Author(s): Korpivaara, Ida; Tuunanen, Tuure; Seppänen, Ville

Title: Performance Measurement in Scaled Agile Organizations

Year: 2021

Version: Published version

Please cite the original version:

Korpivaara, I., Tuunanen, T., & Seppänen, V. (2021). Performance Measurement in Scaled Agile Organizations. In Proceedings of the 54th Hawaii International Conference on System Sciences (HICSS 2021) (pp. 6912-6921). University of Hawai'i at Manoa. Proceedings of the Annual Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2021.830

Performance Measurement in Scaled Agile Organizations

Ida Korpivaara
University of Jyväskylä
ida.e.korpivaara@student.jyu.fi

Tuure Tuunanen
University of Jyväskylä
tuure@tuunanen.fi

Ville Seppänen
University of Jyväskylä
ville.r.seppanen@jyu.fi

Abstract

Performance metrics are a key tool for organizations to direct and motivate their members toward desired outcomes. Despite their central role also in Agile development, little is known about how Agile organizations set performance objectives and metrics in practice and balance between business and software engineering goals. In particular, there is a knowledge gap regarding performance measurement and its challenges in scaled Agile organizations. This paper applies an exploratory case study method to examine performance objectives and metrics in two business units of a scaled Agile organization and suggests a framework for selecting and structuring performance objectives within them. We offer five performance measurement dimensions that can be applied to provide guidance for scaled Agile organizations in selecting and prioritizing performance objectives.

1. Introduction

Performance objectives and metrics are essential tools for organizations to direct and motivate their members toward desired outcomes. Metrics steer activities, inform management on progress, and enable improvement, which makes them a particularly valuable tool for Agile software development organizations that continuously seek opportunities to learn and enhance performance. Prior research has found that high-performing organizations use performance metrics more actively than their peers [1]. In Agile organizations, performance measurement can also be used as a tool to support Agile practices and behavior [2].

Despite the importance of metrics, existing studies on performance measurement in Agile organizations are scarce. Although research has been conducted on metrics in specific areas such as software [3, 4], processes [5], and maturity [6, 7], few studies have addressed performance objectives and their measurement on the organizational level beyond Agile teams. Furthermore, little is known on how organizations combine and align the different measurement areas. Holistic understanding, however, is of growing importance, as Agile methods are being increasingly adopted at scale. In the 14th State of Agile survey, 66% of respondents represented software organizations having more than 100 employees [8].

This article aims to fill the research gap by investigating how scaled Agile development organizations measure performance on different organizational levels. The study focuses on SAFe framework for scaling Agile and seeks to answer the following research questions:

(1) How do scaled Agile organizations determine and measure performance objectives on different organizational levels?

(2) What challenges are associated with performance measurement in scaled Agile organizations?

Traditionally, software organizations evaluate performance based on individual projects. Due to the iterative nature of Agile development, however, performance evaluation in Agile organizations cannot be based on projects to the same extent as in traditional software development. By reviewing performance measurement approaches suggested by prior literature, this study identifies five dimensions to determine performance objectives in scaled Agile organizations: customer value, financial value, internal process efficiency, collaboration, and learning. The study's empirical findings provide support for these dimensions but also show that their prioritization differs between organizational levels, as internal process efficiency, collaboration, and learning are only measured within Agile teams and release trains (ARTs). Based on these findings, the study suggests a framework for selecting and structuring performance objectives in scaled Agile organizations.

As its research method, the study employed an exploratory case study design because of its suitability for studies focusing on "how" aspects [9]. The primary data consist of ten semi-structured interviews

HłCSS

conducted in two scaled Agile development units of a large Nordic bank. The interviews were analyzed using qualitative content analysis guided by findings from previous studies. To increase the validity of the results, findings from the interviews were triangulated by using documents as a secondary data source.

The article is structured as follows. The following section discusses approaches for determining performance objectives and metrics in Agile organizations suggested by extant literature. Next, the case organizations and the research design are presented, followed by the case study findings. Finally, the study's findings and implications are discussed before the concluding remarks.

2. Performance Metrics in Agile Software Organizations

The ability to measure performance is crucial for any organization aiming to learn and improve. Since software development traditionally is structured in projects, performance measurement has focused on evaluating individual projects against their success criteria. As with projects in general, key objectives of software projects are often described through three dimensions: time restrictions, budget restrictions, and fulfillment of project scope. While criticized for only considering internal success instead of external measures such as customer satisfaction, these key objectives, also known as the "Iron Triangle," have been widely adopted [10].

Project-based traditional software metrics have been found to support Agile development poorly [11]. When moving from plan-driven traditional to value-driven Agile software development, organizations are therefore required to shift mindset also regarding their performance metrics. Instead of measuring the realization of pre-defined plans, Agile organizations need to understand how they create value and measure performance accordingly. Prior literature has suggested three approaches for understanding value creation and deriving objectives in Agile context: strategic business objectives [12, 13], stakeholder needs [14, 15], and Agile principles [16, 17].

