

Received 4 December 2022, accepted 27 December 2022, date of publication 5 January 2023, date of current version 11 January 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3234424

## APPLIED RESEARCH

# A Systemic Approach to Evaluating the Organizational Agility in Large-Scale Companies

MERAB GOGICHATY<sup>1</sup>, VLADIMIR IVANOV<sup>2</sup>, ARTEM KRUGLOV<sup>ID 2</sup>,  
WITOLD PEDRYCZ<sup>ID 3</sup>, (Life Fellow, IEEE), ALIYA SAMATOVA<sup>4</sup>,  
GIANCARLO SUCCI<sup>5</sup>, (Member, IEEE), AND RAPHAEL VALEEV<sup>6</sup>

<sup>1</sup>AK Bars Bank, 420066 Kazan, Russia

<sup>2</sup>Institute of Software Development and Engineering, Innopolis University, 420500 Innopolis, Russia

<sup>3</sup>Electrical and Computer Engineering Department, University of Alberta, Edmonton, AB T6G 2R3, Canada

<sup>4</sup>Innopolis University, 420500 Innopolis, Russia

<sup>5</sup>Department of Computer Science and Engineering, University of Bologna, 40126 Bologna, Italy

<sup>6</sup>AK Bars Digital Technologies, 420066 Kazan, Russia

Corresponding author: Giancarlo Succi (g.succi@unibo.it)

This work was supported by the Russian Science Foundation under Grant 22-21-00494.

**ABSTRACT** This paper presents action research to analyze an approach for assessment of the alleged agile transformation. This approach was implemented at AK Bars Digital Technologies, an IT spin-off of one of the largest banks in Russia using the Scaled Agile Framework. The approach is based on the Goal-Question-Metric approach, non-invasive measurement collection, and systemic analysis. It uses data from several different sources, including interviews, code repositories, user ratings in the play stores, and templates for agile assessment. The effectiveness of the approach is subjectively validated by the adoption of the proposed recommendations by the banks' senior management. Details are provided on the approach, the required effort from the side of both those assessing and of the people being assessed and the results. The final part of the paper is devoted to the discussion of its generalizability and the plan for future experimentation and refinement.

**INDEX TERMS** Action research, systemic analysis, assessment, agility, software development, software professionals, measurement.

## I. INTRODUCTION

Agile development was introduced around two decades ago. Nowadays more and more companies are claiming to successfully adopt this approach, making it a mainstream [1], [2], [3].

In fact, most companies are still exhibiting a “traditional” plan-based approach superficially implementing an agile methodology [4]. Some studies indicate that “pretending” to be agile is common at the organizational level (in 59% of the cases [5]) and at the team level (in 35% of the cases [6]).

The associate editor coordinating the review of this manuscript and approving it for publication was Mahmoud Elish<sup>ID</sup>.

Implementing agile methodology in large enterprises can be challenging because of mature cultures, complex infrastructures, inconsistent outdated practices and processes across teams, etc. [1]. For this reason, companies' managers often resort to “experts” or have assessments performed by independent entities [7], [8], [9].

The assessment of any software development process is a wicked problem, because there is no objective criteria of “goodness” or “correctness” of a particular process. In general, one can claim the process implementation effective when it yields the expected result [10]. Some researchers define the effectiveness of a development process as the ability to produce software with desired functionality and

quality in time [11]. In this paper we define and assess the overall efficiency as the conformity of the actual process with its design. It requires an analysis of the overall culture of an organization and the mindset of its stakeholders and employees, and how such culture and mindsets are translated in the principles guiding software development, and then to the daily practices [12]. There are several approaches to assess agile frameworks and their implementation:

- checklists,
- adoption models,
- transformation models, and
- frameworks' analyses tools.

**Checklists** are used to determine the presence or absence of some practices without deep analysis of underlying principles. They are adequate for the personal self-assessment and team's brainstorming while doing retrospectives [13], [14], [15]. They are inadequate, however, for a comprehensive analysis at the overall organizational level [16]. Also, checklists are usually tailored to specific agile methods and are hard to generalize.

**Adoption models** are step-by-step approaches to verify the presence of specific agile practices and guide efforts towards their implementation. These models ignore the overall big picture of the company, which is both their strength and weakness [17], [18], [19], [20], [21], [22]. It is their strength because they are quite simple. Since they focus on individual practices, they can be used across agile methods. It is their weakness as agility is not just a set of practices but a comprehensive and pervasive model of organization [23].

**Transformation models** do not only focus on agile practices, but also consider agile values, principles, and interrelations between them in the organization. This assessment is more complex than the previous one, and so far only few approaches exist, such as the *Comparative Agility Framework*, which compares the agility of a company with other agile organizations [24]. Another model, *Integral Agile Transformation Framework* [25], provides an exhaustive analysis of a company by assessing it from four different perspectives: mindset of leaders, organizational culture, practices and competencies, and organizational architecture. This model helps to evaluate the true level of agility but requires full support and involvement of the top management in the analysis process.

**Frameworks' analysis tools** [16], [26] are not related to the assessment of agile framework implementations, i.e. real case studies. Their aim is to examine the framework itself, its compliance with agile values and principles and the effectiveness in terms of meeting the expected outcomes [23].

The problem addressed in our study is inadequacy of the existing approaches due to their strong reliance on the interviews as the primary (and in most cases only) source of information. Interviews may indeed provide a substantial amount of information, however, they require a *reasonable* number of subjects. In order to avoid issues, the interviews are need to be conducted by professionals because respondents tend to provide biased answers based on their expectations

of the outcome of a questionnaire [27], [28]. Therefore, it is advisable to complement interviews with other approaches as demonstrated in [7] and [9].

In the past, we drafted a possible approach to such assessment for small operations based on the GQM (Goal, Question, Metrics) model and systemic theory [7]. In this paper, we present a refined approach suitable for large companies and we analyze its implementation on the IT spin-off of a bank in the form of an action research [29].

Altogether, the key contribution (i.e., novelty) of our work is as follows:

- it proposes an approach to assess the effectiveness of an agile process in use in a large organization,
- it experiments the approach on a real industrial case and validates its results with the followup actions undertaken by the analysed organization.

This paper is organized as follows. In Section II we provide a brief overview of Scaled Agile Framework (SAFe) for those not familiar with it, we detail the context of the assessment, including the history of the company where the assessment was run, AK Bars Digital Technologies, and the rationale for the assessment. In Section III we describe the methodology of the assessment, which is based on systemic analysis and a connected GQM model. The GQM modelling itself is presented in Section IV. The the procedure and results of the assessment, specifying the outcomes of each stage, tools, and methods applied, are described in Section V and Section VI consecutively. The final stage of the assessment—systemic analysis—is described in Section VII. The results of the assessment are summarized in Section VIII. In Section IX we analyze the effectiveness and quality of the assessment, giving some additional remarks on the validity of the overall research in Section X. Eventually, in Section XI we draw conclusions and outline the future research.

## II. BACKGROUND

### A. BRIEF OVERVIEW OF SAFe

This section provides a short introduction to SAFe. It is based on the set of principles, mindset and four core values: alignment, built-in quality, transparency and program execution. The implementation of this framework is based on the seven core competences [30]:

*Team and Technical Agility* describes critical skills and Lean-Agile principles and practices for Agile teams to create high-quality solutions.

*Agile Product Delivery* defines an approach to build and release a continuous flow of valuable products.

*Enterprise Solution Delivery* describes how to apply Agile principles and practices to develop large sophisticated software and cyber-physical systems.

*Lean Portfolio Management* aligns company's strategy and execution by applying systems thinking.

*Organizational Agility* allows the company to change quickly in order to respond to marketing challenges and opportunities.

*Continuous Learning Culture* describes values and practices that encourage individuals to continually increase knowledge, competence, performance, and innovation.

*Lean-Agile Leadership* helps leaders develop leaners' ways of thinking and operating so that team members learn from their example, mentoring and support.

SAFe targets the whole software development process, organizing it into four levels (Figure 1): a) team, b) program, c) stream, and d) portfolio.

The **team level** is where the actual development takes place. At this level SAFe adopts the same roles as Scrum: Product Owner, Scrum Master, and Developers [31].

The **program level** integrates the different teams, who release their work incrementally in 8-12 weeks long sprints with the metaphor of the *train*. This level introduces additional roles:

*Product Management* provide guidelines to all the product owners working for individual teams.

*Release Train Engineer* acts like chief impediment officer for all the teams in the Agile Release Train (ART).

*System Architect* who defines, communicates and shares an architectural and technical vision for a comprehensive release.

*Business Owner* is the key stakeholder who has the primary business and technical responsibility for a comprehensive release.

*System Team* assists in building and supporting the process of continuously delivering releases.

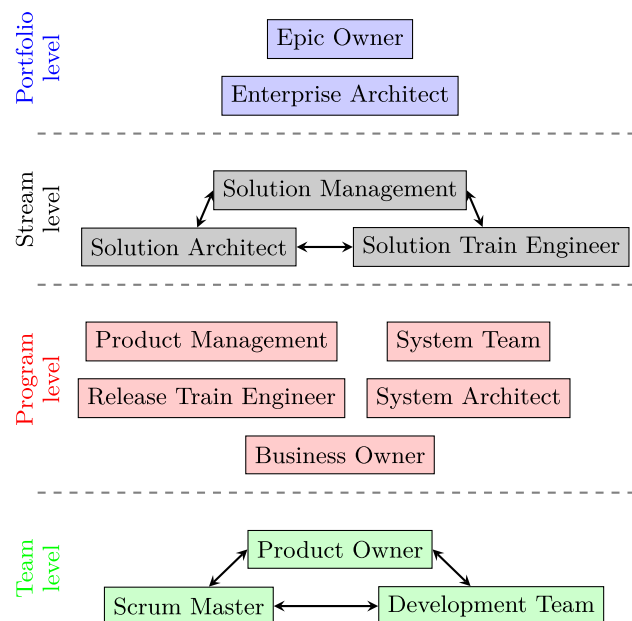


FIGURE 1. SAFe framework.

The **stream level** handles the management of self-standing products or product families, using traditional software

engineering jargon. Thus, self-standing products or product families are called “solutions” emphasizing the alleged customer-centric approach of SAFe. This level also includes the following roles:

*Solution Management* defines the vision and the roadmap for each solution.

*Solution Architect* is responsible for the overall architecture of each solution.

*Solution Train Engineer* synchronizes all releases coming from the teams in single value streams for the customers.

The **portfolio level** is where the general strategy of the company takes place by coordinating value streams. The strategy centered on the concept of “epic.” It includes the development and the deployment of one or more self-standing products or of a product line and generates a tangible economic value for the organization. To elaborate and finalize the portfolio, this level also features the following roles:

*Epic Owner* defines the overall view for the epic.

*Enterprise Architect* designs the overall architecture across streams, to maximize the overall economic value of the company [32].

