# Challenges and Success Factors for Metrics in Large-Scale Agile Development

Pascal Philipp
*Technical University of Munich*, pascal.philipp@tum.de

Franziska Tobisch
*Technical University of Munich*, franziska.tobisch@tum.de

Florian Matthes
*Technical University of Munich*, matthes@tum.de

# Challenges and Success Factors for Metrics in Large-Scale Agile Development

*Completed Research*

**Pascal Philipp**
Technical University of Munich
pascal.philipp@tum.de

**Franziska Tobisch**
Technical University of Munich
franziska.tobisch@tum.de

**Florian Matthes**
Technical University of Munich
matthes@tum.de

## Abstract

Contemporary organizations widely use agile software development to react to unpredictable changes in their business environment. Due to the success of agile methods in contexts similar to the agile sweet spot, organizations have been applying them on a larger scale. However, maintaining oversight in large-scale agile development remains a problem. Metrics can tackle this problem by increasing transparency, but organizations have struggled with their adoption. Furthermore, extant research on large-scale agile development lacks publications investigating metric challenges and success factors. Against this backdrop, we conducted an expert interview study with 23 experts from 13 organizations. The most mentioned challenges are data collection challenges, lack of metric usefulness, and metric calculation challenges. On the other hand, the most occurring success factors are context-dependent metric adoption, implementing an agile metric management process, and ensuring understanding of the metric purpose.

### Keywords

Agile software development, large-scale agile development, metrics, challenges, success factors.

## Introduction

Highly dynamic business environments require organizations to agilely deal with unpredictable internal and external changes (Van Oosterhout et al. 2006). Organizations enable business agility by implementing an agile IT and process architecture (Van Oosterhout et al. 2006). Consequently, developing software agile is crucial. Nowadays, agile software development prevails itself as the dominant software development approach (Dybå and Dingsøyr 2008). Well-known agile methods, such as Scrum and eXtreme programming (XP) are designed for co-located, self-organizing, and small teams that develop software in close collaboration with customers, using regular feedback and rapid development iterations (Kettunen 2007). Agile methods succeed in contexts identical or very similar to the so-called "agile sweet spot" (Kruchten 2013). This sweet spot consists of ideal conditions comprising co-located teams of 5–12 members, frequent deliveries, stable underlying architecture, and medium to low system criticality (Nord et al. 2014). Due to the success of agile methods in the sweet spot, organizations have applied agile software development methods on a larger scale (Digital.ai 2021; Dikert et al. 2016; Dingsøyr et al. 2018; Uludağ et al. 2021a) characterized by single or multiple multi-team settings consisting of at least two agile teams (Dikert et al. 2016; Dingsøyr et al. 2018; Uludağ et al. 2021a). Agile at scale must deal with challenges, such as inter-team coordination (Brown et al. 2013; Dikert et al. 2016; Uludağ et al. 2018) and communication (Dikert et al. 2016; Talby and Dubinsky 2009; Uludağ et al. 2018). Since agile methods, in their original design, cannot deal with these challenges, they must be extended (Uludağ et al. 2021a). Consequently, multiple scaling agile frameworks have been proposed and continuously improved (Uludağ et al. 2021b). The Scaled Agile Framework (SAFe) and Scrum@Scale are the most widely adopted frameworks in the industry (Digital.ai 2021). Adopting metrics beyond the team level has become a relevant research topic in large-scale agile development (Dingsøyr and Moe 2014). Several scaling agile frameworks (e.g., SAFe

(Scaled Agile Inc. 2021) and Scrum@Scale (Sutherland and Scrum Inc. 2021)) are aware of the relevance of metrics and therefore propose metrics and best practices. For example, Scrum@Scale recommends metrics to support data-driven decision-making, streamline the work of teams, stakeholders, or leadership, and reduce decision latency (Sutherland and Scrum Inc. 2021). Despite the relevance of metrics, their implementation remains challenging. Several studies regarding challenges and best practices for implementing metrics in large-scale agile development have been conducted (Korpivaara et al. 2021). However, since the scope of this research included only two cases belonging to the same organization, additional research is required to validate the findings and test their applicability in other organizations and industrial domains (Korpivaara et al. 2021). Therefore, we fill this gap by raising the following research questions:

*RQ1: What challenges have been reported for metrics in large-scale agile development?*

*RQ2: What success factors have been reported for metrics in large-scale agile development?*

## Background and Related Work

In 2001, the agile manifesto was proposed to improve software development at that time (Beck et al. 2001). In the manifesto, particular importance was assigned to individuals and interactions, working software, customer collaboration, and responding to change. Agile methods, such as Scrum, XP, feature-driven development, and crystal methodologies adhere to the manifesto to varying degrees (Dingsøyr et al. 2012). These methods were created for and succeeded in contexts close to the agile sweet spot (Kruchten 2013). According to Kruchten (2013), more than 70% of software projects operate in contexts comparable to the agile sweet spot. However, agile methods fail in contexts different from what they were created for if they are not adapted (Kruchten 2013). Thus, scaling agile frameworks extend the agile methods to deal with challenges arising in larger contexts (Uludağ et al. 2021a). Typical challenges in larger settings include additional coordination and inter-team dependencies (Dikert et al. 2016). Currently, there are more than 20 frameworks (Uludağ et al. 2021b). The most widely adopted frameworks are SAFe, Scrum@Scale, and Enterprise Scrum (Digital.ai 2021).

