Questionnaire Development for a Scientifically Founded Agile Assessment Model

Doruk Tuncel[1], Christian Körner[1] and Reinhold Plösch[2]

[1]*Siemens AG, Otto-Hahn-Ring 6, Munich, 81739, Germany*

[2]*Department of Business Informatics Software Engineering Johannes Kepler University Linz, Altenberger Str. 69, Linz, 4040, Austria*

Abstract

Agile software development methodologies have been a focal point of industry and academia over the past two decades. Organizations are interested in extracting the value of agile methodologies and increasing their business success, however, it is evident that merely applying agile practices does not necessarily make organizations agile. We are building a context agnostic agile assessment model based on agile values and principles to enable contextually appropriate self-assessments, and support organizations' agile transformation endeavors. In this paper, we share the results of our expert interviews with the focus on two principle pillars: *Human Centricity* and *Technical Excellence*. Results of this study show that the proposed assessment questionnaire appropriately addresses highly relevant aspects of agility. It is found practically useful by the experts, yet the coverage of human centricity targeted questions can be improved.

Keywords
agile, maturity assessment, process improvement

1. Introduction

As we approached the end of second decade mark of the publication of the Agile Manifesto [1], agile software development methodologies have been investigated by academics, and advocated by many of the industry leading organizations over the past years [2]. Whether named agile transition or transformation, programs towards becoming an organization that can cope easier with the changing market demands, customer needs and technology stacks have been frequently promoted. Some of these programs are also called digital transformation programs which, in essence, highlight the relevance and importance of the notion of change and transformation for surviving in the highly competitive volatile, uncertain, complex and ambiguous (VUCA) business landscape. While these programs motivate organizational units to adopt certain agile frameworks, or to adhere to certain best practices of agility, it is evident that focusing merely on practices misses the essence behind agile adoption journeys.

Therefore we focus on the values and principles of the Agile Manifesto in building an agile assessment model (AAM). With this agile assessment model, we aim to support organizations

in their transformation journeys and help them identify how agility is lived within their units, irrespective of any agile framework. This model, in our opinion, has the potential to support organizations in conducting contextually appropriate (self)-assessments, judging their own capabilities with respect to their own targets, and identifying areas of improvement that are not pre-imposed as global optima by any of the agile frameworks.

During our previous research [3], we have identified a comprehensive set of agile assessment models, and noticed that the models with scientific foundations do not address some of the aspects that are very important and relevant for real life application scenarios such as large scale organizations, required industry standards for safety critical systems, application domains where hardware development is an inseparable element of the value delivery. While many contributions provide valuable insights towards which agile practices are useful for an organization to adopt, the underlying meta-model structures of the proposed agile assessment models tend to overlook foundational aspects of agility, as most of them do not explicitly reflect on the values and principles of the Agile Manifesto.

Although the industrial relevance of the problem is clear, there have been only a few attempts for practically validating the existing models. Given the additional limitations of the existing models we discussed in [3], we decided to develop a model that is scientifically founded, yet is capable of providing practical use to the industry in varying contexts [4]. It is also important to mention that, agile at large scale was listed among the top 10 burning research questions [5]. For taxonomic clarity throughout the paper, we keep an alignment with [6]. To continue our endeavour in further validating our model elements, we focus on the following research questions (RQ) in this paper:

- RQ1: Do the questions belonging to Human Centricity and Technical Excellence principles cover the concepts related to the respective principles they represent?
- RQ2: Is the proposed mechanism for addressing the large scale product development applicable from a questionnaire perspective?

The main contribution of this research is the result of the expert interviews we conducted across different business units of Siemens towards the validation of the questions for Human Centricity and Technical Excellence principles. We select these two principles because if we can successfully design a robust question development mechanism to serve the needs of these two principles with two distinct stances, this will guide us for the remaining three principles, whose stances fall in between human centricity and technical excellence. The validation focuses on (1) the understandability and feasability of the questions for these principles and (2) the completeness of the questions with respect to the principles. This validation is necessary and valuable for our future research, as it allows us to adjust our self-defined questionnaire development guidelines, and to incorporate the feedback in the development of the questions for the remaining principles.