Undoubtedly, SAFe and all scaled agile approaches are complex and time-consuming to implement. This complexity could be confirmed by existing anecdotal evidence [33], [34], [35] and by a Gartner survey, which shows that only 30% of 124 companies have considered SAFe for their operations. Out of these 30%, more than 12% of companies have already abandoned this framework. The other methods show similar or worse performances [3]. Furthermore, existing empirical analyses evidence that their main challenges include [36] and [37]:

- the lack of autonomy at the team level,
- difficulty in finding effective product owners,
- resistance to changing the development culture and to adapting to new roles, and
- integration with teams that have not yet transitioned to SAFe and are still using the traditional development methods.

The first two challenges refer specifically to SAFe and scaled agile approaches, while the rest can exist in the transition to any methodology. In some cases the problem does not lie in a particular agile method to adopt, but in the lack of effective control over the transition process [38], [39], which hinders the desired improvement [40].

The perceived lack of autonomy and the complexity of finding people matching the role of a product owner are topical issues which should be taken into careful consideration during our assessment.

Nevertheless, SAFe looks attractive for companies' senior management because firstly this framework covers the whole operations—from business to technical. Secondly, it standardizes planning and execution process synchronization between many agile development teams offering a comprehensive and strategic view of a company [41]. Moreover,

SAFe can be introduced quickly without changing the company's structure, being a compromise between agility and vertical organizational structure.

### B. AK BARS DIGITAL TECHNOLOGIES (ABDT)

AK Bars is one of the leading banks in the Republic of Tatarstan. It is ranked in the top 20 banks of Russia in terms of assets and the top 10 digital banks of Russia according to evaluation done by the Skolkovo and VR\_Bank rankings.

AK Bars Digital Technologies (ABDT) opened in Fall 2016 was given a task of developing and managing 6 main projects tightly integrated with the Bank's business [42]:

- *OK.NET* – a proprietary digital banking platform that provides a unified API for personal finance applications;
- *BankOK mobile wallet* and *AK Bars Online* – the online services based on OK.NET platform. Web and mobile services applications are also developed in ABDT;
- *Ak Bars Connect* – a service for direct sales that implements the concept of “Bank goes to business”;
- *Aimee* – a solution for the automation of the Bank's contact center. Artificial intelligence module allows the operator to quickly find the right answer or answer independently in the chat bot mode;
- *akbars.ru* – an updated website of the Bank based on a proprietary content management system;
- *Akbars 365* - a marketplace of non-banking services.

The company has developed its own alleged agile process, which includes prototyping user interfaces, software development, and the diffusion of the finished products.

ABDT is supervised by the Director of the Strategic Development of Ak Bars Bank. The company has 12 products and 2 service teams, tightly integrated with the Bank's business [42].

The reasons for the assessment were:

- aligning the activities of ABDT with the core business of the bank, also providing suitable metrics of success,
- responding effectively to the requests arising from the customers and from the variety of federal regulations,
- integrating better products developed by subcontractors who do not using agile methods, and
- increasing the satisfaction of those employees who were accustomed to non-agile approaches.

Consequently, the following questions were asked:

- Is SAFe implemented correctly?
- Does it fully respond to the needs of ABDT and Ak Bars Bank?
- What are the main causes of complaints from the end customers?
- What actions should be taken to improve the processes in ABDT?

### III. METHODOLOGY OF THE ASSESSMENT

This investigation is based on the paradigm of the action research as applied in Software Engineering [43]. Action research is typically based on two phases [44]: a) situation B) assessment, c) program intervention. This case is peculiar as

the “program intervention” is the execution of an assessment with a specific approach, a refinement of the one proposed in [7]. Our approach is built on a) the paradigm of systemic analysis, and b) GQM as the tool to derive both quantitative and qualitative measures.

In the following paragraph we shortly introduce the GQM and systemic analysis for those readers not yet familiar with this approach.

**Systemic analysis** is a method coming from anthropology, psychology, and sociology. It provides a comprehensive perspective on a person and on the organization based on the internal and external relationships in place [45]. The systemic analysis assumes that an organization cannot be simply decomposed on its components and tries to determine the presence of incoherence between the components and the whole [46]. A special kind of incoherence is the “schismogenesis,” which occurs when the organization promotes a behavior or a key performance indicator that makes it divergent in its goals from the organization. The more this component aims at achieving such key performance indicator, the more, in a positive feedback loop, such behavior is promoted by the organization and the more this divergence becomes significant.

**GQM (Goal, Question, Metrics)** [47] is a framework originally thought to define measures of software systems, which has then become a de-facto standard to define corporate goals and measures to verify the achievements of such goals [7], [48], [49], [50]. The framework is organized at three levels. The conceptual level defines the goal of a research (products, processes, or resources) and the reason for its study. The operational level defines the questions that characterize various aspects of the measurement object. The quantitative level defines a set of measurements that can be used to answer the questions.

To continue with the assessment methodology, the procedure consists of the following steps:

- 1) Conducting the discussion with an organization. By inviting its key stakeholders “on board,” we make them fully aware of the steps we are taking. Applying systemic analysis requires active and open participation of key stakeholders, as it involves time, effort, and mindfulness [51].
- 2) Designing the GQM model for the assessment involving the key stakeholders again and having them agreeing with it. In particular:
  - Defining the goals of the investigation with the stakeholders and formulating the fundamental working hypothesis.
  - Identifying the core questions and metrics, taking into account the available data. On the one hand, metrics are intended to understand how different people perceive at different levels in the organization, how the company is operating and whether their goals and the goals of the organization are aligned. On the other hand, metrics are intended to collect objective facts about the company's

operations that back the discussion coming from the individual interviews, grounding them to the reality, as it is recommended in systemic theory and in critical system heuristics [52].

- As a reference, we use common agile metrics, such as velocity, cycle time, fixed bugs, and so on.

3) Implementing quantitative analysis which includes:

- The collection of the metrics. The fact that we collect and analyze the data about the company's operations non-invasively in order to obtain unbiased results and ensure integrity of the data should be taken into account [53], [54]. The concrete set of metrics to extract has been already defined via the GQM, as discussed above.
- The analysis of preliminary results to understand if these collected metrics provide the results aligned with the GQM and whether there are any problems with the non-invasive approach to pave the ground for the followup systemic analysis.

4) Implementing qualitative analysis which includes:

- The elaboration of the interview script based on the GQM and recommended standards for open interviews [55], [56].
- The interviewing process should be organized in a way to minimize the risk of a researcher or respondents' bias [57], [58], [59]. In addition, the investigator triangulation technique is obligatory for interviewing process to ensure the validity of the process [60].
- The results analysis implementing the tools selected during the GQM modeling stage.

5) Implementing systemic analysis. At this step we determine the incoherence between components of an organization, by comparing results of quantitative and qualitative analyses with officially declared strategy, vision and culture of the company, paying a special attention to incoherence being "schismogenesis".

On the one hand, systemic analysis aims at confirming or rejecting the preliminary results, on the other hand, it aims at defining the alignment of individual company members with the goals of the company, especially with the numeric values coming from metrics associated with goals.

At this stage we finalize the results making final determinations and providing feedback to the organization who requested this assessment.

All the data were collected approximately in the three-month cycle during which ABDT did not have any substantial changes in its operation. In this sense, we can consider the data homogeneous and reliable.

#### IV. GQM MODEL DEVELOPMENT

Before the active assessment phase, preliminary meetings with the head of ABDT and the top management of Ak Bars Bank had been conducted. This step was crucial in terms of

schedule coordination and setting time, scope, and resource constraints of the assessment. The tangible output of this stage was the conceptual level of the GQM model elaborated together with senior managers of the bank. It resulted in the following goals:

- To assess the software development process in ABDT company
- To assess the quality of the BankOK and AK Bars Online 3.0 products

We shaped the following questions out of the goals:

- How can we compare the running software development process against SAFe framework?
- How can we verify the agility at the team and organizational levels?
- What are the available process and product artifacts?
- What does "product quality" mean for ABDT's primary stakeholders?
- How different is the current quality in comparison with the previous, before the introduction of SAFe?
- Are existing measurements adequate to describe fully the processes and the products of ABDT?

To clarify the operational level of the GQM model and understand which metrics are available at the quantitative level, we requested the following data from ABDT:

- current organizational structure and its evolution during the last two years, including personnel holding key positions,
- current processes formally in place and associated official documents,
- repositories of (some of these repositories could be combined): code (also legacy), issues (tracker), customer complaints, user stories, budget and its allocation, resources,
- recent and planned (or expected) decisions of the key stakeholders,
- analogous information from contractors providing code, if it exists,
- current reference metrics: which metrics, what is their usage, how metrics are collected, stored or aggregated, who makes a decision about metrics.

Finally, we developed a comprehensive GQM model for the ABDT assessment project considering the initial request and available data (Figure 2).

#### V. QUANTITATIVE ANALYSIS

To answer the product-centered questions of the GQM model (questions 4 and 5), we collected data from the repositories and performed its quantitative analysis. Records from the issue tracker, as well as customers' rankings in App Store and Google Play allowed evaluating software qualities: efficiency, usability, reliability, and functional suitability. Moreover, data from a project management information system contributes to the analysis of the processes with respect to teams' productivity and development pace [8].

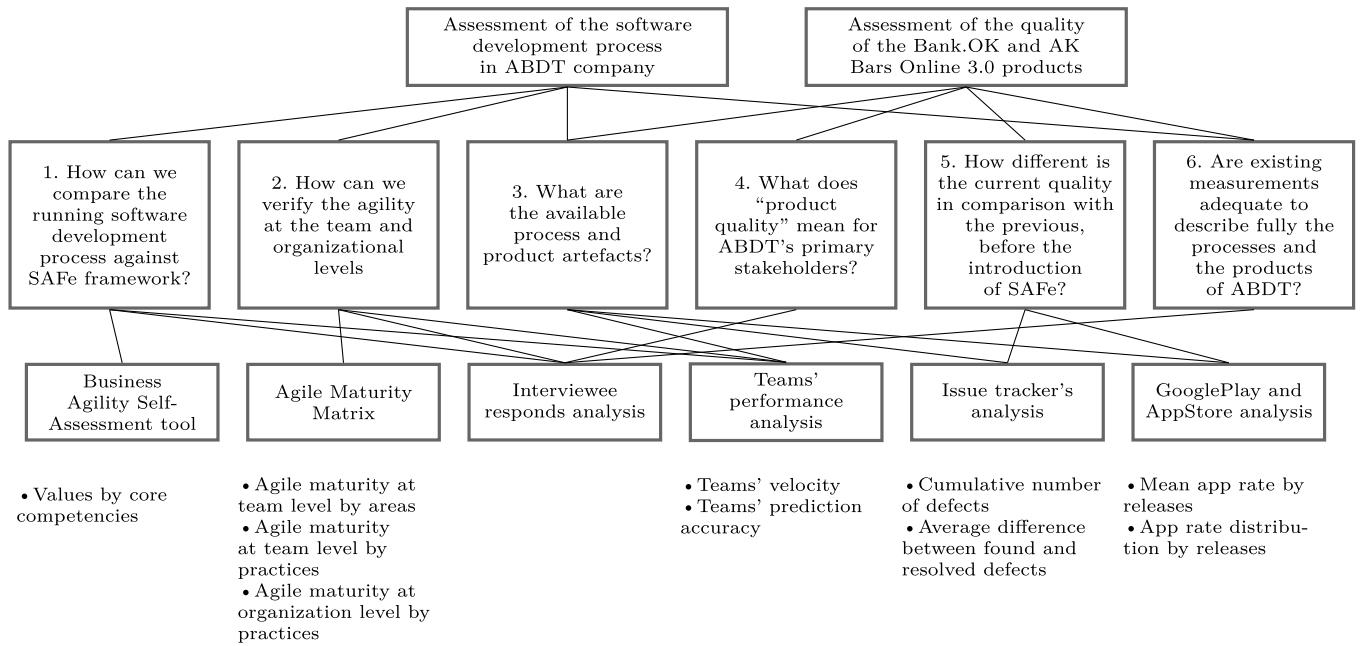


FIGURE 2. QGM model for the assessment.