In traditional software development, metrics were used for decades (Fenton and Neil 2000) to measure or predict attributes (i.e., measurable characteristics) of processes, products, or resources (Fenton and Neil 2000). Several researchers have stressed that successfully implemented metrics in traditional software development cannot always be adopted in agile development (Hartmann and Dymond 2006; Oza and Korkala 2012; Poligadu and Moloo 2014). Whereas metrics applied in traditional software development measure the achievement of a predefined plan, agile metrics must align with the value of embracing change (Korpivaara et al. 2021; Kupiainen et al. 2015). Metrics implementation is challenging in agile (Ram et al. 2018) and large-scale agile development (Brown et al. 2013; Korpivaara et al. 2021; Stettina and Schoemaker 2018). Several researchers have revealed challenges and success factors for metrics in scaling agile environments. According to Ram et al. (2018), challenges faced in operationalizing metrics include difficulty in deriving actionable inputs, existing processes inhibiting change, and the lack of data or appropriate tools for data production. Similarly, Korpivaara et al. (2021) identified inadequate data availability as a challenge and highlighted challenges including the disconnection between the unit Key Performance Indicators (KPIs) and the teams' work, visibility between team metrics and unit KPIs, and strong external factors influencing the metrics. Additionally, they highlighted inadequate data consistency and a lack of skills and guidance for setting metrics, data processing, and lack of standardization and organizational issues as challenges related to performance metrics (Korpivaara et al. 2021). In their study Oza and Korkala (2012) present the key challenge being that metrics can become interests of individuals rather than contributing to organizational improvement. According to Korpivaara et al. (2021), best practices for performance measurement are regular communication and discussion on unit KPIs and agile metrics. Also, Ram et al. (2019) revealed several success factors for adopting metrics in agile environments, including applying an iterative and agile approach to the metrics program or using familiar metrics to establish trust. More generally, Dikert et al. (2016) identified challenges and success factors for agile transformation such as lack of investment or change resistance which can also apply to the metric adoption.

## Study Design

*(1) Study Design and Overview:* To answer our research questions, we conducted semi-structured interviews and analyzed the collected data. The chosen research design is appropriate since semi-structured interviews can be used to investigate practical problems, such as the adoption of metrics in large-scale agile software development (Seaman 1999). To ensure rigor, we designed our research methodology using the guidelines of Myers and Newman (2007) for semi-structured interviews. Our study, although exploratory, includes descriptive elements. Table 1 provides an overview of the study participants. We conducted 23 interviews with experts from 13 organizations. We recruited the study subjects via an email sent to practitioners and researchers working in the domain of agile software development. Note that three interviews at two organizations were excluded since we only included interviewees that had sufficient topic-related knowledge. The included interviewees are characterized by a broad spectrum of industry backgrounds and job roles, ensuring a large "variety of voices" covering multiple viewpoints (Myers and Newman 2007). All except for one expert organization (TeleCo) have their headquarters in Germany. To the best of our knowledge, the experts had no formal training on scaled agile frameworks.

| No. | Company | Role(s), Large-scale agile software development experience (years) | Scaling Agile Framework | Program Size | No. | Company | Role(s), Large-scale agile software development experience (years) | Scaling Agile Framework | Program Size |
|---|---|---|---|---|---|---|---|---|---|
| **E1** | CarCo1 | Agile Coach, 11–15 | SAFe | 750 | **E11** | ConsultCo | Manager, 3–5 | LeSS. Scrum-of-Scrums | 50 |
| **E2** | BankCo1 | Manager, 3–5 | SAFe | 6000 | **E12** | CarCo2 | Scrum Master and Agile Coach, 3–5 | Scrum-of-Scrums | 100 |
| **E3** | HealthCo | Manager, 6–10 | SAFe, LeSS | 450 | **E13** | RetailCo1 | Agile Coach, 6–10 | SAFe | 120 |
| **E4** | ConsultCo | Scrum Master and Agile Coach, 1–2 | LeSS, Scrum-of-Scrums | 50 | **E14** | InsureCo1 | Scrum Master and Agile Coach, 3–5 | SAFe | 75 |
| **E5** | HealthCo | Manager and Agile Coach, 6–10 | SAFe, LeSS | 450 | **E15** | InsureCo1 | Release Train Engineer, 1–2 | SAFe | 80 |
| **E6** | ConsultCo | Developer and Scrum Master, 3–5 | LeSS, Scrum-of-Scrums | 50 | **E16** | InsureCo2 | Scrum Master and Agile Coach, 1–2 | Scrum@Scale | 24 |
| **E7** | ConsultCo | Product Owner and Business Analyst, 1–2 | LeSS, Scrum-of-Scrums | 50 | **E17** | InsureCo1 | Manager and Agile Coach, 3–5 | Nexus, Scrum-of-Scrums | 90 |
| **E8** | ConsultCo | Software Architect, Solution Architect, 3–5 | LeSS, Scrum-of-Scrums | 50 | **E18** | BankCo1 | Agile Coach, 3–5 | SAFe | 400 |
| **E9** | TeleCo | Scrum Master, Agile Coach and Developer, 11–15 | Scrum-of-Scrums | 100 | **E19** | RetailCo2 | Scrum Master and Agile Coach, 6–10 | Scrum-of-Scrums | 70 |
| **E10** | HealthCare Co | Software Architect, 11-15 | Scrum-of-Scrums | 450 | **E20** | TransportationCo | Release Train Engineer, 6-10 | SAFe | 550 |