2. A Novel AAM

The structural components of the proposed agile assessment model have been discussed in our previous paper [4]. In order to increase the understandability of this paper's contribution, we

present the underlying vocabulary in Section 2.1. Section 2.1.3 describes the questionnaire that is built on the assessment model. Finally, Section 2.2 elaborates on the novel aspects of both the assessment model and the method.

2.1. Model Structure

Our model consists of two main structural components: *Principles* and *Clusters*. These structural components correspond to the static aspect of the overall assessment system. Principles are the abstract components that are constructing the areas under which agility should be evaluated. After an extensive literature review, the principles are derived particularly from the model proposed by Sidky et al. [7]. Clusters, on the other hand, represent one lower level of abstraction, containing groups of certain practices of agile software development methodology. We are reusing the principles and clusters as we published and validated in XP'21 [4].

2.1.1. Principles

The proposed model consists of five principles: (1) Embrace change to deliver customer value (2) Plan and deliver software frequently (3) Human centricity (4) Technical excellence and (5) Customer collaboration with no particular order of importance. Principles provide the frame of the assessment model, and are populated with clusters.

2.1.2. Clusters

The proposed model consists of seventeen clusters. As the numbers suggests, clusters are not enforced to be equally distributed across the principles. The clusters came into life as a result of semantic grouping the identified agile software development practices in our literature study [3]. Due to space limitations, the cluster names are excluded in this section, however, particularly relevant clusters for this research, along with their codings, can be found under Section 5.

2.1.3. Questionnaire

The questionnaire is the backbone of the model, as it defines the core of the evaluation. For each cluster there exist multiple questions, and sum of all clusters' questions form the entire questionnaire. In order to systematically develop questions and answer option statements for each capability level-question pair, we setup a meta-structure as a guideline. Questions are therefore either related to specific aspects of artifacts. We also defined, how to bring in indicators, metrics or KPIs at which capability level of a question. We additionally decided to differentiate from SPICE [8] and CMMI [9] to work with specific practices on all capability levels, and therefore to avoid generic practices at higher capability levels. The developed questionnaire contains the questions, capability levels along with their underlying definitions as well as the answer option statements dedicated to each capability level-question pair. The overall capability level definitions enables systematically defining the descriptions for each capability level per question. This additional layer of granularity is an important differentiation point of our design from the existing models. This is because while other models are evaluating whether certain agile practices are performed in the organization, we are going beyond and evaluating the degree

of performing those practices, and quality of the produced artifacts. As our model consists a total of seventeen clusters, any evaluation may easily expand to an impractical amount of questions for each cluster, making this a minimum spanning problem. Therefore, in order to increase the practical applicability of the proposed assessment model our aim is to reach the maximum coverage of agility by means of the minimum set of questions that allow us to maintain practical applicability without sacrificing quality. Due to space limitations, the questions of both the Technical Excellence and Human Centricity Principles are provided at https://bit.ly/3NG8glr.

Figure 1 visualizes a question from a cluster under Technical Excellence. In total, the validation covers eight different clusters consisting total of thirty two questions. For each question, the provided answer option statements per capability level brings in another level of complexity to the overall design with total of hundred sixty statements.

Remaining elements of the overall assessment method (e.g., assessment guidelines, aggregation and evaluation mechanisms) are mentioned under Section 8.2.

Question	L1: Discovering	L2: Experimenting	L3: Measuring	L4: Optimizing	L5: Mastering
Does the source code adhere to coding best practices?	Developers get to know about coding best practices. They gather knowledge about which coding best practices to use and to understand why they are important.	Coding best practices are not explicitly defined but source code partly adheres to coding best practices. Examples: coding style, information hiding, complexity, memory management.	Required coding practices are explicitly defined (documentation or tool configuration). Mandatory practices are part of the DoD. Indicators for source code quality are defined (e.g. Technical Debt, Code Coverage, Complexity, etc.) and systematically measured.	Adherence to coding best practices is checked and required. There exist target values for the defined indicators. Impact of coding best practices onto customer value (e.g., failures) and business value (e.g., #post release defects) is used for optimization.	Adherence to coding best practices is not an issue anymore. Reasons could be: Source code is generated from higher level DSLs, team does not violate the principles anymore, as they fully know how to program with good quality (there are practically no post release defects that can be tracked).