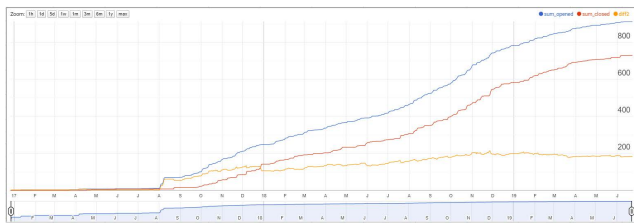


FIGURE 3. Cumulative number of bugs in products.

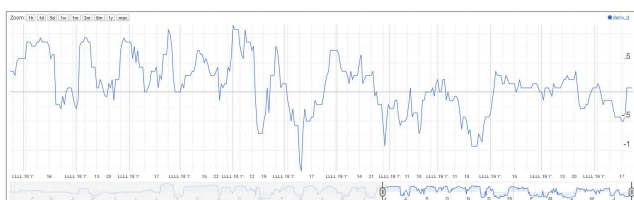


FIGURE 4. The average value of the difference between found and resolved bugs.

**A. DATA COLLECTION**

The data from issue trackers featuring the bug density is shown in Figures 3 to 4. We relied on these metrics as primary indicators of the software quality according to the developed QGM model.

As we already mentioned, user ratings at App Store and Google Play were considered as one of the criteria for product quality assessment. Figure 5 represents a summary of the ratings of the different app versions (the "stars").

The teams applied story points during sprint planning as a common approach for effort estimation. The teams visualized planned sprint activities both virtually, in information

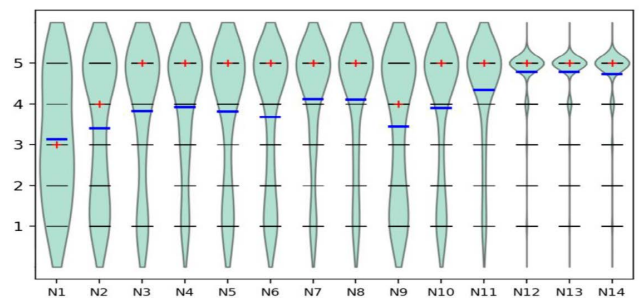


FIGURE 5. Analysis of mobile application rating. The thin (black) bars represent the discrete values of the stars, the thicker (blue) bar is the mean, and the (red) plus sign is the median.

systems (typically, Confluence [61]), and physically in "information radiators". Such "information radiators" as Kanban boards and task boards were implemented for each team and for the value stream in general. To assess the team's performance, we processed the data from the project management system. The indicators we selected to demonstrate the effectiveness and maturity of the teams were velocity (Figure 7) and prediction accuracy (Figure 6).

Thus, during the quantitative analysis, we examined common agile metrics, complementing them with customer satisfaction analysis. This data provided us with insights into the health of ABDT's processes and the current state of agile transformation.

**B. DISCUSSION**

We concluded that the difference between the number of discovered and fixed bugs was growing until 2018. Since 2019,

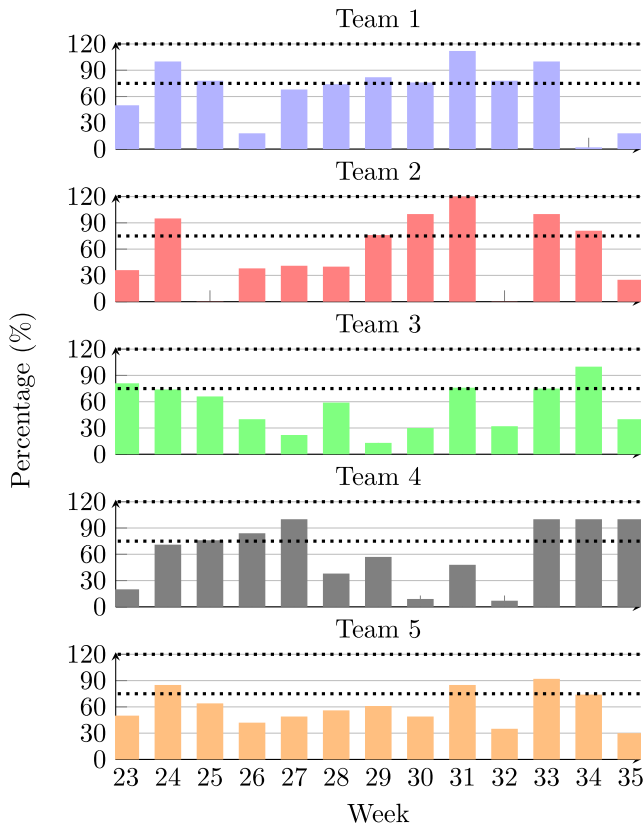


FIGURE 6. Teams' accuracy.

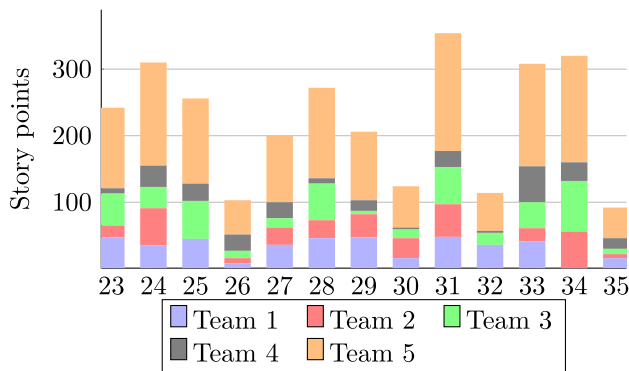


FIGURE 7. Teams' velocity.

when the SAFe framework was introduced, the situation has improved significantly, as shown in Figure 4. The parameter *deriv\_d* indicates the average difference between the number of new and fixed bugs in the range of two weeks (standard sprint length). Until 2018, the bugs were fixed after their mass detection. It led to the presence of a larger number of open and not resolved bugs in the code. Starting from 2019, the detection and resolving of bugs has not led to a dramatic increase in their number. It is worth mentioning that the analysis does not take into account the priority of bugs.

Considering the users' feedback, we concluded that the overall quality of the products has been improving since the introduction of SAFe framework in ABDT and currently

they fulfill high standards for usability and reliability. The mean rating has moved from 3,29 to 4.60—statistically significant at  $\alpha = 0.05$  as evaluated using the non-parametric Mann-Whitney U test [62]. The median has been bumped from 4 to 5—statistically significant at  $\alpha = 0.05$  as evaluated using the non-parametric Kruskal-Wallis test [63]. We have then analyzed the data with the Kolmogorov-Smirnov statistics [64], which assumes a value of 0.37 (rounded to the second decimal), statistically significant at the 0.05 level (actually, with a p-value less than  $10^{-70}$ ).

While analyzing the teams' dynamics, the accuracy of the planning was expected to be increased due to process continuous improvement, team maturity, and scope clarification. The acceptable variability, according to the Cone of uncertainty model [65], [66], lies in a range from 75% to 120% after several sprints. In the given case study, the accuracy is unstable for every team but fourth, which demonstrates 100% accuracy for the last 3 sprints (Figure 6). For the analyzed period of time, most teams were in the so-called "Wormhole of Uncertainty": the level of accuracy did not increase over time because of late discovered bugs or emergent requirements [67].

The indicator of well-established processes in the organization is stable team velocity with the tendency of a slight increase due to implementation of the continuous improvement principle. The fluctuation of the completed scope (Figure 7) may indicate at either poor planning with a strong commitment to the Definition of Done or scope creep and task switching due to the emergence of new critical tasks, assigned to the teams from above.

### C. OUTCOME OF QUANTITATIVE ANALYSIS

After conducting the quantitative analysis, we came up with the following hypothesis about the development process in ABDT.

- The company is at the early stage of transition towards SAFe, its processes immature both at teams and value stream levels.
- The teams have limited authority over requirements management.
- The quality of the products is increasing in terms of general usage scenarios.
- The quality of the products is insufficient for newly added features and uncommon usage scenarios.
- Communication between business and development is low.

### VI. QUALITATIVE ANALYSIS

We conducted a number of in-depth interviews with developers and managers, similar to previous experience [7]; however, this time we enriched the process by taking advantage of the Business Agility Self-Assessment tool [21] and Agile Maturity Matrices [68], which are well-known approaches easy to be accepted by the company. Two assessment tools used simultaneously allowed checking the organization's business agility both through the lens of SAFe framework

**TABLE 1. Characteristics of the interviewing process.**

People interviewed	15
Length of interviews	from 40 minutes to 1 hour and 15 minutes
Average length of interview	50 minutes
Interviewees' positions	PO, SM/RTE, Architect, developer, UX/UI designer, QA engineer, AppSec engineer, DevOps engineer, business owner, director

and against the general agile values and principles. In other words, we observed the agility level of ABDT at micro and macro levels.

The interviews were conducted by two researchers at a time. They were done in-person recorded with the permission of the subject and then the result was processed and sent back for verification. The information about the interviewing process is summarized in Table 1.

We tried to access those participants who were both playing key roles in different company positions and aware of the dynamics in place, from the team to top management level [69]. Overall, we interviewed 11 employees of ABDT (development teams' members, system architects, DevOps specialists, product owners, and RTEs) and four employees of AK Bars bank (top managers, business owners). Then we processed the answers to match them to specific categories of the tools mentioned. To avoid biased responses, the script of the interview did not provide one-to-one matching between questions and tools' practice areas; however, there was a strong interconnection between them.

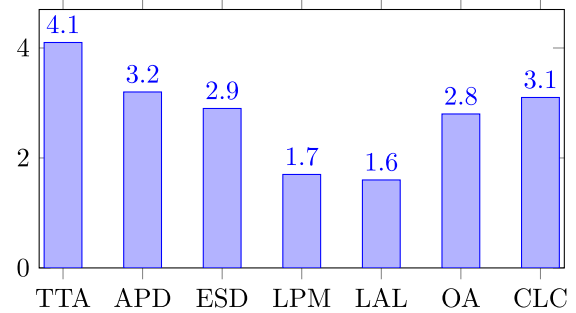
It should also be noted that all ABDT employees have been trained in the basics of SAFe as part of the SAFe transformation. Therefore, it is expected that they are familiar with the terms, concepts, and ideas that form this framework prior to the interview.