2.1. Strategic Business Objectives and Software Value Map

One of the most comprehensive descriptions of value created by software development organizations is the Software Value Map introduced by Khurum et al. [12]. The value map builds on the Balanced Scorecard developed by Kaplan and Norton [13], which aims to provide a tool for organizations "to align business

activities to the vision and strategy of the business" and "monitor business performance against strategic goals." While focusing on IT context, organizational business strategy and vision are also fundamental building blocks for the Software Value Map.

To ensure a broad view of organizational health, the Software Value Map suggests four perspectives for organizational performance: *customer, financial, internal business*, and *innovation and learning*. The *customer* perspective is defined as the "capability to develop and deliver a product that satisfies customer requirements;" [12] it considers both the value perceived and the value obtained by a customer when acquiring a given product. The *financial* perspective focuses on the business owner's perspective and strategies considered to improve the bottom line. Despite its importance for business owners, Khurum et al. [12] suggest that financial objectives often shift focus from long-term investments to short-term goals. The *internal business* perspective focuses on value aspects related to maintaining the competitiveness and quality of the development base, such as production time and cost. Finally, the *innovation and learning* perspective considers the organization's intangible assets, focusing on the skills and capabilities needed to support the organizational value chain in the present as well as in the future [12].

Although the Software Value Map was developed for the software development context, it does not consider Agile methods in particular. In their study of the interpretation of value in Agile organizations, Alahyari et al. [18] examined the Software Value Map in an Agile context. They identified customer value as the most prioritized perspective and delivery time and quality as the most important value constructs. Internal business processes were identified as another priority dimension for Agile organizations, while innovation and learning and financial value received almost no attention in the study.

2.2. Stakeholder-driven Measurement

Despite being widely used, strategic business objectives as a base for measuring performance have been criticized for only considering owners and customers as organizational stakeholders. This has given impetus to the emergence of stakeholder-driven performance measurement [14, 15, 19, 20].

The stakeholder-driven approach argues that best performance is achieved when all stakeholder groups' needs and objectives have been met. While all stakeholder groups may not be equally relevant in all contexts, Oza and Korkala [19] perceive a stakeholder-driven approach as a balanced method that considers different viewpoints and provides

therefore a comprehensive foundation for determining performance objectives. List et al. [15] also highlight the ownership aspect of the stakeholder-driven measurement: once every indicator has been assigned to a stakeholder group with interest in it, poor performance will always be noted and acted upon.

In their application of stakeholder-driven performance measurement in software development, List et al. [15] consider four principal stakeholder groups: investors, employees, customers, and society. While the objectives of customers and investors, are largely aligned with customer and financial value in the Software Value Map, the stakeholder groups of employees and society distinguish the approach from the Balanced Scorecard. Objectives within perspectives of employees and the society may include goals such as job satisfaction or regulatory requirements. However, as List et al. [15] note, there is no universal stakeholder list, but rather the list should be adjusted for each context.

Even if stakeholder-driven performance measurement better captures the value delivered to employees and society than the Software Value Map, it is perceived to be limited by its focus on the present at the expense of the future. It does not consider such performance dimensions as learning or improvement. To mitigate this shortcoming, List et al. [15] suggest adding innovation as a fifth key aspect in the model.

2.3. Measurement of Agile Principles

Instead of applying general performance measurement approaches to the Agile context, previous studies have suggested that performance dimensions for Agile organizations can be derived directly from the Agile Manifesto [16, 17, 21, 22]. This approach has received support, particularly from industry practitioners, while academic studies on the field are scarce.

Prior literature has suggested multiple ways of translating Agile values and principles into performance metrics. In their study of Agile transformations, Olszewska et al. [16] suggest four key metric areas as indicators of Agile success: responsiveness to change, throughput, workflow distribution, and quality. Similarly, Davis [17] distinguish between four dimensions: effective processes, software, requirements, and development teams. Dubinsky et al. [21] propose measuring completed work, frequency of releases, burndown of realized vs. committed work, and number of faults to ensure continuous high-quality deliveries. Despite their differences, the importance of responsiveness to change, productivity, quality, and continuous deliveries as performance objectives are emphasized in all suggested metric frameworks.

Instead of measuring the realization of benefits outlined by the Agile Manifesto, prior studies have suggested that organizations may also benefit from measuring the extent to which they are following Agile practices. This concept, known as Agile maturity, focuses on capabilities rather than outcomes. In their study of Agile maturity, Fontana et al. [7] emphasize the role of subjective capabilities for Agile organizations, since Agile practices, unlike many other development methods, are based on shared principles rather than strictly defined processes. Fontana et al. [7] suggest defining Agile maturity in terms of eight team practices: collaboration, care for the customer, acceptance of requirement changes, knowledge-sharing, use of Agile tools, self-organization, continuous improvement, and generation of perceived outcomes.

Although metrics based on Agile principles and maturity models may address special characteristics of Agile practices better than general performance models, they often focus solely on software development activity. Furthermore, they measure mostly internal delivery efficiency instead of externally generated outcomes. Few studies have shed light on how Agile organizations combine and prioritize internally oriented metrics based on Agile principles with those that measure external outcomes.