Figure 1: Sample Question

2.2. Novelty

Our previous study [3] revealed that, fundamental aspects are not present in the existing models. Therefore we introduce a new model in [4] instead of substantially reworking existing models which are fundamentally different by design. The novelty of our proposal resides under both the static and dynamic components. Features such as having theoretical and practical validation, providing guidance for improvement, showing an explicit awareness of the agile values and principles of the Agile Manifesto or even paying attention to human centric aspects of agility are not very common in the existing assessment models.

2.2.1. Artifact Orientation

Artifact orientation is one of the core guiding aspects of our model development approach. It is important because constructing an assessment model on artifacts contributes to the objectivity of the assessment results. In our model, the artifact orientation is reflected under the questionnaire.

Some artifact examples of the Technical Excellence principle's Testing cluster are "Test Strategy", "Test Plan" and "Unit Test". If we take test strategy, while mere existence of a test strategy itself may hold value to an extent, the provided leveling mechanism allows identifying a range of states: whether the importance of having a test strategy is acknowledged, whether

a test strategy is in-use up to the point where an existing strategy is being reduced down to its essentials and new, contextually supportive strategies are introduced. Evaluating the state of an artifacts not only supports the endeavour of conducting objective evaluations, but also provides a concrete path of improvement to the assessed entity.

2.2.2. Enabling Self-Assessments

Enabling self-assessments is another guiding aspect of our model development endeavor. This aspect is important mainly for two reasons. The first one is about the practical applicability of the assessment model. The questions are designed and validated during the design iterations, to ensure that they are understandable. In industry, assessments hold a great potential in providing units with a realistic image of their current state and actionable insights for improvement opportunities. With our model, we encourage units to conduct their own assessments and identify improvement areas with respect to their own targets. To realize that, we designed the questions of the questionnaire without any framework-specific terminology and provided guiding examples for each question. Second, the leveling mechanism has its own underlying descriptions to ease the assessment process during self-assessments. The aforementioned artifact orientation for objectivity is also a contributing factor to this aspect.

2.2.3. Modularity

The modularity corresponds to the aspect of the assessment model that enables contextually appropriate assessments. To achieve modularity, we separate the concerns of different context defining elements in our design. For example, questions under each cluster are categorically labeled as being relevant only for large scale organizations. This gives our model the flexibility to discard contextually inappropriate questions, and to adapt the questionnaire size. Similar mechanism applies to other context defining elements such as having hardware development concerns, or having geographically distributed units. Having such modular structure enables the questionnaire to be tailored specifically to the changing organizational contexts.

3. Methodology

In our journey towards establishing a practically applicable agile assessment model that is both academically and practically validated, we designed our research method as a design science research methodology as proposed by Hevner et al. [10]. In order to ensure the soundness of the model and questionnaire development activities, we structured our own research in the form of an agile process, where we conduct interview studies with the experts in our organization and integrate the feedback into the method development process. While the design science research methodology provides the overall frame of research, in order to integrate the learning outcomes of the academic and industrial validation steps, we iteratively realize design science research in alignment with action design research defined by Sein et al. [11]. By doing so, feedback is systematically collected from the potential consumers of the design artifact.

4. Related Work

Sidky et al. [7] propose one of the early models for guiding organizations in their agile adoption journey. Authors define an agile adoption framework with two core elements: *agile measurement index* and *four stage process*. These two elements support organizations in adopting agile practices. The agile measurement index consists of agile *levels*, *principles*, *practices* and *indicators*. Five agile principles of the measurement index are derived from the Agile Manifesto and the practices populate the evaluation matrix. The four stage process utilizes the described agile measurement index. The model is explicitly highlighting the importance of *tailorability* of the five levels, by describing the challenges behind reaching a consensus on the assignment of practices to the levels.

Qumer and Henderson-Sellers [12] provide an agile solution framework for software. The model is built upon a conceptual agile model that is accompanied by an analytical tool consisting of four dimensions and an agile toolkit. The core of the method comprises the following five aspects: *Agility*, *People*, *Process*, *Product*, *Tools*, and an additional *Abstraction* aspect to reflect the agile software development methodology. While the agile toolkit consists of seven main components, the provided analytical tool focuses on the following four dimensions: *Method scope*, *Agility characterization*, *Agile value characterization* and *Software process characterization*. In order to support these two framework components, the authors propose the Agile Adoption and Improvement Model (AAIM) that is built on three agile blocks: *Prompt*, *Crux* and *Apex* and six agile levels. The study highlights the importance of supporting situation specific scenarios in software engineering.