**A. ADHERENCE TO SAFe**

During interviews, we analyzed employees' perception of ABDT's adherence to the SAFe framework. From the interviewees' responses, we extracted the descriptors of key SAFe principles according to Business Agility Self-Assessment tool. Then we evaluated the implementation level of each principle, thus obtaining scores for the SAFe practice areas:

- TTA—Team and Technical Agility,
- APD—Agile Product Delivery,
- ESD—Enterprise Solution Delivery,
- LPM—Lean Portfolio Management,
- LAL—Lean-Agile Leadership,
- OA—Organizational Agility, and
- CLC—Continuous Learning Culture.

The overall adherence to SAFe practice areas is shown in Figure 8. The comprehensive exploration of the evaluation in each area is provided in Figure 9.



**FIGURE 8. Analysis of the compliance with SAFe framework by the areas. Score ranges from 0 (absence of the practice) to 5 (full implementation).**

The results of this analysis show moderate maturity of the mechanical adoption of the prescribed practices (TTA and APD areas), and the growth of the learning culture within the organization (CLC area). On the other hand, low scores of management and leadership behaviors (LPM and LAL areas) may indicate a violation of the agile values and principles. For the sake of a thorough analysis of this inconsistency, we performed the additional assessment of the processes' agility at organizational and team levels.

**B. ANALYSIS OF AGILE MATURITY AT THE TEAM LEVEL**

Alongside with the assessment of adherence to SAFe, we perform the overall agility assessment, regardless of the specific framework. The high-level summary analysis at the teams' level is given in Figure 10.

The analysis of the agile maturity by areas shows a sustainable and stable implementation of agile practices and principles in all dimensions but one—product. Detailed analysis of key processes and cultural characteristics by areas for teams' level is given in Figure 11.

In the following, we discuss each area in detail.

**1) TEAM DYNAMICS**

This area characterizes how well team members understand and adhere to the agile methodology; their motivation to work precisely on the tasks given, adherence to the rules and stable work at a given pace. Based on the analysis, the following conclusions can be drawn:

- Teams exist in the current composition from 6 months to 1.5 years, being at norming or performing stages of Tackman's scale.
- The majority of the teams follow the mechanics of the SAFe; there is a steady positive dynamic of its implementation. Nevertheless, agile is not yet implemented as a holistic approach.
- Teams share the adopted regulations and principles, while some team members may ignore individual regulatory procedures and generally do not have sufficient motivation to implement and support the SAFe.

**2) TEAM ENVIRONMENT**

Team environment indicates the level of commitment of team members; the quality of the team's structure in terms of its



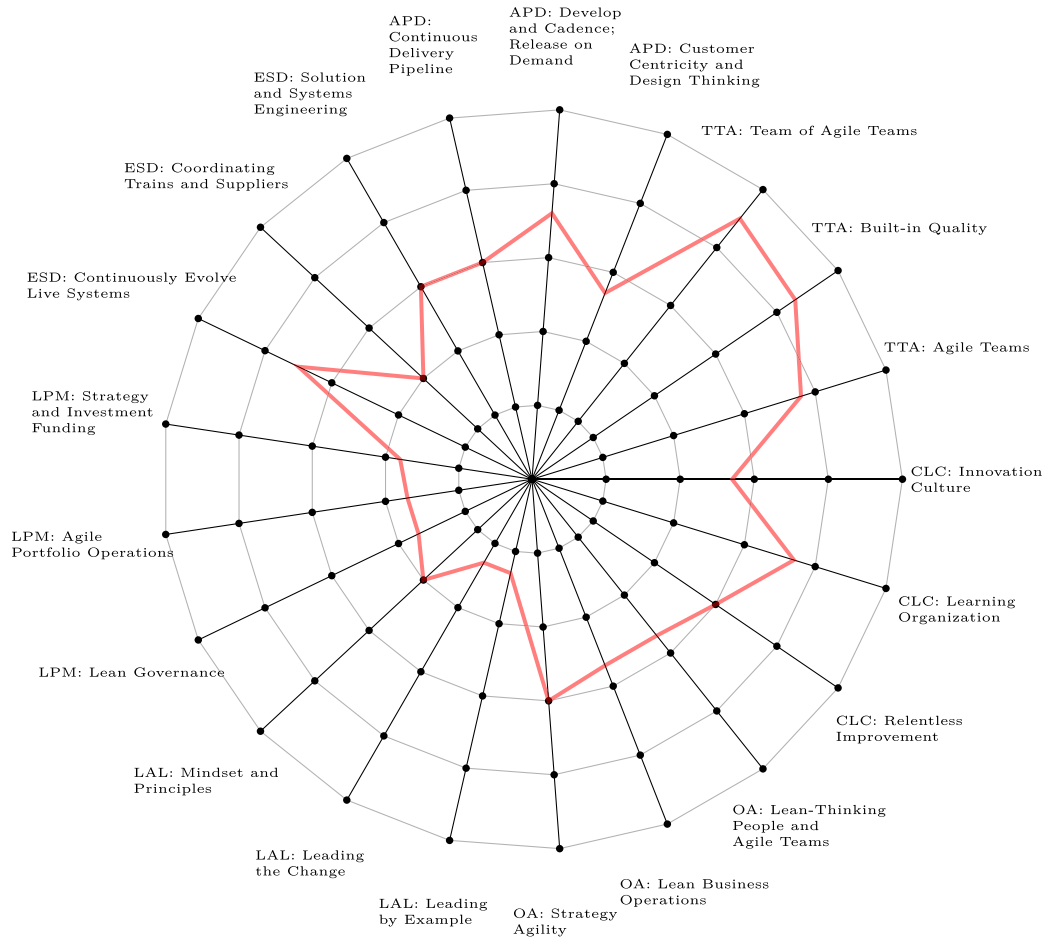


FIGURE 9. Detailed analysis of the compliance with SAFe framework.

consistency and availability of necessary competencies; the authority level of a team as well as the existence of barriers and interference. Based on the analysis, the following conclusions can be drawn:

- Teams work in a common open workspace, which favorably affects their effectiveness and provides osmotic communication.
- The process of identifying and resolving dependencies between teams at the level of quarterly planning has been built, while the situational identification of dependencies and risks is not carried out explicitly.
- Roles in the team are clearly distributed, competencies do not overlap (I-shaped specialists). Working environment do not support the formation of cross-functional teams, which increase dependencies and overall risks.
- There is a personnel shortage, which hinders the implementation of some key practices like peer review. The problem is solved by informally combining competencies into tribes, which should be regulated.
- Teams do not have a sufficient level of authority to make decisions on a product development strategy.

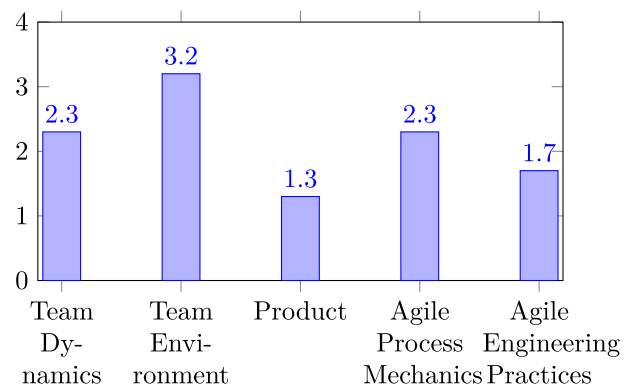


FIGURE 10. Analysis of the agile maturity at the team level by areas. The scale indicates the following levels of maturity: 0 - Impeded, 1 - In Transition, 2 - Sustainable, 3 - Agile, 4 - Ideal.

### 3) PRODUCT

Product is a factor indicating team performance, time from idea to its implementation. We analyzed a working product management strategy at a team’s level and assessed how all team members visualize the product [70]. This area covers aspects of the requirements engineering process by analyzing

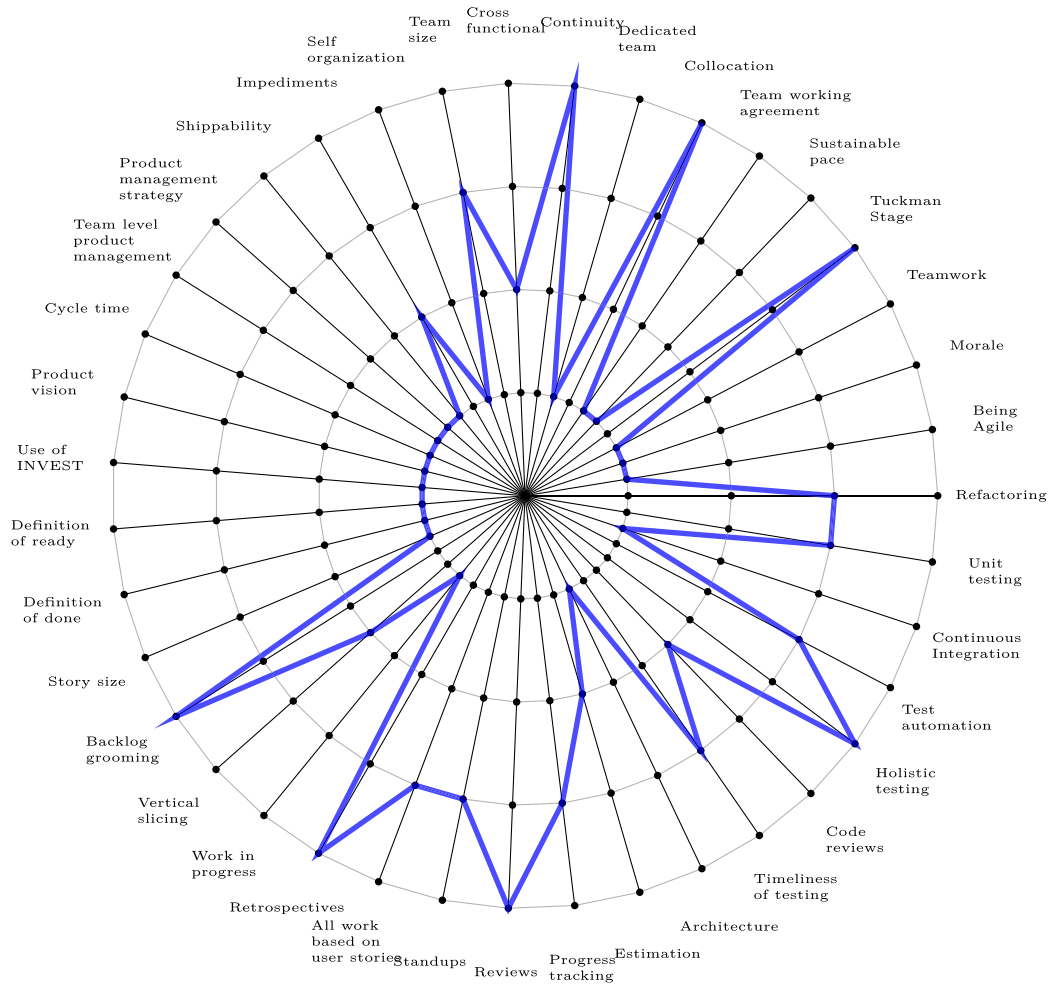


FIGURE 11. Assessment of the agile maturity at the teams' level.

the implementation of the user stories, “Definition of Ready” and “Definition of Done” concepts. The following conclusions can be drawn:

- There is no common understanding of terminology both within teams and between teams and stakeholders.
- The main metric for assessing the effectiveness of teams is the performance per iteration (velocity), the cycle time metric is used by a limited number of teams.
- Teams do not have control over a number of assigned key performance indicators (KPI), which negatively affects their morale [71].