2.4. Overview of Performance Metrics

To form an overview of the different measurement approaches, performance dimensions highlighted by them were summarized as presented by Table 1.

Table 1. Performance dimensions for Agile development organizations

Dimension	Examples of objectives	Example studies
Financial value	Return on investment	Khurum et al. [12]; Alahyari et al. [14]
	Production cost	
Customer value	Customer satisfaction	Alahyari et al. [14]; Mahnic & Zabkar [16]; Neely et al. [18]; Fontana et al. [7]
	Product quality	
	Product usability	
	Delivery time	
	Delivery quality	
Internal process efficiency	Productivity	Khurum et al. [12]; Alahyari et al. [14] Olszewska et al. [19]; Davis [20]; Dubinsky et al. [21]
	Predictability	
	Production quality	
	Responsiveness to change	
Collaboration and culture	Employee satisfaction	Mahnic & Zabkar [16]; List et al. [17]; Fontana et al. [7]; Neely et al. [18]
	Collaboration	
	Acceptance of requirement changes	
	Care for the customer	
	Use of Agile tools	
	Self-organization	
Innovation and learning	Capability development	Khurum et al. [12]; Fontana et al. [7]
	Continuous improvement	

Overall, there were identified five performance dimensions: financial value, customer value, internal process efficiency, collaboration and culture, and innovation and learning. While these included all perspectives from the Software Value Map, performance areas suggested by stakeholder-driven and Agile principles-based approaches were added only if they were not covered by the existing dimensions. The five dimensions were used as input for designing the interview questions in the study.

3. Research Design

This study approached the research problem through an exploratory case study. The method was chosen for its suitability for studying real-life contexts and its ability to provide in-depth insights. Furthermore, Yin [9] suggests that the case study method supports well research focusing on "how" aspects.

The case units chosen for the study were two software development units of a large Nordic bank. The units were perceived to fit the study well, as they provided typical examples of international scaled Agile organizations. One unit creates digital banking solutions for personal customers and the other one for businesses ranging from entrepreneurs to large institutions. Combined, the units have over 1,000 employees located around the Nordic countries and in Poland and India. While both units formally use the scaled Agile framework (SAFe), the extent to which the framework was followed differed between departments and was flexible particularly in the unit focusing on corporate customers. Apart from software developers, the units had adopted Agile methods only two years ago and hence still considered themselves to be in transition with their measurement practices.

Following SAFe, the units were structured in three levels: team, program, and portfolio. On the lowest level, development took place in Agile teams. On the program level, the teams were organized in ARTs that shared a common purpose and development domain. On the portfolio level, the ARTs were further grouped into hubs based on the customer need that they focused on, but these structures were more of a strategic nature than part of SAFe. Parallel to the SAFe structure, the units had a line organization designed to manage people that was structured according to capabilities.

The data were primarily collected through semi-structured interviews. Purposive sampling [23] was used to ensure an information-rich and diverse sample. Interview participants were primarily selected based on their expertise on performance metrics, without compromising a balance between Agile roles, organizational levels, genders, and nationalities

among the participants to capture different perspectives. Altogether, ten people were interviewed for the study, as shown in Table 2. All interviews were conducted in English individually, recorded, and transcribed for the purpose of analysis. On average, the interviews lasted 45 minutes. Performance dimensions from the literature review (Table 1) were used to guide the interview structure and questions.

Table 2. Interview participants by role

	Case unit 1	Case unit 2
Portfolio-level Manager	1	1
Operational-level Manager	1	1
Business Developer	1	1
Agile Coach	1	
Product Owner		1
Software Developer		1
Subject-matter Expert	1	
Total	**5**	**5**

Data analysis was carried out following the six-step process for thematic analyses presented by Braun and Clarke [24]. The process started by reading through the interview transcripts and writing initial ideas down. Next, all quotes of interest were systematically transformed into initial named codes, which were then grouped into code categories or themes. Before refining and finalizing, the themes were checked against the text extracts to make sure the interpretation fitted the data. Atlas.ti software for qualitative analyses was used to assist in the coding. While the analysis was primarily carried out by the first author, the emerging themes and categories were regularly reviewed and discussed collectively.

According to Singleton et al. [25], the weaknesses of single methodological choices and measures can be avoided by using multiple data sources, also known as triangulation. In this study, this was done by using documents as a secondary data source to provide details on metrics described in the interviews. The documents consisted primarily of performance reports and presentations to executive decision boards.

4. Findings

To understand performance measurement in the case units holistically, the use of performance objectives and metrics was examined separately for unit- and program-level metrics. The interview participants were also asked to describe the use and interplay of the different metrics. The findings are presented below.

4.1. Metrics on the Portfolio Level

The most important performance objective for both case units was customer satisfaction. As described by

one of the interviewees, *"we measure our performance on if we are able to provide digital solutions that make our customers' life easier."* Customer satisfaction was perceived both as an end goal as well as a lead indicator for income generation, since satisfied customers were assumed to have higher future engagement and to attract more new customers than unsatisfied customers. The emphasis on customer satisfaction is aligned with prior findings by Alahyari et al. [18] that identified customer value as the single most important performance dimension for Agile organizations.