**Table 1. Interview Partners**

*(2) Data Collection*: We conducted the expert interviews using videotelephony between January and March 2021. At least two researchers participated in each interview to ensure observer triangulation (Runeson and Höst 2009). All interviews were recorded and had an average duration of 82 minutes. The interviews summed to 1.640 minutes and differed in length due to varying extents of metric usage within the expert organizations. Furthermore, the interview durations exclude the introduction and clarification times at the start of the interview. The research goal was recalled during the introduction, and the clarifications addressed the interview structure. The interview protocol was not reviewed by an ethics committee or human subjects review board. Each interview had the same outline and the interview questions remained unchanged. Though, due to the nature of semi-structured interviews, slight variations regarding the question order or wording occurred. We separated the interview questions into two sections. In the first

section, we asked experts questions on their professional background and large-scale agile development environment. Some interviewees answered questions of the first section upfront and sent their written responses to us. The second section included questions regarding metric adoption. Moreover, we included files shared by the interviewees to facilitate the triangulation of data sources. The results communicated here are a subset of the overall collected data. We plan to publish the remaining empirical evidence in further publications.

*(3) Data Analysis:* Two researchers took the transcripts, and a third was the quality gate reviewing the completeness and consistency of each transcript. We dismissed sensitive information and ensured anonymity by assigning unique acronyms to each expert. Furthermore, we used deductive and inductive coding (Miles et al. 2020) to code the transcripts, prefilled questionnaire parts, and additional files. We aligned our coding with the two-cyclic approach proposed by Saldaña (2021). Thus, we assigned descriptive first-cycle codes (Miles et al. 2020) to the text segments assigned to the higher-order second-cycle codes. We based the second-cycle deductive coding on an initial list of higher-order codes (Saldaña 2021) and aligned the list with the interview questionnaires. Additional second-cycle codes were inductively created whenever new findings that could not be assigned to an existing code category occurred (Saldaña 2021). Introducing new codes necessitated a re-codification of the previously coded data. Accordingly, we created the final coding incrementally. As tools for coding and analyzing the transcripts we used MAXQDA[1] and Excel[2]. Furthermore, we discussed and resolved all conflicts by mutual consent to enhance result validity. Also, the experts were contacted to resolve ambiguities and clarify interview statements.

# Results

## *Challenges of Metrics in Large-Scale Agile Development*

This section answers the first research question: *What challenges have been reported for metrics in large-scale agile development?* We identified eleven challenges, as summarized in Table 2.

| Challenge (Occurrences) | Experts | # Experts (Occurrences) |
|---|---|---|
| Data collection challenges (21,13%) | E1–E3, E5–E15, E17–E19 | 17 (85%) |
| Lack of metric usefulness (15,21%) | E1–E12, E14, E15, E17, E18, E20 | 17 (85%) |
| Metric calculation challenges (13,24%) | E1–E5, E7, E8, E10, E11, E13, E15, E17 | 12 (60%) |
| Metric-based decision-making challenges (10,99%) | E1–E11, E12, E14, E16–E20 | 18 (90%) |
| Negative effects for employees (10,70%) | E1, E2, E4, E6–E9, E11–E17, E19, E20 | 16 (80%) |
| Lack of metric understanding (7,04%) | E1–E3, E9, E10, E12–E15, E19 | 10 (50%) |
| Change resistance (6,48%) | E1, E2, E7, E10, E12, E15–E17, E19, E20 | 10 (50%) |
| Focus on local instead of global optimum (5,07%) | E1, E2, E7–E10, E13–E15, E17 | 10 (50%) |
| Metric definition challenges (5,07%) | E1–E3, E5–E7, E9, E11, E14, E16 | 10 (50%) |
| Manipulation of metric values (3,10%) | E6–E8, E11, E15, E17, E19 | 7 (35%) |
| Imbalance between control and team autonomy (1,97%) | E9, E12, E19, E20 | 4 (20%) |

**Table 2. Challenges of Metrics in Large-Scale Agile Development**

*(1) Data collection challenges:* Data collection for the metric is notably a common problem. Several experts mentioned the dependence on data quality (E1–E3, E5, E7, E9, E10, E12–E15, E17–E19), data availability (E1–E3, E6, E7, E10–E14, E18), data collection complexity (E1, E2, E8, E10), and missing anonymity of data collection (E6, E17) as related challenges. In particular, the dependence on data quality is experienced as the main challenge since the metric can lose its value, for example, due to irrelevant data (E1) or biased human data sources (E2). Furthermore, employees missing discipline for entering data was mentioned as a challenge in data availability (E13, E18).

*(2) Lack of metric usefulness:* Ensuring metric usefulness is experienced as a problem by several experts. Deriving value from the metrics (E1–E7, E10–E12, E14, E17, E18, E20) and error-prone metric-based

[1] https://www.maxqda.com
[2] https://www.microsoft.com/de-de/microsoft-365/excel

planning (E7–E11, E15, E20) were mentioned as related challenges. For deriving value from the metrics, the experts experience a lack of insights on reasons for changes or deviations (E1, E2, E5, E10, E14, E16) and that metrics were only a starting point for problem identification (E10, E14, E22). Concerning the challenge of error-prone metric-based planning, an expert stated that it is difficult to derive stable predictions based on the metric (E8).