#### 4) AGILE PROCESS MECHANICS

The mechanics of the process is a factor that indicates the alignment with the common agile events like planning, synchronization, review, and retrospective meetings. The format of the progress tracking is also considered in this area. Based on the analysis, it can be concluded that most team members understand the benefits of the agile process mechanics and follow them on the regular basis.

#### 5) AGILE ENGINEERING PRACTICES

Engineering practices are a factor that indicates the maturity of design and programming culture and CI\CD processes automation. Based on the analysis, the following conclusions can be drawn:

- The refactoring process is the cultural norm of the organization; additional activities to improve product quality have been introduced.
- Introduction of the so-called “Feature Freeze” activity for exhaustive testing and refactoring indicates the improper usage of Definition of Done and the absence of test-driven development.
- Due to the narrow specialization of team members, peer review and team review have been reduced within teams.

### C. ANALYSIS OF AGILE MATURITY AT THE ORGANIZATIONAL LEVEL

Analysis of the organizational agility gave us the following characteristics of ABDT structure (Figure 12):

### 1) ORGANIZATIONAL STRUCTURE

Top management of ABDT understands that building a process around products, teams, and delivery is more effective; however, there is a hierarchical functional vertical at the moment.

### 2) FUNDING

The budget is formed on an annual basis for the whole stream, but there is no correlation between the funding and the formation of teams within streams. A comprehensive statement of work with detailed estimates required prior to funding, and project success depends on fully implementing all requirements.

### 3) PROJECTS IN PROGRESS

Top management performs the project's prioritization annually during funding sessions. An intake process has been established to determine the project alignment with goals. There is no clear understanding of how many projects can be in progress at a time.

### 4) METRICS

Technical and business metrics are not properly correlated; moreover, business metrics are not detailed enough and product metrics are based on high-level increments (epic stories) that do not provide an accurate evaluation of the current progress.

### 5) COMPENSATION

Contrary to agile values, teams cannot influence the performance of some KPIs and perceive them as an inequitable penalty.

### 6) IMPEDIMENTS

Dependencies are proactively recognized during the program increment planning sessions and actively resolved by teams within the stream.

### 7) TOOLS AND TECHNOLOGY

Teams implement both high-tech (JIRA, Confluence) and high-touch (task boards, stickers) tools as information radiators. However, they are used in a limited format, a culture of the proper tools usage in the agile environment is being implemented.

### 8) AGILE ADAPTION TRACKING

The main metric for evaluating the effectiveness of implementing agile processes is Time to Market. The benefits of the transition towards the SAFe are not obvious for the management of the company.

### 9) AGILE ADOPTION GUIDING COALITION

ABDT has implemented steps 1-3 of the John Cotter change model [72].

### 10) MANAGEMENT OF AGILE TEAMS

Teams are managed as functional units, not product ones. Product decisions are often forwarded by different stakeholders in a directive manner, bypassing the product owner. Self-organization is not encouraged.

### 11) BUSINESS/DEVELOPMENT RELATIONSHIP

There is no mutual desire for collaborative work within agile ideology.

### 12) EXECUTIVE SUPPORT

The bank's top management support the implemented development model but poorly understand the ideas and values of agile. It results in the misconception of agile adoption and agile transformation at the organizational scale

### 13) MIDDLE-MANAGEMENT SUPPORT

Middle-level managers of the bank's business department prefer the functionally-oriented model because it is more familiar to them.

### 14) TEAM MEMBER MANAGEMENT SUPPORT

Teams support the agile practices and are trained to work effectively according to the chosen methodology, although a few employees tend to the traditional development model.

### 15) MULTI-TEAM SYNCHRONIZATION

The teams' work is synchronized. There are common retro, demo, common iteration start/end times for the whole stream. There is a practice of release during the iteration.

## VII. SYSTEMIC ANALYSIS

The next step in our assessment is analyzing the company as a "system," focusing on understanding whether the company is really overall aligned toward SAFe, as it claims to be, or there are situations when the company promotes behaviours or uses KPIs that make the operation diverging from its goals.

Using the outcome of quantitative and qualitative analyses, we were able to compare principles and values declared by the company with actual processes and behavior. We used official documentation of ABDT as a source of declared culture and processes in place. Such an analysis is the key idea of the systemic approach which helps to highlight the inconsistencies which can be in form of:

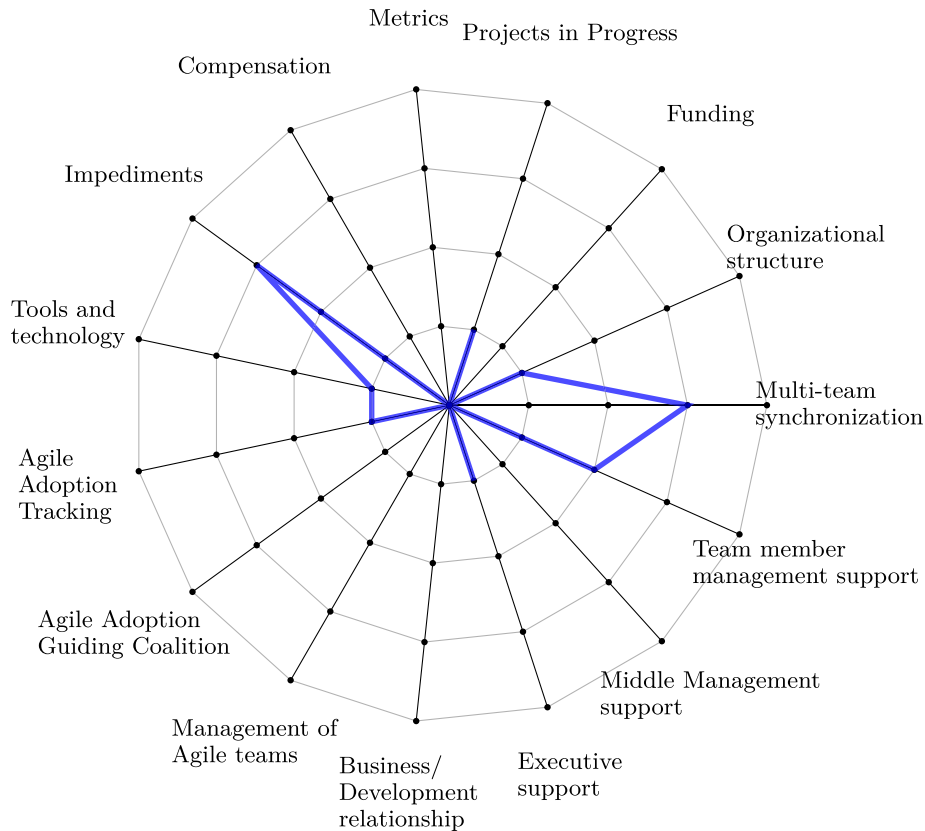
*Misalignment*, when inconsistency resulted in a significant negative impact on the company but the situation is stable,

*Schismogenesis*, when inconsistency has a cumulative negative effect and diverges over time.

The result of the analysis is given in Table 2.

The systemic analysis has confirmed the initial hypotheses and highlighted the most urgent problems which inhibit the effectiveness of the agile process implemented in ABDT:

- Schismogenic relationship between the business and development sides as a result of the contradiction



**FIGURE 12.** Assessment of the agile maturity at the organization level. The scale indicates the following levels of maturity: 0 - Impeded, 1 - In Transition, 2 - Sustainable, 3 - Agile, 4 - Ideal.

between a functionally oriented approach and the agile values and principles.

- The lack or inappropriate usage of metrics to assess the effectiveness of the established processes: misalignment of measurement system with objectives and responsibilities of the roles involved in the development process.

**VIII. RESULT OF THE ASSESSMENT**

Finally, we were able to decide on the formulated hypotheses:

- Organizational agility indicates the early stage of transition towards SAFe with significant impediments from the middle management. However, basic mechanics of the agile process are performed properly both at the team and stream levels. Moreover, there is sustainable support of the ideas and values of agile development at the team level.
- Teams act like functionally oriented units rather than product- or feature-based ones due to the lack of authority and empowerment. Moreover, the existing system assumes business metrics for evaluating teams’ performance that they cannot influence. As a result, the motivation among the team members is low.
- Implementation of testing (unit and integration ones), code reviews, and refactoring helps to discover and fix bugs thus increasing the quality of the products.

- On the other side, the absence of Definition of Done and TDD—practices that guarantee the reliability of product increment—confirms the insufficient quality of newly added and personalized functionality.
- Communication between business and development is highly disrupted: both parties misunderstand and mistrust each other. It results in frequent tensions and other symptoms of an uncontrollable rivalry.

The results of the assessment were presented to the key stakeholders of ABDT and AK Bars bank and they appeared to be endorsed by the head of ABDT. The bank’s management agreed with the strategy for the further transition of the ABDT toward the SAFe methodology and the implementation of the suggested corrective actions. Thus, the proposed approach to the assessment of a company’s agility has shown its effectiveness in the given case study. It is approved by the customer’s resolution. Giving the practical and pragmatic nature of the described method, we would recommend its appli-ance for the analysis or audit procedures within large-scale organizations.

**IX. QUALITY OF THE RESEARCH**

As mandated by action research [44] and as already mentioned in Section III, we now need to perform an assessment of our action, which was itself an assessment.