On the unit level, customer satisfaction was measured through two main key performance indicators (KPIs): Net Promoter Score (NPS) and Customer Experience Index (CEI). While the use of NPS and CEI was a company-wide standard, the customer base that they were measured on varied between organizational levels. On the unit level, the case units measured the indices for the service channels that they were responsible for.

Financial value was another key objective on the unit level. On income, the main KPIs included generated income, turnover, and sales and were measured for the service channels that the units were in charge of. For cost, the objectives were two-fold. First, the units had their cost budgets to adhere to. Second, there was a strong focus on enabling cost savings and efficiencies elsewhere in the organization. The units might, for example, aim to digitalize a workflow that previously required physical meetings or to automate a manual report to save postal expenses and free up advisory time. Since the time and resources saved by these activities could either be converted into cost savings or used for new sales, the activities were described as enablers rather than cost savings. The activities were also considered to have an impact on income generation through being a competitive advantage and had therefore a link to the internal process value described by Khurum et al. [12]. The metrics used for these initiatives varied from activity to activity and were not necessarily aggregated on a unit level as a KPI.

Tightly linked to customer satisfaction and financial value, IT stability formed another important performance objective. IT stability was perceived to signal overall quality in product development and implementation. Another motivation for its selection was its impact on customer satisfaction and financial value: *"If we have downtime all the time and people are not able to use the channel, we don't earn any money."* IT stability was measured by the number of reported incidents and application downtime.

The units also had objectives related to the growth of their user base and the services offered. These objectives, however, were not necessarily perceived as end goals but rather as supporting metrics. Even if there was no formalized categorization, a distinction was often made between primary and secondary objectives. Primary objectives focused on outcomes and were measured by reactive or lagging indicators, while the secondary objectives emphasized future performance potential. As an example, income generated by a specific digital platform might be set as a primary KPI, whereas the net change in active users could be used as a supporting metric to ensure that development was on the right track. One interviewee observed: *"Overarching should be the value creation. The others are just indicators on our ability to do that.*"

The unit-level objectives were mainly determined by the company's overall strategic priorities. Strategic business objectives were regarded as a primary motivation for measuring performance, in line with the findings of Khurum et al. [12]. Apart from the group priorities, market factors such as changing customer requirements, competitors, and regulation influenced the selection of unit-level performance objectives.

Interestingly, Agile development methods were not perceived to have any role in setting the units' performance objectives. While the SAFe framework included suggestions for metrics to maximize realized benefits from its use, these metrics were not used as unit KPIs. This was explained by the primary role of unit KPIs as a tool to communicate contribution toward group objectives.

4.2. Metrics on the Program and Team Levels

Within the case units, organizational structure had a major impact on performance measurement. As described in section 3, execution organization was separated from line organization in both units, and the two dimensions were evaluated on separate metrics. In addition, performance follow-up within the execution organization differed between hubs and ARTs. The overall structure and the main performance metrics are presented in Figure 1.
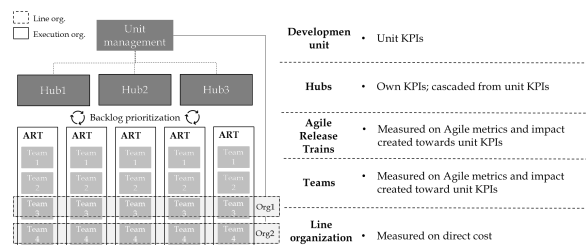

Figure 1. Measurement structure in case units

The highest level of the execution organization, hubs, had performance goals that were directly cascaded from the unit objectives. If a hub focused, for example, on financing housing, it was measured on customer satisfaction with housing loans, income from housing loans, and so on. The key purpose of the hubs was to contribute to the unit KPIs and therefore catalyze strategy implementation.

Within the hubs, the focus in performance evaluation shifted from the "what" aspect toward "how." Since the backlog prioritizations process ensured that ARTs were working on deliveries that created the most value to the unit objectives, there was no need to measure ARTs in terms of delivered impact. Instead, their performance evaluation focused on efficiency. This was done using so-called Agile metrics recommended or inspired by the SAFe framework. Rather than merely becoming faster at execution, the Agile metrics were designed to support the organization in maturing and optimizing its ways of working in a holistic manner. Being designed to support Agile principles, the metrics aligned well with other metric models based on Agile principles suggested by prior studies [16].

The Agile metrics consisted of seven performance objectives: productivity, time to market, quality, continuous improvement, customer satisfaction, employee engagement, and alignment to strategic objectives. *Productivity* aimed to reduce feature cycle time and was primarily measured by the number of completed features in a given time, also known as throughput. Closely related, predictability was important for optimizing planning and was followed up by the burndown ratio, meaning the share of completed work against committed work. *Time to market* focused on the frequency of releases and was measured by feature lead time, covering both time from idea to execution and that from execution to implementation. *Quality in development* was measured by the number of defects and the test automation ratio, while *quality in production* was evaluated through the number of business-critical incidents and the mean time to restore service.