*(3) Metric calculation challenges:* Metric calculation is another common problem. Several experts mentioned the incorporation of all relevant variables (E2, E4, E7, E8, E10, E11, E13, E15), provision of calculation resources (E3, E5), aggregation of team-level values into program-level values (E5, E8, E10, E17), complexity of metric calculation (E1, E3, E11, E17), and lack of adequate tool support (E5) as challenges related to this problem. Notably, incorporating all relevant variables was echoed as a challenge because not considering variables such as fluctuations in velocity planning (E4, E8, E10, E11), holiday times (E7, E8, E11, E13), and size changes of epics (E2, E4, E7, E8, E11) lead to reduced a metric value. Moreover, the high time effort for writing metric tests (E5) and measuring the metric (E3) is perceived as challenging.

*(4) Metric-based decision-making challenges:* Decision-making based on metric values was perceived as a challenge. For example, multiple experts mentioned the difficulty of deciding on next actions (E1, E3–E5, E8, E9, E12, E14, E17), ineffective management based on metric values (E2, E4, E6, E7, E9, E10, E16, E18–E20), and misleading information derived from metrics (E1, E11) as related challenges. The organization ConsultCo struggled, for instance, to identify actions and responsible persons to resolve blockers (E8). Additionally, ineffective management due to incorrect metric values occurs, for example, when managers conduct unfeasible comparisons between teams (E2, E4, E6, E7, E10, E20) or programs (E19). Expert E10 stressed the importance of avoiding inter-team comparisons based on the velocity: *"[…] reflect that the team velocity is different for every team. So that's one of the biggest things that you can do wrong when it comes to agile […] to say that every team has the same velocity. But if you reflect that already in your planning you can simply sum up all story points and over all teams, the burn-down will go in the right direction in theory, practically of course always something goes wrong."*

*(5) Negative effects for employees:* Negative effects for employees are another common problem and include challenges, such as negative psychological effects (E1, E2, E4, E7–E9, E11, E13–E17, E19, E20), internal competition (E6, E8, E9, E11), and feeling of being controlled (E1, E12, E15, E16) as described by expert E1: *"[…] you start measuring people, so, there is […] disturbance about the big transparency […] there is always the fear, that the measurement will leads us to control the people"*. An example related to negative psychological effects is increased pressure and stress due to deviations from planned metric values (E2, E13). Also, sometimes metric-based criticism is perceived as a personal attack (E19). Furthermore, such metric-based comparisons lead to undesirable internal competition, thereby decreasing collaboration between teams (E9).

*(6) Lack of metric understanding:* A lacking metric understanding is another shared problem among the experts. A lack of metric understanding is caused by missing awareness of dependencies (E1–E3, E9, E15), a general lack of metric understanding (E1, E2, E12, E14, E19), lack of understanding the calculation (E10, E12), providing insufficient resources to understand metrics (E12, E13), and complexity (E10, E12). Sometimes missing knowledge on metrics' existence (E12) is another related challenge. Often a lacking understanding leads to misuse of the metric (E10). In particular, the managers' lack of understanding (E2, E19) is experienced as a problem.

*(7) Change resistance:* Change resistance in the organization is also a shared problem among the experts. Challenges related to change resistance are difficulty in convincing stakeholders (E1, E2, E15, E17, E19), lack of management commitment (E2, E22), establishing a common metric understanding within the organization (E12, E16, E17, E20), investment in change (E1, E7), and lack of common reporting (E10). Convincing stakeholders can be a problem because, for instance, achieving openness (E15, E17) and acceptance (E15), convincing managers with a traditional mindset to use agile metrics (E19), and justifying metrics before the work council (E1, E2) is difficult.

*(8) Focus on local instead of global optimum:* Working toward goals that do not contribute to an organization-wide optimum is another identified challenge (E1, E2, E7–E10, E13–E15, E17). For example, several experts mentioned that metric target values often did not align with essential goals like customer satisfaction (E14, E15, E17), product quality (E8, E10, E13), or cost-efficiency (E2).

*(9) Metric definition challenges:* Several experts mentioned challenges during the metric definition. Especially the definition of appropriate target values (E1, E2, E5, E9, E16), complexity of metric definition (E1, E6, E11, E14), definition of appropriate tolerance values (E3, E5), defining appropriate granularity of calculation parameters (E7), and defining appropriate calculation rules (E1, E11) were experienced as challenges. Problems related to defining appropriate target values include setting unrealistic target values (E5) and high variations between programs when setting target values on the portfolio level (E1). Additionally, the metric definition can be complex due to difficulty in harmonizing measurement units across teams (E11).

*(10) Manipulation of metric values:* Employees' manipulation of metric values is another challenge mentioned by several experts. For example, conservative estimating is a circumvention tactic to reach the metric target value (E15). Driving factors for metric value manipulation are competition between teams (E6, E8, E11, E17) and employees' feeling of being controlled (E15).