**TABLE 2. Inconsistencies revealed by the systemic analysis.**

Declared values and principles	Actual behavior	Type of inconsistency
The company has a flat structure. There are no team leaders who distribute tasks, and no managers who control the order of the team's work	The company has hierarchical and functional-oriented structure	Misalignment
There are several separate and motivated self-organizing teams focused on primary business goals	Working environment do not support the formation of cross-functional teams; employees are selected into teams by the HR department based on necessity, not their preferences	Misalignment
The decision-making facing business problems is completely run by teams	Teams do not have any authority to make decisions on a product development strategy; lack of control over their work negatively affects teams' morale	Schismogenesis. The more tasks assigned to the teams from above the less commitment to the product and sprint goals which results in continuous divergence with agile culture [70]
At least once every two weeks, each team together with the business decides what needs to be worked on next	There is no mutual desire for collaborative work within agile philosophy, middle-level management tends to the functionally oriented model	Schismogenesis. Managers of the bank act from the position of traditional project management. The more directive discussion, the more latent dissent exhibit developers [12]
The teams use three types of metrics that are displayed on special information panels. They pay attention to product metrics, system health metrics, and process metrics (production metrics)	Business metrics are not detailed enough: product metrics are based on high-level increments and Time-to-market, team performance assessed by velocity. Teams do not have control over a number of assigned KPI	Schismogenesis. The measurement system is not aligned with the teams, their goals and responsibilities, which leads to frustration. The more metrics are introduced, the less accurate they are collected and the less transparent the development process becomes [71]
In addition to INVEST, a number of criteria were developed for the Definition of Ready of User Story	Most of them unaware of INVEST criteria at all. A task can be assigned at any moment throughout the sprint obviously violating Definition of Ready	Misalignment
Definition of Done defines a checklist of what needs to be done in order for a task to be considered complete and ready for release	Improper usage of Definition of Done and omitting of this practice by some teams	Misalignment

We analyzed the validity of the assessment based on the following criteria to its theoretical and practical implications [73]: a) veracity—the credibility of the results, b) consistency—the dependability and reliability of the research, c) neutrality—the presence of any bias among observers that has an impact on the interpretation of the results, d) applicability of the results.

#### A. VERACITY

This characteristic refers to the degree of accuracy and precision of the research. To enhance the veracity of systemic analysis, we implemented the triangulation technique, which is the usage of two types of questionnaires, and conducting and analysis of interviews by several researchers.

The sample size may be insufficient for statistical measurement, and the lack of randomization is the limitation of the assessment. To soften this factor, we performed the matching of candidates for interviews based on their roles and teams. As a result, we have interviewed representatives of each team in the value stream as well as business unit representatives and top management (Table 1).

In case of quantitative analysis, we can talk about applying a time series experiment, where the kickoff of the SAFe transformation is the treatment. According to the characteristics of this quasi-experimental design [74], history is the only potential threat to internal validity. To some extent it can

be controlled by analysis of several teams; still, this would mitigate this threat but not eliminate it.

#### B. CONSISTENCY

Consistency refers to dependability and reliability. The systemic analysis can be traced back to the number of artifacts: interview scripts, recordings, and transcriptions. This comprehensive documentation stands for the high dependability of the approach.

Consistency, stability, and repeatability of quantitative analysis are approved by the reliability of the data collectors (information systems) and procedure of the measurement (statistical analysis). The only weakness is the process of collecting data, which involved manual data input in Confluence. In the future, this vulnerability can be solved by applying automatic non-invasive data collection [75], [76].

#### C. NEUTRALITY

This aspect of research quality is mostly referred to confirmability in qualitative research, as far as quantitative research relies on statistical analysis, thus having a high level of objectivity. Whether the researchers have a priori assumptions that may bias the implementation of the study? The fact that the results of quantitative and systemic analyses complement each other characterize the company's processes and culture in a similar way, decreases the possibility of such bias.

Moreover, quantitative research and interviews conducted by different people also decreases the risk of biased evaluation.

#### D. APPLICABILITY

Will the results be similar for other populations in the given context of large-scale enterprises? We believe that the problems revealed in this study are typical to some extent to any large-scale enterprise, which is approved by a number of relevant examples [77].

The nature of the proposed assessment presupposes its deep customization, starting from the elaboration of GQM model and further along with quantitative data analysis and systemic analysis. Thus, this approach is vulnerable to external validity threats:

- Reactive effects of experimental arrangements, which involves the inability to generalize from the setting where the experiment occurred to another setting.
- Interaction of selection and treatment, which involves the inability to generalize beyond the groups in the experiment.
- Interaction of history and treatment, which involves the inability to generalize findings to past and future situations.

However, the assessment method itself can be applied to another contexts: its limitations depend on the expertise of an assessment team only. Several types of potentially problematic design decisions which the team has to make are sampling, data collection, analysis, interpretation, and presentation.

It is worth noting that the proposed approach has no limitations with respect to the size of the analyzed company, the point is the rationality of its application on a small scale. For small companies the evaluation can be simplified to the evaluation of a single team with a focus on the applied framework (usually Scrum). This issue has been extensively studied from theoretical and practical points of view [78], [79].

#### X. VALIDITY OF THE ACTION RESEARCH

The general notion of validity is becoming more complex nowadays, and we therefore need to be aware of the validity of the approach we are proposing. We evaluate the validity of the research study as a whole based on the ideas suggested by M. Straton in his work on application action research in software engineering [80]. Thus, the validity considered separately from the lenses of a) construct validity, b) internal validity, c) conclusion validity, and d) external validity.

From this perspective, veracity and neutrality resembles somehow to *internal validity*. Nevertheless, the role of the observer should be further considered. In our case, the observers are independent people not part of the organization. They attend meetings, help develop the GQM, and collect and study the data, perform the interviews, and help to identify misalignments and scismogenesis. In these contexts having external observers instead of using people of ABDT for performing these tasks is the best insurance against biased deductions.

The *construct validity* is referred to consistency. Still, it refers to the assessment and not, for the time being, to the overall action research. Given the single instance of experimentation that is being considered, most of the effort is concentrated on how the data is collected [81], [82]. To this end we follow the approach proposed by Turnock and Gibson [83] and we consider the following critical aspects:

- Open or hidden data collection, with respect to the people involved, and associated possible interference with the results. All the collection of data is completely open to employees of the company, including the non-invasive data collection, due to the legal environment in which ABDT operates, which prevents hidden collections of private data of employees. Moreover, it would be impossible to conduct a deep analysis, like the one required for systemic analysis without explaining its goals and needs to all the people involved [84].
- Mode and timeline of data collection. This aspect was thoroughly discussed while describing the methodology of the assessment (Section III)

Please note that rather than a “fully fledged validation” the proposal by Turnock and Gibson is a check of self-consistency of an action research, the only that is possible in these contexts [83].

The *external validity* of the research is addressed by the analysis of the applicability of the proposed approach.

The *conclusion validity* refers to drawing correct conclusions from observations. In this work, it is guaranteed by two core elements of the assessment methodology: the GQM model and systemic analysis. The former refers to setting properly the objective, the latter ensures that the right causes are distinguished from the observations. [80]

#### XI. CONCLUSION AND FURTHER RESEARCH

The transition from the traditional organizational culture to agile is tremendously difficult. Many organizations claiming that they have completed successfully an agile transformation, in fact, have just glued a variety of agile practices to the existing traditional processes and culture [85], [86], often worsening the overall organizational performances [4]. However, true agile transformations have the potentials of carrying significant strategic benefits, including increased productivity and quality and also positive changes in mindsets, behaviors, structures, and policies. However, it is not easy to evaluate such changes and the proper assessment of the agile transformation is another complex task that does not have (yet) a standardized solution.

In this paper we have presented action research centered on the assessment of the effectiveness of an agile process in large organisation, using as concrete experimental field the implementation of SAFe in the ABDT [87]. The assessment is based on GQM and systemic approach, which allowed us to effectively exploit the advantages of quantitative and qualitative tools and techniques. As a result, we have reviewed the current software development process in a comprehensive way considering different perspectives and assessing its

maturity with respect to SAFe and the overall agility. This enabled the identification of the key existing issues in the way ABDT implements SAFe.

Through this approach, we have found the undoubtedly positive trends since the introduction of SAFe. Testing and bug fixing processes have become more stable: fluctuations of differences between open/closed issues tend to zero. The overall quality of the products being released as evidenced by the user satisfaction ratings are presented in the play stores. Thus, we assume that the introduction of SAFe has a significant improvement of ABDT's development process.

We have also found open issues. First of all, the fundamental mechanics and attributes of SAFe are in place, but most employees have not yet realized its rationale. Even though people in ABDT and in Ak Bars Bank claim to support the key principles of agile development, the reality is that in the minds of many of them there is still a traditional process, just enriched with several "best practices" coming from agile.

Another issue, strictly connected to the previous, is the limited communication between roles in the organization, which could be caused by the lack of proper control and management in the implementation of SAFe. We have found this well exemplified in the relationships between a product owner and a business owner. Such limited communication may result in a very dangerous, significant, and constantly increasing mismatch of mutual expectations, even when sharing the paramount and collective business goals. The result of this was the initial dissatisfaction of the banks' management in the performance of ABDT.

The third issue is also strictly connected with the previous two points. The empowerment of teams and their motivation are insufficient. The definition and allocation of tasks are functional and hierarchical. The senior management of the bank do not delegate the decision-making to the teams in ABDT. As a result, the teams in ABDT evidence limited desires (and even opportunities) for exchanging knowledge across them. It leads to the problematic consequences for the overall development process.

Overall, the model of assessing an agile process in a large organisation that we have proposed in this action research has been found useful in its context. Moreover, we believe that it could be applied also in another context. In the future we aim at validating this approach further using more systematic research design models, as common in action research [88]. We also think that the issues that we have found are quite common while introducing agile approach in large organizations. Therefore, it can be used as criteria or reference situations for future planning and assessments, regardless of the specific planning or assessment model. Our next research will be concentrated on empirical studies of such findings.

## ACKNOWLEDGMENT

The authors would like to thank AK Bars Digital Technologies for their help in this work. They also thank Kruglova Evgenia (Institute of Humanities and Social Sciences) for proofreading.

