams/kanban-pizza-game.

C. Argyris. Single-loop and double-loop models in research on decision making. Administrative science quarterly, pages 363-375, 1976.

H. I. Brink. Validity and reliability in qualitative research. Curationis, 16(2):35-38, 1993.

J. C. De Winter. Using the student's t-test with extremely small sample sizes. Practical Assessment, Research, and Evaluation, 18(1):10, 2013.

K. Dikert, M. Paasivaara, and C. Lassenius. Challenges and success factors for large-scale agile transformations: A systematic literature review. Journal of Systems and Software, 119:87-108, 2016.

T. Dingsoyr, M. Mikalsen, A. Solem, and K. Vestues. Learning in the large - an exploratory study of retrospectives in large-scale agile development. In International Conference on Agile Software Development, pages 191-198. Springer, Cham, 2018a.

T. Dingsoyr, N. B. Moe, T. E. Faegri, and E. A. Seim. Exploring software development at the very large-scale: a revelatory case study and research agenda for agile method adaptation. Empirical Software Engineering, 23(1):490-520, 2018b.

T. Dingsoyr, N. B. Moe, and E. A. Seim. Coordinating knowledge work in multiteam programs: Findings from a large-scale agile development program. Project Management Journal, 49(6): 64-77, 2018c.

M. Fowler, J. Highsmith, et al. The agile manifesto. Software Development, 9(8):28-35, 2001.

J. Heintz. Agile airplane game. URL http://gistlabs.com/2011/06/agile-airplane-game/.

M. Hummel and A. Epp. Success factors of agile information systems development: A qualitative study. In 2015 48th Hawaii International Conference on System Sciences, pages 5045-5054.

IEEE, 2015.
V. Hurbungs and S. D. Nagowah. A practical approach to teaching agile methodologies and principles at tertiary level using student-centred activities. Agile and Lean Concepts for Teaching and Learning: Bringing Methodologies from Industry to the Classroom, pages 355-389, 2019.
L. R. Ireland. PMI Lexicon of Project Management Terms. Project Management Institute, Inc., 2002.
M. Leininger and C. Reynolds. Culture care diversity and universality theory. Nursing, 1991.
M. Paasivaara, C. Lassenius, and V. T. Heikkila. Inter-team coordination in large-scale globally distributed scrum: Do scrum-of-scrums really work? In Proceedings of the ACM-IEEE international symposium on Empirical software engineering and measurement, pages 235-238, 2012.
A. Przybylek and M. K. Olszewski. Adopting collaborative games into open kanban. In 2016 Federated Conference on Computer Science and Information Systems (FedCSIS), pages 1539-1543. IEEE, 2016
H. Salameh. What, when, why, and how? a comparison between agile project management and traditional project management methods. International Journal of Business and Management Review, 2(5):52-74, 2014.
S. Schulze. The usefulness of reflexive photography for qualitative research: A case study in higher education. South African journal of higher education, 21(5):536-553, 2007. 14
K. Schwaber and J. Sutherland. The scrum guide. Scrum Alliance, 2020.
P. Serrador and J. K. Pinto. Does agile work? A quantitative analysis of agile project success.
International Journal of Project Management, 33(5):1040-1051, 2015.
J. Shao, R. Mueller, and J. R. Turner. Measuring program success. Project Management Journal, 43(1):37-49, 2012.
C. J. Stettina and L. Schoemaker. Reporting in agile portfolio management: routines, metrics and artefacts to maintain an effective oversight. In International Conference on Agile Software
Development, pages 199215. Springer, Cham, 2018.
D. E. Strode, S. L. Huff, B. Hope, and S. Link. Coordination in co-located agile software development projects. Journal of Systems and Software, 85(6):1222-1238, 2012.
D. Toegl, T. Huygh, and S. De Haes. Retrospectives in scaled agile: A systematic literature review. Working paper, 2023.
H. Treiblmaier and L.-M. Putz. Gamification as a moderator for the impact of intrinsic motivation: Findings from a multigroup field experiment. Learning and Motivation, 71:101655, 2020.