*(11) Imbalance between control and team autonomy:* Finding the right balance between control and ensuring team autonomy is another common challenge. Particularly balancing governance and team autonomy (E12, E19) and avoiding too high management involvement (E9, E20) are related challenges. Balancing governance and team autonomy is challenging because self-organization and control may be experienced as contradictory (E19). Regarding the involvement of management, the intervention of the program management keeps developers from working (E20) and, also, metric-based micro-management hinders people from working (E20).

## *Success Factors for Metrics in Large-Scale Agile Development*

This section answers the second research question: *What success factors have been reported for metrics in large-scale agile development?* We identified 22 success factors. Subsequently, we explain eleven in detail and summarize the remaining in the last paragraph. The success factors are summarized in Table 3.

| **Success Factor (Occurrences)** | **Experts** | **# Experts (Occurrences)** |
|---|---|---|
| Context-dependent metric adoption (16,15%) | E1–E3, E5–E20 | 19 (95%) |
| Implementing an agile metric management process (15,32%) | E1, E3, E5–E7, E9, E10, E12–E17, E20 | 14 (70%) |
| Ensuring understanding of metric purpose (8,70%) | E1, E2, E4, E6, E8–E13, E16, E17, E19, E20 | 14 (70%) |
| Managing interplay of metric and its environment (8,28%) | E1, E4, E6–E8, E10–E12, E20 | 9 (45%) |
| Keep metric adoption simple (7,66%) | E3, E5–E8, E10–E12, E14, E15, E17, E18, E20 | 13 (65%) |
| Measures for ensuring effective data collection (7,25%) | E1–E3, E5, E6, E8, E10–E12, E14, E17, E19, E20 | 12 (60%) |
| Enabling metric adoption by providing sufficient resources (6,83%) | E1–E5, E7–E11, E14, E17, E20 | 13 (65%) |
| Empowerment of teams in metric adoption (5,59%) | E4, E7–E12, E15, E17, E20 | 10 (50%) |
| Standardization of metric adoption (4,97%) | E5, E7–E12, E14 | 8 (40%) |
| Combining metric with qualitative feedback (3,11%) | E1, E9, E14, E17, E20 | 5 (25%) |
| Ensuring goal-orientation of metric (3,11%) | E1, E9, E10, E11, E12, E18, E19, E20 | 8 (40%) |
| Other (13,04%) | E1–E7, E9–E17, E19, E20 | 18 (90%) |

**Table 3. Success Factors for Metrics in Large-Scale Agile Software Development**

*(1) Context-dependent metric adoption:* A successful metric adoption must be context-specific. According to the experts, the context plays a crucial role during metric configuration (E1–E3, E5–E8, E10–E12, E14–E17, E19, E20), deciding next actions (E1, E3, E5, E11, E14, E15, E19, E20) based on metric values, metric value interpretation (E1, E2, E5, E7, E9, E13), metric measurement (E16–E18), and metric selection (E7). Furthermore, organizations must consider the requirements of the large-scale agile development context by adhering to agile values and principles (E1, E13, E17).

*(2) Implementing an agile metric management process:* Several experts identified adhering to an agile metric management process as a crucial success factor. According to the experts, continuous improvement (E1, E3, E5, E7, E10, E12, E14, E16, E20), collaborative inspection (E1, E6, E7, E9, E15, E16, E20), collaborative decision-making of stakeholders regarding metric lifecycle (E1, E6, E10, E15, E16), ensuring metric transparency (e.g., by using visualizations) (E7, E9, E10, E13, E16), and using lightweight tools (E12,

E17, E20) are crucial. Collaborative inspection can be facilitated by including metric discussions in regular events (E9, E15, E19).

*(3) Ensuring understanding of metric purpose:* Another success factor is ensuring an understanding of the purpose of the metric, thereby reducing change resistance (E6, E9, E12, E13, E16, E20) and avoiding comparisons between teams (E4, E6, E8, E10, E11, E17, E19, E20). Also, the experts highlighted the importance of increasing management understanding (E1, E2, E16, E19) and organization-wide understanding (E1, E10, E12). Furthermore, creating an understanding of the metric target values is considered important (E2).

*(4) Managing interplay of metric and its environment:* Each metric is embedded in an environment, and managing its dependencies within this environment is critical for its successful adoption. According to the experts, it is important to manage the relationships between metrics (E1, E4, E6–E8, E10, E12, E20) and their dependencies to the environment in general (E1, E4, E6, E8, E10, E11, E20). Also, combining related metrics (E12, E20) and raising the awareness of the difference between causality and correlation (E10) was considered important.

*(5) Keep metric adoption simple:* Another success factor considered by the experts is keeping the metric adoption simple. Experts highlighted the importance of achieving simplicity of metric measurement (E7, E10–E12, E14, E17, E18) and simplicity of metrics in general (E7, E8, E11, E12, E18, E20). They highlighted automation (E3, E5, E10, E20) and choosing metrics with low overhead (E12) as measures to simplify metric adoption. A simple metric adoption is crucial and can be achieved by adopting only a few metrics (E6, E10, E15). Another important factor is ensuring the simplicity of metric scaling (E10).