Fontana et al. [13] suggest an agile maturity framework based on the complex adaptive systems theory. Authors emphasize ambidexterity as a fundamental attribute towards maturity and the described framework focuses on outcomes rather than agile practices. The fundamental positioning of people in the software development processes is made explicit, while the trade-off between exploitation and exploration is mapped to the trade-off between the specific outcomes and adopting new practices. As a result of their cross-case analysis, the following six outcomes were provided: *Practices*, *Team*, *Deliveries*, *Requirements*, *Product* and *Customer*. The study concludes by underlining the importance of allowing context-specific practices in the maturing process while considering the values of the Agile Manifesto.

5. Evaluation Results

This section discusses the results of the expert interview study conducted with 7 experts. The demographics of the experts can be found in Figure 2.

Each interview was designed as 45-60 minutes, per principle. When the area of expertise fell more towards ethics, communication, collaboration and not necessarily to software engineering on the implementation level, only the human centricity principle was evaluated. The data collection was both in the form of likert scale questions, as well as open text for further descriptions. Likert scale questions along with their codings as described below, represent the quantitative results. These are provided in the form of heatmaps in the Figures [3 - 6]. Open text questions provide the qualitative results in accordance with [14] and are either part of

	Professional Experience (Years)	Experience in Agile (Years)	Role	Division	Location
E1	>20	>10	Scrum Master, Agile Coach	Mobility	UK
E2	>20	>10	Sr. Consultant	Technology	GER
E3	>10	>10	Scrum Master, Agile Coach&Trainer	Consultancy	TR
E4	>20	>10	Sr. Consultant, Agile Expert	Consultancy	USA
E5	>20	>10	Sr. Software Engineer, Test Expert	Technology	GER
E6	>10	>10	Sr. Sofware Engineer, Agile Expert	Technology	GER
E7	>10	>10	Sr. Sofware Engineer, DevOps Expert	Technology	GER

Figure 2: Demographics

Section 5 or form the basis of Section 6 and Section 8.

In order to deliver the results in a coherent way, the following subsections will share the results regarding the questionnaire, human centricity principle and technical excellence principle under Section 5.1, Section 5.2 and Section 5.3 respectively.

The coding mechanism of results in the following figures will be as follows: *Strongly Agree:* 5, *Agree:* 4, *Neither/Nor:* 3, *Disagree:* 2, *Strongly Disagree:* 1, while E1 to E7 refer to the interviewed experts. For the cluster mappings of the Human Centricity Principle, we use Psychological Safety (HC1), Effective Communication (HC2), Unit Empowerment and Autonomy (HC3), Unit Collaboration (HC4) and Personal Growth (HC5). Same mechanism applies for the Technical Excellence Principle as, Design and Coding Practices (TE1), Testing (TE2) and Data Driven DevOps (TE3).

5.1. Generic Results

The overall results show that the relevance of the artifacts used in the questionnaire, and the coverage of the questions with respect to the clusters they are asked under, do differ. As it can be observed in Figure 3, all but one of the experts that are involved in software development on the level of implementation agree, if not strongly, that the provided artifacts are relevant for representing their associated clusters. On the other hand, as it can be seen in Figure 4, the coverage of the questions with respect to the cluster they are associated shows that the questions of Psychological Safety (HC1) cluster lack the expected coverage.

Promising results with respect to coverage show that the rules we used for developing the questions as described under Section 2.1.3 were reliable reference points, as judging the coverage of a cluster requires full understanding of the questions and their formulations.

Another encouraging result is regarding the evolutionary leveling mechanism, where all 7 experts agreed that the leveling mechanism is suitable for differentiating between the capabilities of organizations with respect to agile software development methodology.