Complementing the quantitative metrics, trains and teams also had several qualitative Agile metrics. *Customer satisfaction* was measured by ratings and satisfaction scores on the digital features that the ARTs developed. *Employee engagement* was followed up through a group-wide quarterly satisfaction survey complemented with additional questions on empowerment, purpose, and collaboration to capture the culture dimensions most critical for Agile development. *Continuous improvement* was not yet measured in the units in a formalized way, but it was evaluated through self-assessments in several ARTs

with plans to scale the practice. The performance objectives for Agile trains with sample metrics are presented in Figure 2. The objectives followed recommendations of the SAFe framework [22].

Productivity	Reduced feature cycle times	Completed features; iteration velocity
Time to market	More frequent releases	Feature lead time
Quality (development and production)	Reduced defects and share of downtime	Number of defects; mean time to restore
Continuous improvement	Improvement on other metrics	Self-assessments
Employee engagement	Increased employee satisfaction	Surveys on satisfaction and empowerment
Customer satisfaction	Improved likelihood to recommend	Net promoter score
Alignment	Progress and impact on strategic objectives	Objectives and key results

Figure 2. Objectives, expected benefit and sample metrics for ARTs and teams

Even if agile teams were not measured or evaluated on organizational objectives, they were encouraged to measure the impact of their features. This was also addressed in the Agile metrics with the objective of *alignment to strategic objectives* that aimed to ensure that the trains and teams created impact in terms of the units' KPIs. However, setting metrics for impact and following them up was entirely up to the teams, and the results were also primarily used only within them. To guide the teams in setting metrics for their impact, the units had recently introduced an "Objectives and Key results" (OKR) framework. Introduced by Intel and used among others by Google, OKR required all features to have an overall objective as well as 2–5 quantitative key results that can be used to measure progres toward the objective. Apart from saving time and effort from development teams, the framework was adopted to increase alignment across the teams.

In both units, learning received special attention due to its important role in enabling other performance areas. Systematic plans to improve in the area included, for example, regular retrospectives, knowledge-sharing meetings, hackathons, innovation days, developer forums, and in some teams even "cultural goals." Despite this focus, learning was not formally measured. This is in line with earlier findings that found learning to be among the least prioritized performance areas in Agile organizations [18].

Overall, efficiency and productivity were perceived as the most important performance objectives for the execution organization. While management had the responsibility for ensuring that the ARTs used their time on the most value-adding features, ARTs and teams had responsibility for doing it as efficiently as possible. The line organization was

used only for managing individuals and therefore had a minor role in performance measurement in the units.

4.3. Measurement Challenges

In addition to performance objectives and metrics, the interviews provided insights into measurement practices and challenges in the case units. The most frequently mentioned challenge was a feeling of disconnect between the unit KPIs and the work delivered in the teams. It was perceived as difficult for the Agile teams and ARTs to know how, if at all, they contributed to the strategic unit's KPIs. One explanation for the disconnect was that the unit KPIs were on such a macro level that it was difficult to see the impact of micro-level features on them. Another perspective was that the teams were simply unable to select and implement metrics that would demonstrate the connection. "*The translation of [the unit] objectives or helping to understand how they work in everyday life is not always working*," one of the interviewees summarized.

External factors influencing the metrics formed another measurement challenge in the units. Many teams felt that the unit KPIs were influenced by such powerful factors beyond their control that their role in the outcome was negligible. Even if the challenge was recognized by those setting the unit KPIs, it was found to be difficult to solve, since the managers also saw downsides with setting more specific and actionable metrics, as these made it harder to grasp organizational performance. To cover all the teams and activities, the organizations would need to have tens of KPIs, which was impractical for top management. "*It is a hard problem to solve. I haven't found a good solution even though I have been thinking about it for very very long time*," one of the interviewees commented on the trade-off between actionable and holistic metrics.

Despite their different roles, all interview participants perceived the visibility between team metrics and unit KPIs as crucial for the organization to deliver optimal results. This observation is in line with earlier findings by Oza and Korkala [19]. As possible solutions, the interviewees suggested establishing clear connections between the metrics, increasing training on how to measure the impact of features in the teams through the OKR framework, and frequent communication on unit performance in the teams.

Related to the difficulties in seeing how their work contributed to the unit objectives, some interview participants expressed concern about whether the backlog prioritization process judged development features objectively. In particular, interviewees from development teams were concerned that when a feature had a higher impact on customer satisfaction than financials, the leaders prioritized features with higher financial impacts despite customer satisfaction being a primary objective. Even if the suspicion was difficult to prove, it was shared by several people and had an impact on their motivation. "*My perception from the things that I find on these hub boards is that business impact is the one that is deciding*," one of the interviewees described.

In addition to challenges related to selecting performance objectives, the units faced several challenges of a more practical nature. Inadequate data availability or consistency limited the selection of performance metrics and required teams to either only measure what was available or invest in data infrastructure. The time and effort required to collect and process data for performance metrics formed another practical challenge that also limited the selection of metrics. Finally, lack of skills and guidance for setting metrics and processing data also hindered the units' abilities to benefit from performance measurement. The lack of standardization of measurement practices also challenged managers, as they were not able to benchmark the teams and often had to take decisions based on poor-quality data.