*(6) Measures for ensuring effective data collection:* Multiple experts suggested implementing measures for effective data collection. Therefore, establishing a data collection strategy (E1, E2, E6, E14, E17), ensuring data quality (E1, E8, E11, E19, E20), and an effective test environment (E3, E5, E10) are crucial factors to consider. Moreover, safeguarding the anonymity of data collection (E6, E14, E17) and assigning data collection responsibilities (E12) were considered important. Furthermore, clarifying the effort for data collection (E1) is a success factor.

*(7) Enabling metric adoption by providing sufficient resources:* Providing sufficient resources like appropriate tool support (E2, E4, E5, E7, E10, E11, E14), the provision of a metric support team (E3, E20), ensuring the availability of experts (E1, E7, E8, E11), granting a proper time (E5, E8, E9, E11, E17) and money budget (E3, E5) were considered important.

*(8) Empowerment of teams in metric adoption:* Empowering the team during metric adoption is another success factor. Thus, assigning extensive team responsibility (E4, E7, E9, E20) is critical. Further, granting the team autonomy regarding selecting and unselecting (E9, E10, E15), the metric calculation (E8, E11, E12), and configuration (E7, E17) is important. Finally, taking a team-centric view (E10) during the metric adoption is another key factor highlighted by experts.

*(9) Standardization of metric adoption:* Multiple experts recommend using best practices (E5, E7–E9, E11, E12, E14) for the metric adoption. Further, it is considered crucial to achieving standardization regarding calculation (E5, E11), configuration (E10), and reporting (E5, E10). Tools serve to standardize the metric adoption (E5).

*(10) Ensuring goal-orientation of metric:* Using metrics to support goals (E1, E9, E11, E12, E18, E20) can be achieved if the target values of the metrics are goals or if the metric values create transparency regarding goal-fulfillment. Experts mentioned setting realistic target values as a key factor (E10, E19, E20). Also, implementing Objectives and Key Results (E9) and measuring them using metrics can help organizations. Furthermore, metric target values should focus on the outcome instead of output (E18).

*(11) Combining metric with qualitative feedback:* Another success factor lies in combining the quantitative information of metrics with further qualitative feedback to enhance inspection (E1, E9, E14, E17, E20) as described by a Scrum Master (E20): *"The number can of course only be an indicator at this point. It always depends on what I do with it. So ultimately, as a Scrum Master […] I must then also look at a qualitative analysis, what were the problems […] in the sprint and must look at how we can make it better through measures that I generate in the retrospective."* Including qualitative feedback can either help to understand the quantitative values (E20) better or get a complete view of the situation. Surveys can serve to collect qualitative feedback (E14, E17, E20).

*(12) Other:* Further success factors include providing training, guidelines, and documentation (E1, E3, E9, E12, E14, E20), understanding what the metric is measuring (E2, E6, E7, E10, E12, E16, E19, E20), using metrics on higher organizational levels (E7, E9–E11, E13, E15, E20), communicating metrics' value to reduce change resistance (E1, E2, E15, E17, E19), only approved stakeholders can access metric information (E2, E6, E9, E10, E17, E19), assigning clear responsibilities (E5, E15), collaborating teams to achieve shared metric target values (E11), establishing a community of practice for metrics (E4), facilitating alignment of management and teams regarding metrics (E20), acceptance of tools by employees (E15), ensuring realistic expectation management based on metrics (E9).

# Discussion

To answer RQ1: "What challenges have been reported for metrics in large-scale agile development?" we identified eleven challenges summarized in Table 2. The challenges with the highest number of total occurrences are data collection challenges (21,13%), lack of metric usefulness (15,21%), and metric calculation challenges (13,24%). Similarly, Korpivaara et al. (2021) identified the time and effort required to collect data, inadequate data quality, and data availability as data collection challenges. Also, Ram et al. (2018) identify a lack of tools as a cause of data collection problems. Regarding metric usefulness, Oza and Korkala (2012) and Korpivaara et al. (2021) revealed that strong external factors influence metrics to a degree where their outcome is no longer controllable. Concerning metric calculation, Korpivaara et al. (2021) identify the lack of skills and guidance as problems. Our findings show that adopting metrics in scaling agile environments is challenging, as often tools easing the data collection and calculation are missing or are used for other purposes. Since many of the expert organizations are in the early phases of the agile transformation, the challenges can also be attributed to a low metric maturity level which will be improved at later stages in the transformation. We identified 22 success factors summarized in Table 3 to answer RQ2: "What success factors have been reported for metrics in large-scale agile development?". The success factors with the highest number of total occurrences are context-dependent metric adoption (16,15%), implementing an agile metric management process (15,32%), and ensuring understanding of metric purpose (8,70%). These success factors confirm the findings of extant research. Similar to the context-dependent metric adoption, Ram et al. (2019) highlight the need to implement flexible development processes to accommodate new practices, enabling the adoption of metrics. Moreover, Korpivaara et al. (2021) and Ram et al. (2019) highlight the importance of applying an iterative and agile way of working to the metrics program, including discussions and short feedback cycles. Finally, Oza and Korkala (2012) and Ram et al. (2019) show the importance of the metric purpose by highlighting that metrics serve as means for trend analysis and propose to establish a long-term vision to drive the technical benefits of metrics. Our success factors indicate that the metric adoption will not succeed if metrics are implemented without further customization. Contrarily, a careful and continuous ensuring of transparency, inspection, and adaption of the metric and its context and safeguarding an understanding of why metrics are implemented is crucial to their success in large-scale agile development.