For the interpretation of Figures [5-6] the following question mappings will be considered: **Q1:** I think the set of clusters sufficiently reflects the underlying principle in an agile organizational

	E1	E2	E3	E4	E5	E6	E7
HC1	3	3	5	1	4	5	4
HC2	3	5	5	1	4	4	5
HC3	4	4	5	4	4	4	5
HC4	4	4	5	2	4	4	5
HC5	4	3	5	1	4	5	5
TE1	-	-	5	4	5	4	5
TE2	-	-	4	4	4	5	4
TE3	-	-	5	2	4	4	5

Expert = E, Human Centricity = HC, Technical Excellence = TE, Strongly Agree= 5, Agree= 4, Neither/Nor = 3, Disagree = 2, Strongly Disagree = 1, Not asked = "-"

Figure 3: Relevance of the artifacts used in the questionnaire, per cluster

unit. **Q2:** I think this mechanism of adding a question for larger scales is suitable for this principle. **Q3:** I think an evaluation based on this questionnaire would help my unit to identify improvement areas. **Q4:** I can imagine myself using this model for a self-assessment of my unit. **Q5:** This model would be useful for capturing the state of agility of my unit.

Results of our interviews show that not only the structural components, but also the provided questions for both principles have been separately evaluated and agreed to be providing sufficient coverage for a practical evaluation as well. It is important to mention again that the evaluation is not on the questions only, but also on the statements we provide for each capability level(L1-L5) as depicted in Figure 1. Overall, only one expert(E4) disagreed to use Human Centricity Questionnaire. 5 out of 7 experts agreed that they would be willing to use one of the questionnaires in real life while 4 out of 7 stated the same for both of the questionnaires.

	E1	E2	E3	E4	E5	E6	E7
HC1	2	2	4	2	2	3	3
HC2	2	3	4	1	4	4	5
HC3	4	2	5	2	3	3	4
HC4	4	4	5	2	4	2	3
HC5	4	3	5	2	4	3	4
TE1	-	-	3	4	4	4	4
TE2	-	-	4	4	4	2	3
TE3	-	-	5	2	4	2	3

Expert = E, Human Centricity = HC, Technical Excellence = TE, Strongly Agree= 5, Agree= 4, Neither/Nor = 3, Disagree = 2, Strongly Disagree = 1, Not asked = "-"

Figure 4: Coverage of the questions used in the questionnaire, per cluster

5.2. Human Centricity Specific Results

As it can be observed from Figure 5, questionnaire belonging to this principle, excluding one expert, received particularly positive feedback with respect to its industrial applicability potential. In addition to that, Figure 5 highlights the coverage of the Human Centricity Principle by the proposed five clusters was very well received by the experts. It can also be observed in the same figure, on the level of a principle, the mechanism for addressing the large scale development aspects with respect to human centricity found to be suitable by most of the experts. On the level of clusters, addressing the scaled aspects were evaluated as sufficient by three experts for Psychological Safety and Effective Communication. For Unit Autonomy and Empowerment, and for Personal Growth, the decision of not including an additional question for addressing large scale was also found appropriate. Further, combining autonomy and empowerment concepts under one cluster instead of separately addressing them received positive feedback. As a side note, although keywords such as purpose, meaning, engagement or valuing someones work, were articulated during the discussions on the Unit Autonomy and Empowerment, the current placement of these keywords under the Psychological Safety was found appropriate.

	E1	E2	E3	E4	E5	E6	E7
Q1 (Principle Relevance)	4	4	5	5	5	4	5
Q2 (Scale)	2	4	5	4	4	3	4
Q3 (Improvement)	4	4	5	2	5	4	5
Q4 (Self-Assessment)	3	3	5	2	5	4	4
Q5 (Suitability)	4	3	5	2	5	4	4

Figure 5: Representativeness of human centricity clusters and addressing human centricity at scale

Regarding Human Centricity Principle, "failure" was highlighted as an important missing aspect of Psychological Safety by one of the experts. Moreover, it was suggested for large scale product development organizations that if there exists a misalignment between teams with respect to the artifacts and methods associated with psychological safety, the questionnaire should be able to address those and make them visible to the members of the organization. Another remark with respect to Psychological Safety was the importance of capturing abstract notions such as trust, integrity and transparency. In terms of Effective Communication, the domain of communication engineering along with its features (e.g., structure, construct, time, type) was commented as a useful component that can be additionally considered. Importance of in person gatherings was also discussed in relation to COVID-19 related developments and their consequences. It was also reported that while tool support is very essential, the utiliziation of tools may differ among different scales of an organization. A critical view on Unit Autonomy and Empowerment was highlighting that the existence of organizational structures supporting autonomy do not necessarily mean that the organization can operate autonomously, as it relies on people and their intrinsic motivation to own certain tasks and work in an autonomous way. Additionally, even the individuals are intrinsically motivated towards self-organization and working autonomously, one of the experts commented that the resource limitations may hinder the process, even if the will is there. For Unit Collaboration, one of the expert's recommendation