Performance measurement was also impacted by general organizational issues. One of these was collaboration between units with different ways of working. Many support units did not work Agile or had different cadences, which made it difficult for the case units to have support in reasonable time. "*Where I see a big issue is when the support organizations become Agile. You know, we don't have three-month PIs but just two-week sprints. But when a support organization where you have a dependency uses three-month sprints, you immediately have not a dependency but an interdependency*," one interviewee exemplified. Despite being challenging, enhanced collaboration across units was of increasing importance in the case units. There was an aim to improve especially the flow between processes by setting performance objectives and measuring them end to end rather than separately for each process.

In addition to challenges, there were also identified couple of best practices in the case units. Frequent review and reprioritization of metrics, especially within the ARTs, had been found to be ensure that the most relevant metrics were continuously measured, since, for example, the Agile maturity level could change the relative importance of the metrics quickly. Many of the Agile practices built on each other, and hence, if the basics were not in place, there was no use measuring an advanced practice. Furthermore, regular communication and discussion on both unit KPIs and

Agile metrics were found to increase the motivation and feeling of purpose within ARTs and teams.

5. Discussion

This study has investigated performance measurement in scaled Agile organizations. The review of the extant literature suggested that, due to its iterative nature and constantly changing priorities, traditional project-based performance metrics support Agile development poorly. Instead, the literature had presented three alternative foundations for deriving performance objectives in Agile development organizations: strategic business objectives [12, 13], stakeholder needs [14, 15], and benefits outlined in the Agile principles [16, 17]. Based on these approaches, we identified five performance dimensions for Agile development: 1) customer value, 2) financial value, 3) internal process efficiency, 4) collaboration, and 5) learning. Of these, customer value and internal process efficiency were identified as the most relevant dimensions for Agile organizations.

The case analysis was conducted separately on two different organizational levels. On the unit level, performance objectives were determined based on strategic business objectives and emphasized outcomes. There were four main objectives: customer satisfaction, product quality, income growth, and operational efficiency. These objectives mainly addressed only the customer and financial value of the five dimensions suggested by the literature.

On the program level, the focus in performance measurement was found to shift from externally created value to internal efficiency. Performance objectives were derived from Agile principles using the SAFe framework recommendations and included productivity, time to market, quality in development, continuous improvement, customer satisfaction, and employee engagement. While these objectives addressed all the performance dimensions apart from financial value, there was a clear emphasis on internal process efficiency, demonstrated, for example, by the number of metrics it was measured on.

The principal tool for aligning unit-level performance objectives with activities in the execution organization was backlog prioritization. Since the prioritization process was designed to make sure that the organization continuously worked on features with the most impact on strategic objectives, no need was seen to measure the created impact within the execution organization. Agile teams were, however, encouraged to individually measure the impact that their features had on unit KPIs through the OKR framework.

The differences between the unit- and the program-level performance objectives in the cases demonstrated a clear structure and prioritization between the performance dimensions (cf. Figure 3). First, the dimensions of customer and financial value addressed value created for external stakeholders and were therefore regarded as the primary objectives. These dimensions were measured by unit KPIs. Second, internal process efficiency was mainly perceived as important because of its key role in achieving the primary objectives. It was regarded as a secondary objective and measured with Agile metrics on the execution organization only. On the tertiary level, there were performance dimensions of learning and collaboration that did not directly influence the primary objectives but still had a role in enabling performance. These were addressed through several initiatives but not necessarily formally measured.
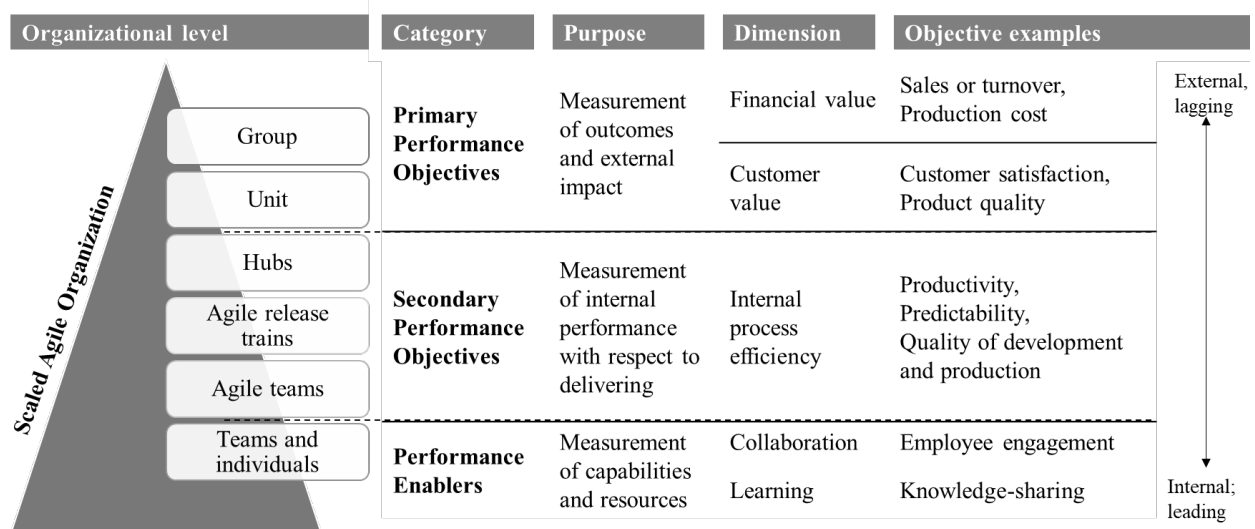
Organizational level		Category	Purpose	Dimension	Objective examples	
Group		**Primary Performance Objectives**	Measurement of outcomes and external impact	Financial value	Sales or turnover, Production cost	External, lagging
Unit				Customer value	Customer satisfaction, Product quality	
Hubs		**Secondary Performance Objectives**	Measurement of internal performance with respect to delivering	Internal process efficiency	Productivity, Predictability, Quality of development and production	
Agile release trains						
Agile teams						
Teams and individuals		**Performance Enablers**	Measurement of capabilities and resources	Collaboration	Employee engagement	Internal; leading
				Learning	Knowledge-sharing	