## REFERENCES

- [1] VerionOne CollabNet. (2021). *15th annual State of Agile report*. Accessed: May 15, 2022. [Online]. Available: [https://info.digital.ai/rs/981-LQX-968/images/RE-SA-15th-Annual-State-Of-Agile-Report.pdf?\\_ga=2.142275953.2113799532.1652644027-1107505881.1651758513](https://info.digital.ai/rs/981-LQX-968/images/RE-SA-15th-Annual-State-Of-Agile-Report.pdf?_ga=2.142275953.2113799532.1652644027-1107505881.1651758513)
- [2] KPMG. (2019). *Agile Transformation. Survey on Agility*. Accessed: May 12, 2020. [Online]. Available: <https://assets.kpmg/content/dam/kpmg/be/pdf/2019/11/agile-transformation.pdf>
- [3] Gartner. (2019). *Results Summary: Agile in the Enterprise*. Accessed: Sep. 5, 2020. [Online]. Available: [https://circle.gartner.com/Portals/2/Resources/pdf/Agile%20in%20the%20Enterprise%202019%20-%20Results%20Summary%20\(updated\).pdf](https://circle.gartner.com/Portals/2/Resources/pdf/Agile%20in%20the%20Enterprise%202019%20-%20Results%20Summary%20(updated).pdf)
- [4] V. Ivanov, S. Masyagin, A. Rogers, G. Succi, A. Tormasov, J. Yi, and V. Zorin, "Comparison of agile, quasi-agile and traditional methodologies," in *Proc. Int. Conf. Softw. Eng. Defence Appl.* (Advances in Intelligent Systems and Computing). Cham, Switzerland: Springer, Mar. 2019, pp. 128–137.
- [5] AmbySoft. (2017). *Agile Governance Survey Results*. Accessed: Jun. 5, 2020. [Online]. Available: <http://www.ambysoft.com/surveys/agileGovernance2017.html>
- [6] AmbySoft. (2013). *How Agile Are You? Survey Results*. Accessed: Jun. 5, 2020. [Online]. Available: <http://www.ambysoft.com/surveys/howAgileAreYou2013.html>
- [7] V. Ivanov, M. Mazzara, W. Pedrycz, A. Sillitti, and G. Succi, "Assessing the process of an eastern European software SME using systemic analysis, GQM, and reliability growth models," in *Proc. 38th Int. Conf. Softw. Eng. Companion*, May 2016, pp. 251–259.
- [8] J. A. Farris, R. L. Groesbeck, E. M. V. Aken, and G. Letens, "Evaluating the relative performance of engineering design projects: A case study using data envelopment analysis," *IEEE Trans. Eng. Manag.*, vol. 53, no. 3, pp. 471–482, Aug. 2006.
- [9] P. Middleton and D. Joyce, "Lean software management: BBC worldwide case study," *IEEE Trans. Eng. Manag.*, vol. 59, no. 1, pp. 20–32, Feb. 2012.
- [10] J. Munch, O. Armbrust, M. Kowalczyk, and M. Soto, *Software Process Definition and Management* (The Fraunhofer IESE Series on Software and Systems Engineering), 1st ed. Berlin, Germany: Springer-Verlag, 2012.
- [11] P. S. Setiawan, M. I. Jambak, and M. I. Jambak, "The effectiveness of using software development methods analysis by the project timeline in an Indonesian media company," in *Proc. Int. Conf. Data Softw. Eng. (ICoDSE)*, Nov. 2017, pp. 1–6.
- [12] D. A. Tamburri, F. Palomba, and R. Kazman, "Success and failure in software engineering: A followup systematic literature review," *IEEE Trans. Eng. Manag.*, vol. 68, no. 2, pp. 599–611, Apr. 2021.
- [13] George Dinwiddie. (2009). *DIY Project/Process Evaluation Kit*. Accessed: Sep. 5, 2020. [Online]. Available: <http://blog.gdinwiddie.com/2009/08/18/diyprojectprocess-evaluation-kit/>
- [14] Joe Little. (2007). *The Nokia Test*. Accessed: Sep. 5, 2020. [Online]. Available: <http://agileconsortium.blogspot.com/2007/12/nokia-test.html>
- [15] Michael James. (2007). *An Example Checklist for Scrummasters*. Accessed: Sep. 5, 2020. [Online]. Available: [https://scrummasterchecklist.org/pdf/ScrumMaster\\_Checklist\\_12\\_unbranded.pdf](https://scrummasterchecklist.org/pdf/ScrumMaster_Checklist_12_unbranded.pdf)
- [16] S. Soundararajan, "A methodology for assessing agile software development approaches," 2011, *arXiv:1108.0427*.
- [17] A. Qumer and B. Henderson-Sellers, "A framework to support the evaluation, adoption and improvement of agile methods in practice," *J. Syst. Softw.*, vol. 81, no. 11, pp. 1899–1919, Nov. 2008.
- [18] C. Patel and M. Ramachandran, "Agile maturity model (AMM): A software process improvement framework for agile software development practices," *J. Softw. Eng.*, vol. 2, no. 1, pp. 3–28, Jan. 2009.
- [19] T. J. Gandomani, "Agility assessment model to measure agility degree of agile software companies," *Indian J. Sci. Technol.*, vol. 7, no. 7, pp. 955–959, Jul. 2014.
- [20] A. Sidky, "A structured approach to adopting agile practices: The agile adoption framework," Ph.D. thesis, Virginia Polytech. Inst. & State Univ., Blacksburg, VA, USA, 2007.
- [21] Scaled Agile. (2019). *Business Agility Self-Assessment*. Accessed: Mar. 15, 2020. [Online]. Available: <https://www.scaledagileframework.com/metrics/#EN1>
- [22] J. Humble and R. Russell. (Sep. 2009). *The Agile Maturity Model Applied to Building and Releasing Software*. Accessed: Sep. 14, 2020. [Online]. Available: [https://info.thoughtworks.com/rs/thoughtworks2/images/agile\\_maturity\\_model.pdf](https://info.thoughtworks.com/rs/thoughtworks2/images/agile_maturity_model.pdf)

- [23] M. Z. Nafchi and T. J. Gandomani, "Agile assessment methods and approaches," in *Advances in Systems Analysis, Software Engineering, and High Performance Computing*. Hershey, PA, USA: IGI Global, 2016, pp. 165–174.
- [24] L. Williams, K. Rubin, and M. Cohn, "Driving process improvement via comparative agility assessment," in *Proc. Agile Conf.*, Aug. 2010, pp. 3–10.
- [25] K. M. Spayd. (2014). *Coaching the Agile Enterprise*. Accessed: Sep. 5, 2020. [Online]. Available: <https://pdfs.semanticscholar.org/5efa/4d23e1beaaa4d181a0f3366dfcb920651822.pdf>
- [26] A. Qumer and B. Henderson-Sellers, "An evaluation of the degree of agility in six agile methods and its applicability for method engineering," *Inf. Softw. Technol.*, vol. 50, no. 4, pp. 280–295, 2008.
- [27] K. Bogner and U. Landrock, "Response Biases in Standardised Surveys," GESIS Survey Guidelines, Leibniz Inst. Social Sci., Mannheim, Germany, 2016, doi: [10.15465/gesis-sg\\_en\\_016](https://doi.org/10.15465/gesis-sg_en_016).
- [28] E. Wetzel, J. R. Böhnke, and A. Brown, *Response Biases*. Oxford, U.K.: Oxford Univ. Press, 2016, pp. 349–363.
- [29] C. Peter and H. Sue, *Action Research*. Boston, MA, USA: Springer, 2007, pp. 3–17.
- [30] R. Knaster and D. Leffingwell, *SAFe 5.0 Distilled; Achieving Business Agility With the Scaled Agile Framework*. Boston, MA, USA: Addison Wesley, 2020.
- [31] ScrumGuides.org. (2020). *The Scrum Guide*. Accessed: Sep. 13, 2020. [Online]. Available: <https://www.scrumguides.org/scrum-guide.html>
- [32] A. Oberoi. (2019). *SAFe vs LeSS vs DaD: Comparing the Three Frameworks to Scale Agile*. Accessed: Jun. 17, 2020. [Online]. Available: <https://insights.daffodilsw.com/blog/safe-vs-less-vs-dadcomparing-the-three-frameworks-to-scale-agile>
- [33] A. Krush. (2018). *5 Success Stories that Will Make you Believe in Scaled Agile*. Accessed: Sep. 9, 2020. [Online]. Available: <https://www.objectstyle.com/agile/scaled-agile-success-story-lessons>
- [34] M. Laanti and P. Kettunen, "SAFe adoptions in finland: A survey research," in *Agile Processes in Software Engineering and Extreme Programming Workshops*. Cham, Switzerland: Springer, 2019, pp. 81–87.
- [35] D. Leffingwell. (2020). *Category Archives for: Case Studies*. Accessed: Sep. 9, 2020. [Online]. Available: <https://www.scaledagileframework.com/blog/category/case-studies/>
- [36] K. Conboy and N. Carroll, "Implementing large-scale agile frameworks: Challenges and recommendations," *IEEE Softw.*, vol. 36, no. 2, pp. 44–50, Mar. 2019.
- [37] A. Putta, M. Paasivaara, and C. Lassenius, "Benefits and challenges of adopting the scaled agile framework (SAFe): Preliminary results from a multivocal literature review," in *Product-Focused Software Process Improvement*. Berlin, Germany: Springer, 2018, pp. 334–351.
- [38] S. Dexter. (2020). *Beware SAFe (The Scaled Agile Framework for Enterprise), an Unholy Incarnation of Darkness*. Accessed: Jun. 15, 2020. [Online]. Available: <https://medium.com/@seandexter1/beware-safe-the-scaled-agile-framework-for-enterprise-an-unholy-incarnation-of-darkness-bf6819f6943f>
- [39] K. Evans. (2019). *The Major Problems With SAFe*. Accessed: Jun. 17, 2020. [Online]. Available: <https://productcoalition.com/the-major-problems-with-safe-1e797f7e48f8>
- [40] H. B. Dale, *Total Quality Management*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1995.
- [41] K. C. GmbH. (2019). *Scaled Agile Frameworks: Your Complete Guide to Which, Why and How*. Accessed: Jun. 10, 2020. [Online]. Available: <https://kruschecompany.com/scaled-agile-agile-approachesextended-beyond-the-team-level/>
- [42] TAdviser. (2020). *AK Bars Digital Technologies Report*. Accessed: Sep. 5, 2020. [Online]. Available: <https://www.tadviser.ru/a/337799>
- [43] P. S. M. D. Santos and G. H. Travassos, "Action research use in software engineering: An initial survey," in *Proc. 3rd Int. Symp. Empirical Softw. Eng. Meas.*, Oct. 2009, pp. 414–417.
- [44] M. Staron, *Action Research in Software Engineering*. Berlin, Germany: Springer, 2020.
- [45] A. S. Olmos, A. J. Roffé, and S. Ginnobili, "Systemic analysis and functional explanation: Structure and limitations," in *History, Philosophy and Theory of the Life Sciences*. Berlin, Germany: Springer, 2020, pp. 209–229.
- [46] H. M. E. Abdelsalam and H. P. Bao, "A simulation-based optimization framework for product development cycle time reduction," *IEEE Trans. Eng. Manag.*, vol. 53, no. 1, pp. 69–85, Feb. 2006.
- [47] V. R. B. G. Caldiera and H. D. Rombach, *The Goal Question Metric Approach*, vol. 1. Hoboken, NJ, USA: Wiley, 1994.
- [48] V. R. Basili, M. Lindvall, M. Regardie, C. Seaman, J. Heidrich, J. Munch, D. Rombach, and A. Trendowicz, "Linking software development and business strategy through measurement," *Computer*, vol. 43, no. 4, pp. 57–65, Apr. 2010.
- [49] J. Munch, F. Fagerholm, P. Kettunen, M. Pagels, and J. Partanen, "Experiences and insights from applying GQM+strategies in a systems product development organisation," in *Proc. 39th Euromicro Conf. Softw. Eng. Adv. Appl.*, Sep. 2013, pp. 70–77.
- [50] L. Corral, A. Sillitti, and G. Succi, "Using a partially instantiated GQM to measure the quality of mobile applications," in *Proc. 25th Int. Conf. Softw. Eng. Knowl. Eng.*, Boston, MA, USA, 2013, pp. 520–524.
- [51] E. Gatarik, "The value of knowledge sharing in decision-making and organisational development: A model-theoretic systemic analysis of an intervention in an Austrian SME," *Econ. Research-Ekonomska Istraživanja*, vol. 32, no. 1, pp. 148–167, Jan. 2019.
- [52] W. Ulrich and M. Reynolds, "Critical systems heuristics," in *Systems Approaches to Managing Change: A Practical Guide*, M. Reynolds and S. Holwell, Eds. London, U.K.: Springer, 2010, pp. 243–292, doi: [10.1007/978-1-84882-809-4\\_6](https://doi.org/10.1007/978-1-84882-809-4_6).
- [53] I. D. Coman, A. Sillitti, and G. Succi, "A case-study on using an automated in-process software engineering measurement and analysis system in an industrial environment," in *Proc. IEEE 31st Int. Conf. Softw. Eng.*, May 2009, pp. 89–99.
- [54] I. Fronza, A. Sillitti, G. Succi, and J. Vlasenko, "Toward a non invasive control of applications—A biomedical approach to failure prediction," in *Proc. 13th Int. Conf. Enterprise Inf. Syst.*, vol. 2. Beijing, China: SciTePress, Jun. 2011, pp. 83–91.
- [55] P. M. Bentler, D. N. Jackson, and S. Messick, "Identification of content and style: A two-dimensional interpretation of acquiescence," *Psychol. Bull.*, vol. 76, no. 3, pp. 186–204, 1971.
- [56] M. N. Bradburn and S. Sudman, *The Current Status of Questionnaire Research*. Hoboken, NJ, USA: Wiley, 2004, ch. 2, pp. 27–40.
- [57] J. A. Williams, "Interviewer-respondent interaction: A study of bias in the information interview," *Sociometry*, vol. 27, no. 3, pp. 338–352, 1964.
- [58] A. Furnham, "Response bias, social desirability and dissimulation," *Personality Individual Differences*, vol. 7, no. 3, pp. 385–400, Jan. 1986.
- [59] A. J. Nederhof, "Methods of coping with social desirability bias: A review," *Eur. J. Social Psychol.*, vol. 15, no. 3, pp. 263–280, Jul. 1985.
- [60] V. Wilson, "Research methods: Triangulation," *Evidence Based Library Inf. Pract.*, vol. 9, pp. 74–75, Mar. 2014.
- [61] Atlassian. (2020). *Confluence Home Page*. Accessed: Oct. 30, 2020. [Online]. Available: <https://www.atlassian.com/software/confluence>
- [62] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *Ann. Math. Statist.*, vol. 18, no. 1, pp. 50–60, 1947.
- [63] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis," *J. Amer. Stat. Assoc.*, vol. 47, no. 260, pp. 583–621, 1952.
- [64] A. N. Kolmogorov, "Sulla determinazione empirica di una legge di distribuzione," *Giornale dell'Istituto Italiano Degli Attuari*, vol. 4, pp. 83–91, Jan. 1933.
- [65] W. Tengeler. (2014). *Cone of Uncertainty for Agile Projects*. Accessed: Oct. 5, 2020. [Online]. Available: <http://www.themotionstudio.nl/en/cone-of-uncertainty-for-agile-projects/>
- [66] P. Aroonvatanaporn, C. Sinthop, and B. Boehm, "Reducing estimation uncertainty with continuous assessment: Tracking the cone of uncertainty," in *Proc. IEEE/ACM Int. Conf. Automated Softw. Eng. (ASE)*, 2010, pp. 337–340.
- [67] S. Luyten. (2014). *The Cone of Uncertainty and How to Avoid it Turning Into a Wormhole*. Accessed: Oct. 5, 2020. [Online]. Available: <https://medium.com/@stefanluyten/the-cone-of-uncertainty-82d21e99fcc2>
- [68] Department of Information Technologies. (2019). *Agile Maturity Matrix*. Accessed: Mar. 14, 2020. [Online]. Available: <https://doit.maryland.gov/SDLC/Pages/templates-phases.aspx>
- [69] M. Q. Patton *Qualitative Research & Evaluation Methods*. Thousand Oaks, CA, USA: SAGE, 2002, pp. 209–259.
- [70] G. Marzi, F. Ciampi, D. Dallì, and M. Dabic, "New product development during the last ten years: The ongoing debate and future avenues," *IEEE Trans. Eng. Manag.*, vol. 68, no. 1, pp. 330–344, Feb. 2021.
- [71] H. D. Frederiksen and L. Mathiassen, "A contextual approach to improving software metrics practices," *IEEE Trans. Eng. Manag.*, vol. 55, no. 4, pp. 602–616, Nov. 2008.