The results of this study contribute to research on large-scale agile software development. First, in extension to extant literature (cf. Dikert et al. (2016); Korpivaara et al. (2021); Oza and Korkala (2012); Ram et al. (2018); Ram et al. (2019)) and to the best of our knowledge, we present novel challenges and success factors for metrics in scaling agile environments. We identified the three challenges negative effects for employees, manipulation of metric values, and imbalance between control and team autonomy. Practitioners can benefit from these findings by counteracting these challenges. For example, empowering the team during metric adoption is a suitable countermeasure since it demands managers to present metrics as optional means only implemented if compatible with the team's interests. Concurrently, it motivates managers to act as servant leaders by supporting teams during the metric adoption while partly relinquishing decision-power. We identified the three success factors empowerment of teams in metric adoption, combining metrics with qualitative feedback, and ensuring goal-orientation of metrics. Practitioners can benefit from these findings since each success factor offers a concrete starting point to facilitate the metric adoption. We recommend practitioners assess the necessity of introducing new or improving existing success factors based on their context-specific requirements. Second, our research makes a confirmatory contribution since the remaining challenges and success factors are consistent with extant research (cf. Dikert et al. (2016); Korpivaara et al. (2021); Oza and Korkala (2012); Ram et al. (2018); Ram et al. (2019)).

We used the assessment scheme of Runeson and Höst (2009) to evaluate possible validity threats. Since our study is exploratory with descriptive elements and does explain casual relationships, we did not consider internal validity a threat. To counteract the threats to construct validity, we included several sources, i.e., semi-structured interviews and internal documents, during data collection. Whenever the interpretation of the evidence required clarification, we contacted the interviewees again to resolve ambiguities. We countermeasure the threat of external validity by clearly outlining how our results relate to the experts and organizations of our study. We dealt with threats to reliability by implementing three countermeasures. First, each interview had at least two researchers. Second, each transcript was reviewed by a second researcher. Third, whenever the interpretation of the evidence required clarification, we resolved the ambiguities by contacting the experts again.

## Conclusion and Future Work

In this paper, we presented an expert interview study exploring challenges and success factors for metrics in large-scale agile development. We conducted an interview study with 23 experts from 13 organizations. The challenges with the highest occurrences are data collection challenges, lack of metric usefulness, and metric calculation challenges. The success factors with the highest occurrences are context-dependent metric adoption, implementing an agile metric management process, and ensuring understanding of the metric purpose. As a future research topic, we suggest using the identified challenges and success factors to inform the design of artifacts that support metric adoption in large-scale agile development. Further, researchers could use the success factors to develop a maturity grid to assess the metric adoption in scaling agile environments.

## REFERENCES


Beck, K., Beedle, M., Van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., Grenning, J., Highsmith, J., Hunt, A., and Jeffries, R. 2001. "Manifesto for Agile Software Development."

Brown, A.W., Ambler, S., and Royce, W. 2013. "Agility at Scale: Economic Governance, Measured Improvement, and Disciplined Delivery," *Proceedings of the 35th International Conference on Software Engineering (ICSE)*, San Francisco, CA, USA: IEEE, pp. 873-881.

Digital.ai. 2021. "15th State of Agile Survey." Retrieved 16. February, 2022, from https://info.digital.ai/rs/981-LQX-968/images/RE-SA-15th-Annual-State-Of-Agile-Report.pdf_ga=2.241760322.1357558838.1645022323-2137180703.1645022323

Dikert, K., Paasivaara, M., and Lassenius, C. 2016. "Challenges and Success Factors for Large-Scale Agile Transformations: A Systematic Literature Review," *Journal of Systems and Software* (119), pp. 87-108.

Dingsøyr, T., and Moe, N.B. 2014. "Towards Principles of Large-Scale Agile Development," *Proceedings of the 15th International Conference on Agile Software Development (XP 2014),* T. Dingsøyr, N.B. Moe, R. Tonelli, S. Counsell, C. Gencel and K. Petersen (eds.), Rome, Italy: Springer, Cham, pp. 1-8.

Dingsøyr, T., Moe, N.B., Fægri, T.E., and Seim, E.A. 2018. "Exploring Software Development at the Very Large-Scale: A Revelatory Case Study and Research Agenda for Agile Method Adaptation," *Empirical Software Engineering* (23), pp. 490-520.

Dingsøyr, T., Nerur, S., Balijepally, V., and Moe, N.B. 2012. "A Decade of Agile Methodologies: Towards Explaining Agile Software Development," *Journal of Systems and Software* (85:6), pp. 1213-1221.

Dybå, T., and Dingsøyr, T. 2008. "Empirical Studies of Agile Software Development: A Systematic Review," *Information and Software Technology* (50:9-10), pp. 833-859.

Fenton, N.E., and Neil, M. 2000. "Software Metrics: Roadmap," *Proceedings of the Conference on the Future of Software Engineering*, Limerick, Ireland: ACM, pp. 357-370.

Hartmann, D., and Dymond, R. 2006. "Appropriate Agile Measurement: Using Metrics and Diagnostics to Deliver Business Value," *Proceedings of the AGILE 2006 (AGILE'06)*, Minneapolis, MN, USA: IEEE, pp. pp. 6 pp.-134.