Figure 3. Performance objective structure in scaled Agile organizations

Overall, our findings from the case study were aligned with prior studies by proposing customer value and internal process efficiency as the key performance dimensions to measure in Agile development organizations. The empirical findings, however, suggested the dimensions to be emphasized differently on distinct organizational levels, with internal process efficiency being of primary importance only on the program level and below. Financial value was identified as a key performance dimension in the case units, even if Alahyari et al. [18] suggested it to be only of secondary value for Agile organizations. Conversely, the importance of learning was highlighted more in previous studies than in the case units, for whom it was primarily an enabler for achieving other objectives, even if an important such.

The study also examined challenges associated with performance measurement in scaled Agile organizations. Interestingly, the most frequently mentioned challenges were found to arise from the measurement structure that emphasized distinct objectives on different organizational levels. As the most prominent challenge, the study identified a feeling of disconnect between the unit KPIs and work delivered in the teams since, due to the measurement structure, ARTs and Agile teams had few metrics measuring their contributions to the unit KPIs. This disconnect resulted in disengagement, which was further amplified by the many external factors impacting the unit KPIs, thus making it difficult for teams to control them. Prior studies have also identified the challenge of establishing the visibility of the impact of micro metrics on macro objectives [19]. In addition to establishing clear links between metrics, the findings suggest training on the use of the OKR framework and enhanced communication as potential remedies for this challenge.

Other performance measurement challenges in the case units included lack of available data and time and skills for processing them. Increased training on measuring impact in teams was perceived as necessary, not just for Agile teams but also for managers who lacked quality data for decision-making and had few possibilities to benchmark teams due to misaligned measurement practices. Furthermore, dependencies on support units with different cadences and ways of working formed a major challenge that was also recognized by prior research [18].

Our study provides existing literature on performance measurement with two contributions for further investigation. First, the study presents an example of how scaled Agile organizations can determine and structure performance objectives on portfolio and program levels. Since prior studies examined metrics primarily within Agile teams, little is known about how large organizations balance and structure strategic, software, process, behavioral, and other metrics on distinct organizational levels. Our findings suggest that emphasizing and prioritizing objectives differently between organizational levels may support scaled Agile organizations in optimizing outcomes from Agile development methods.

Second, the study suggests that financial performance objectives may also be relevant for scaled Agile organizations. Although earlier studies suggested business value as a key metric for Agile development organizations [11, 19], there is contradicting evidence of its prioritization [18]. Our findings provide support for its applicability.

Our study also has implications for practice. The presented framework (cf. Figure 3) can be used to provide guidance for scaled Agile organizations to select and prioritize performance objectives. By selecting objectives from all three categories, organizations can address performance holistically and achieve a balance between business and software development metrics. Furthermore, the study findings allow industry practitioners to understand, avoid, and mitigate challenges associated with scaling performance measurement in Agile organizations.

6. Conclusions and Future Work

This study examined how scaled Agile organizations determine and measure performance objectives on different organizational levels. The study findings contribute to existing research on performance measurement by providing insights into how strategic, software, process, and other performance objectives can be combined and balanced in scaled Agile organizations. The created framework on performance objectives in scaled Agile organizations can also assist practitioners in selecting and prioritizing performance objectives and understanding challenges related to structuring and measuring them.

The main limitation of the study was its narrow empirical data that consisted of only ten interviews in two case units belonging to the same organization. Additional research is therefore needed to validate the findings and test their applicability within other organizations and industrial domains. Furthermore, the study was limited to consider only SAFe framework from the various approaches for scaling Agile. Hence, there is a need for further studies to gain an understanding on the potential implications of scaling approaches on performance measurement.