[72] J. P. Kotter, *Leading Change*. Boston, MA, USA: Harvard Business Press, 1996.

[73] L. Curry and M. Nunez-Smith, *Mixed Methods in Health Sciences Research: A Practical Primer*. Thousand Oaks, CA, USA: SAGE, 2015.

[74] T. D. Campbell and C. J. Stanley, *Experimental and Quasi-Experimental Designs for Research*. Chicago, IL, USA: Rand McNally College, 1963.

[75] S. Ergasheva, M. Gudkov, A. Kruglov, G. Succi, X. Vasques, and A. Zagidullina, "Obtaining data from the third-party systems for software development process analysis," in *Proc. 3rd Int. Conf. Inf. Manag. Manag. Sci.*, Aug. 2020, pp. 82–86.

[76] D. Atonge, V. Ivanov, A. Kruglov, I. Khomyakov, A. Sadovykh, D. Strugar, G. Succi, X. Z. Vasquez, and E. Zouev, "The development of data collectors in open-source system for energy efficiency assessment," in *Proc. IFIP Int. Conf. Open Source Syst.* (IFIP Advances in Information and Communication Technology). Cham, Switzerland: Springer, 2020, pp. 14–24.

[77] K. Dikert, M. Paasivaara, and C. Lassenius, "Challenges and success factors for large-scale agile transformations: A systematic literature review," *J. Syst. Softw.*, vol. 119, pp. 87–108, Sep. 2016.

[78] N. B. Moe and T. Dingsøy, "Scrum and team effectiveness: Theory and practice," in *Agile Processes in Software Engineering and Extreme Programming* (Lecture Notes in Business Information Processing). Berlin, Germany: Springer, 2008, pp. 11–20.

[79] C. Verwijs, J. Schartau, and B. Overeem, *Zombie Scrum Survival Guide*, 1st ed. Reading, MA, USA: Addison-Wesley, 2020.

[80] M. Staron, "Validity evaluation," in *Action Research in Software Engineering*. Berlin, Germany: Springer, Nov. 2019, pp. 169–189.

[81] K. E. Watkins, *Validity in Action Research*. Washington, DC, USA: ERIC Clearinghouse, 1991.

[82] A. Feldman, "Validity and quality in action research," *Educ. Action Res.*, vol. 15, no. 1, pp. 21–32, Mar. 2007.

[83] C. Turnock and V. Gibson, "Validity in action research: A discussion on theoretical and practice issues encountered whilst using observation to collect data," *J. Adv. Nursing*, vol. 36, no. 3, pp. 471–477, Nov. 2001.

[84] A. Wiek, B. Ness, P. Schweizer-Ries, F. S. Brand, and F. Farioli, "From complex systems analysis to transformational change: A comparative appraisal of sustainability science projects," *Sustainability Sci.*, vol. 7, no. S1, pp. 5–24, Feb. 2012.

[85] M. Sahota, *An Agile Adoption and Transformation Survival Guide*, 2012.

[86] E. Amr, *Agile Adoption Patterns: A Roadmap to Organizational Success*, 1st ed. Reading, MA, USA: Addison-Wesley, 2008.

[87] W. L. French and C. Bell, *Organization Development: Behavioral Science Interventions for Organization Improvement*. Upper Saddle River, NJ, USA: Prentice-Hall, 1999.

[88] A. Styhre and M. Sundgren, "Action research as experimentation," *Systemic Pract. Action Res.*, vol. 18, no. 1, pp. 53–65, Feb. 2005.



**ARTEM KRUGLOV** received the graduate degree from Ural Federal University, in 2013, and the Ph.D. degree, in 2017. He is an Assistant Professor with the Faculty of Computer Science and Software Engineering, Innopolis University. His research interests include aspects of software development processes, agile methodologies, product and project management, and empirical methods.



**WITOLD PEDRYCZ** (Life Fellow, IEEE) is a Professor and the Canada Research Chair (CRC) of computational intelligence with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada. He is also with the Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland. He is a Foreign Member of the Polish Academy of Sciences and a fellow of the Royal Society of Canada. He was a recipient of several awards, including Norbert Wiener Award from the IEEE Systems, Man, and Cybernetics Society, the IEEE Canada Computer Engineering Medal, the Cajastur Prize for Soft Computing from the European Centre for Soft Computing, the Killam Prize, the Fuzzy Pioneer Award from the IEEE Computational Intelligence Society, and the 2019 Meritorious Service Award from the IEEE Systems Man and Cybernetics Society. He is an Editor-in-Chief of *Information Sciences* and *WIRES Data Mining and Knowledge Discovery* (Wiley) and a Co-Editor-in-Chief of *Granular Computing* (Springer) and *Journal of Data Information and Management* (Springer).



**ALIYA SAMATOVA** received the graduate degree from the Tatar-American Regional Institute, in 2006. She worked at Tattfondbank for ten years and worked her way up from a Recruiter to the Head of the Department. Since 2015, she has been the Vice-Rector for Development and Human Resources Policy, Innopolis University.



the Deputy Chairperson of the Chairman-Management Board with AK Bars Bank, where he is in charge of information and technological support.

**MERAB GOGICHATY** received the undergraduate degree from North Caucasus State Technical University and the graduate degree from Samara State University. From 2011 to 2016, he was at the Povolzhsky Bank of Sberbank, Russia. Since 2017, he has been the Managing Director of Sberbank-Service LLC. From April to November 2018, he was the Director of the Operations and Technology Directorate, Ak Bars Bank OJSC. Since November 2018, he has been



natural language processing, information extraction, text mining, software reliability, and software metrics.

**VLADIMIR IVANOV** received the Ph.D. degree in physics and mathematics from Kazan State University, in 2009. He was a Researcher/an Assistant Professor at Kazan Federal University, from 2005 to 2013, the Head of the Big Data and Text Analytics Laboratory, Kazan Federal University, from 2013 to 2015, and a Researcher at Innopolis University, from 2015 to 2018. His research interests include data analysis, machine learning, software engineering, computational linguistics,



including open source development, agile methodologies, experimental software engineering, software engineering over the internet, and software product lines and software reuse.

**GIANCARLO SUCCI** (Member, IEEE) is a Professor with the University of Bologna, Italy. He was a Full Professor at Innopolis University, Russia, a Professor with tenure at the Free University of Bolzano-Bozen, Italy, a Professor with tenure at the University of Alberta, Edmonton, AB, Canada, an Associate Professor at the University of Calgary, AB, Canada, and an Assistant Professor at the University of Trento, Italy. His research interests include multiple areas of software engineering,



• • •  
**RAPHAEL VALEEV** received the graduate degree from Kazan State Technical University named after A.N. Tupolev, in 2009. Since December 2015, he has been the Head of the Division of the Processing Center, IT Department, PJSC "AK BARS." Since October 2016, he has been the General Director of AK BARS Digital Technologies.