Kettunen, P. 2007. "Extending Software Project Agility with New Product Development Enterprise Agility," *Software Process: Improvement and Practice* (12:6), pp. 541-548.

Korpivaara, I., Tuunanen, T., and Seppänen, V. 2021. "Performance Measurement in Scaled Agile Organizations," *Proceedings of the 54th Annual Hawaii International Conference on System Sciences (HICSS)*, Kauaii, Hawaii, USA: University of Hawai'i at Manoa, pp. 6912-6921.

Kruchten, P. 2013. "Contextualizing Agile Software Development," *Journal of Software: Evolution and Process* (25:4), pp. 351-361.

Kupiainen, E., Mäntylä, M.V., and Itkonen, J. 2015. "Using Metrics in Agile and Lean Software Development – a Systematic Literature Review of Industrial Studies," *Information and Software Technology* (62), pp. 143-163.

Miles, M.B., Huberman, A.M., and Saldaña, J. 2020. *Qualitative Data Analysis: A Methods Sourcebook*. SAGE Publications, Inc.

Myers, M.D., and Newman, M. 2007. "The Qualitative Interview in Is Research: Examining the Craft," *Information and Organization* (17:1), pp. 2-26.

Nord, R.L., Ozkaya, I., and Kruchten, P. 2014. "Agile in Distress: Architecture to the Rescue," *Proceedings of the 15th International Conference on Agile Software Development (XP 2014),* T. Dingsøyr, N.B. Moe, R. Tonelli, S. Counsell, C. Gencel and K. Petersen (eds.), Rome, Italy: Springer, Cham, pp. 43-57.

Oza, N., and Korkala, M. 2012. "Lessons Learned in Implementing Agile Software Development Metrics," *Proceedings of the UK Academy for Informaiton Systems Conference (UKAIS)*, Oxford, UK, p. 38.

Poligadu, A., and Moloo, R.K. 2014. "An Innovative Measurement Programme for Agile Governance," *International Journal of Agile Systems and Management* (7:1), pp. 26-60.

Ram, P., Rodriguez, P., and Oivo, M. 2018. "Software Process Measurement and Related Challenges in Agile Software Development: A Multiple Case Study," *Proceedings of the International Conference on Product-Focused Software Process Improvement*, Wolfsburg, Germany: Springer, Cham, pp. 272-287.

Ram, P., Rodriguez, P., Oivo, M., and Martínez-Fernández, S. 2019. "Success Factors for Effective Process Metrics Operationalization in Agile Software Development: A Multiple Case Study," *Proceedings of the International Conference on Software and System Processes (ICSSP)*, Montreal, QC, Canada: IEEE, pp. 14-23.

Runeson, P., and Höst, M. 2009. "Guidelines for Conducting and Reporting Case Study Research in Software Engineering," *Empirical Software Engineering* (14), pp. 131-164.

Saldaña, J. 2021. *The Coding Manual for Qualitative Researchers*. London: SAGE Publications.

Scaled Agile Inc. 2021. "Metrics." Retrieved 29. January, 2022, from https://www.scaledagileframework.com/metrics/

Seaman, C.B. 1999. "Qualitative Methods in Empirical Studies of Software Engineering," *IEEE Transactions on Software Engineering* (25:4), pp. 557-572.

Stettina, C.J., and Schoemaker, L. 2018. "Reporting in Agile Portfolio Management: Routines, Metrics and Artefacts to Maintain an Effective Oversight," *Proceedings of the 19th International Conference on Agile Software Development (XP 2018),* J. Garbajosa, X. Wang and A. Aguiar (eds.), Porto, Protugal: Springer, Cham, pp. 199-215.

Sutherland, J., and Scrum Inc. 2021. "The Scrum at Scale Guide - the Definitive Guide to the Scrum@Scale Framework." Retrieved 29. January 2022, from https://www.scrumatscale.com/scrum-at-scale-guide-online/#Metrics-and-transparency

Talby, D., and Dubinsky, Y. 2009. "Governance of an Agile Software Project," *Proceedings of the ICSE Workshop on Software Development Governance*, Vancouver, BC, Canada: IEEE, pp. 40-45.

Uludağ, Ö., Kleehaus, M., Caprano, C., and Matthes, F. 2018. "Identifying and Structuring Challenges in Large-Scale Agile Development Based on a Structured Literature Review," *Proceedings of the 22nd International Enterprise Distributed Object Computing Conference (EDOC)*, Stockholm, Sweden: IEEE, pp. 191-197.

Uludağ, Ö., Philipp, P., Putta, A., Paasivaara, M., Lassenius, C., and Matthes, F. 2021a. "Revealing the State-of-the-Art in Large-Scale Agile Development: A Systematic Mapping Study," *arXiv preprint arXiv:2007.05578*).

Uludağ, Ö., Putta, A., Paasivaara, M., and Matthes, F. 2021b. "Evolution of the Agile Scaling Frameworks," *Proceedings of the International Conference on Agile Software Development,* P. Gregory, C. Lassenius, X. Wang and P. Kruchten (eds.), Virtual Event: Springer, Cham, pp. 123-139.

Van Oosterhout, M., Waarts, E., and van Hillegersberg, J. 2006. "Change Factors Requiring Agility and Implications for It," *European Journal of Information Systems* (15:2), pp. 132-145.