While it is known that priorities in an organization may vary according to its stage of Agile maturity, the cross-sectional nature of the study limited its ability to shed light on the dynamics between maturity and

performance measurement. In the future, longitudinal studies are therefore necessary to increase understanding on how performance measurement can best be adapted to support Agile practices under changing circumstances. Potential adaptation mechanisms to be investigated could include, e.g., weighting of objectives.

7. References

[1] C. Jones, Applied Software Measurement: Global Analysis of Productivity and Quality, McGraw-Hill Education Group, 2008.

[2] D. R. Greening, "Agile Enterprise Metrics", 48th Hawaii International Conference on System Sciences, IEEE, 2015.

[3] E. Kupiainen, M. V. Mäntylä, and J. Itkonen, "Using Metrics in Agile and Lean Software Development – A Systematic Literature Review of industrial Studies", Information and Software Technology, 62, 2015, pp. 143–163.

[4] B. Kitchenham, "What's Up With Software Metrics? – A Preliminary Mapping Study", Journal of Systems and Software, 83(1), 2010, pp. 37–51.

[5] P. Ram, P. Rodriguez, and M. Oivo, "Software Process Measurement and Related Challenges in Agile Software Development: A Multiple Case Study", International Conference on Product-Focused Software Process Improvement, Springer, Cham, 2018, pp. 272–287.

[6] L. Gren, R. Torkar, and R. Feldt, "The Prospects of a Quantitative Measurement of Agility: A Validation Study on an Agile Maturity Model", Journal of Systems and Software, 107, 2015, pp. 38–49.

[7] R. M. Fontana, I. M. Fontana, P.A. da Rosa Garbuio, S. Reinehr, and A. Malucelli, "Processes Versus People: How Should Agile Software Development Maturity be Defined?", Journal of Systems and Software, 97, 2014, pp. 140–155.

[8] VersionOne, "14th Annual State of Agile Survey", 2020. Available at: https://stateofagile.com/

[9] R. K. Yin, Case Study Research: Design and Methods (ed.)., Thousand Oaks, CA, USA, 2003.

[10] S. Mitra, V. Sambamurthy, and G. Westerman, "Measuring IT Performance and Communicating Value", MIS Quarterly Executive, 10(1), 2011.

[11] D. Hartmann and R. Dymond, "Appropriate Agile Measurement: Using Metrics and Diagnostics to Deliver Business Value", AGILE 2006 (AGILE'06), IEEE, MN, USA, pp. 6–134.

[12] M. Khurum, T. Gorschek, and M. Wilson, "The Software Value Map – An Exhaustive Collection of Value Aspects for the Development of Software Intensive Products," Journal of Software: Evolution and Process, 25(7), 2013, pp. 711-741.

[13] R. S. Kaplan and D. P. Norton, "The Balanced Scorecard-Measures that Drive Performance", Harvard Business Review, 70(1), 1992, pp. 71-79.

[14] V. Mahnic and N. Zabkar, "Using COBIT Indicators for Measuring Scrum-based Software Development", WSEAS Transactions on Computers, 7(10), 2008, pp. 505–1617.

[15] B. List, R.M. Bruckner, and J. Kapaun, "Holistic Software Process Performance Measurement from the Stakeholders' Perspective," 16th International Workshop on Database and Expert Systems Applications (DEXA'05), IEEE, 2005, pp. 941–947.

[16] M. Olszewska, J. Heidenberg, M. Weijola, K. Mikkonen, and I. Porres, "Quantitatively Measuring a Large-scale Agile Transformation", Journal of Systems and Software 117, 2016, pp. 258–273.

[17] W.C. Davis, Agile Metrics in Action: Measuring and Enhancing the Performance of Agile Teams, Manning Publications, CT, USA, 2015.

[18] H. Alahyari, R.B. Svensson, and T. Gorschek, "A Study of Value in Agile Software Development Organizations," Journal of Systems and Software, 125, 2017, pp. 271–288.

[19] N. Oza and M. Korkala, "Lessons Learned in Implementing Agile Software Development Metrics", The UK Academy for Information Systems Conference, 2012, p. 38.

[20] A.D. Neely, C. Adams, and M. Kennerley, The Performance Prism: The Scorecard for Measuring and Managing Business Success," Prentice Hall; Financial Times, London, 2002.

[21] Y. Dubinsky, D. Talby, O. Hazzan and A. Keren, "Agile Metrics at the Israeli Air Force", Agile Development Conference (ADC'05), IEEE, 2005, pp. 12–19.

[22] D. Leffingwell, Metrics—Scaled Agile Framework, Agile Inc, 2019. Available at: https://www.scaledagileframework.com/metrics/

[23] M.Q. Patton, "Two Decades of Developments in Qualitative Inquiry: A Personal, Experiential Perspective", Qualitative Social Work, 1(3), 2002, pp. 261–283.

[24] V. Braun and V. Clarke, "Using Thematic Analysis in Psychology", Qualitative Research in Psychology, 3(2), 2006, pp. 77–101.

[25] R.A. Singleton, B.C. Straits, and M. M. Straits, Approaches to Social Sciences, Oxford University Press, New York, 2005.