was framing the collaboration aspect as an activity of mutual artifact creation and therefore suggested us to inspect joint efforts going into the creation of any software artifact such as revision repositories. When it comes to the Personal Growth, following two aspects were commented as important: first, how growth is perceived in the organization and second, the relevance of decoupling the views of the provider and receiver of any learning component.

5.3. Technical Excellence Specific Results

As it can be observed from Figure 6, questionnaire belonging to this principle received particularly positive feedback with respect to its industrial applicability potential.

As a positive remark, except one expert, questions associated with each cluster of Technical Excellence Principle are found to be easy to understand. Moreover, regarding addressing the large scale product development on a principle level, **Q2** shows that the suitability of the proposed mechanism is agreed by all experts. On the cluster level, the questions tailored for the scaled aspects of Design and Coding Practices was also found to be sufficient in addressing the large scale development by all but one expert. The comment from that one expert was the necessity of including "interface and cross-component connectivity" under this cluster. When it comes to Testing, the scaled aspects received positive response from three out of five experts. The remaining two highlighted the importance of an "overall test architecture" and suggested "test strategy" to be explicitly considered also at large scale. Data Driven DevOps as well received positive response from three of the experts with respect to completeness, while the other two highlighted the importance of "purposeful measuring" and handling metrics and dashboards separately.

	E3	E4	E5	E6	E7
Q1 (Principle Relevance)	4	5	4	4	4
Q2 (Scale)	5	4	4	4	5
Q3 (Improvement)	5	4	4	4	5
Q4 (Self-Assessment)	5	4	4	4	5
Q5 (Suitability)	5	4	4	4	5

Figure 6: Representativeness of technical excellence clusters and addressing technical excellence at scale

Clusters of the Technical Excellence Principle received additional comments regarding their coherence and completeness. Design and Coding Practices was primarily evaluated with high relevance and sufficiently good coverage with its underlying artifacts and methods. Consequently, it did not received any negative feedback, however, one of the experts shared its own preference towards considering software architecture related aspects under this cluster more explicitly. Following comment was emphasizing the importance of reflecting the architectural decisions both on team and large scale product development level. Testing received additional input regarding the level of granularity of its questions. Two experts stated that the joint representation of two questions under this cluster may reduce the assessment time without

compromising the quality of results. In terms of Data Driven DevOps, an improvement sugges-
tion was towards the clarification of the "incident documentation" artifact. It was denoted by
one of the experts that incident documentation may not be a very informative artifact, if the
area of incident is not specified within the questions.

6. Discussion

In order to avoid the potential risk of leading to human fatigue, the scope of this research
focuses on the questionnaires of two principles. In order to deliver the results in a coherent
way, the subsections will address the following: our findings that are not specific to either of
the principles, findings that are specific to the human centricity principle and findings that
are specific to the technical excellence principle, under Section 6.1, Section 6.2 and Section 6.3
respectively.

6.1. Generic Findings

As the main finding, it is important to highlight that both questionnaires are found to be
applicable and showed potential for being utilized in the industry. As mentioned in Section
5, questions Q3, Q4 and Q5 refer to experts' opinion regarding their usage of this model for
their own purposes and the responses to these questions are very promising and encouraging.
Artifact orientation in the questionnaire was very well received, which is an important reference
point and learning outcome for future researchers to consider.

On the other hand, understandability of the questionnaires were very high. Surely, there are
suggestions and minor improvement possibilities for each principle. The results of the expert
interviews show that the questions of the Technical Excellence Principle have a better coverage
than the questions of the Human Centricity Principle. Here, the Psychological Safety stands
out as a cluster with a relatively lower coverage. Our interpretation is that, as a very intangible
and subjective concept, psychological safety becomes hard to evaluate on an artifactual basis.

Another important remark was on setting targets appropriately. The ability of establishing
targets is very fundamental to any improvement activity, however, setting targets carelessly
brings the risk of jeopardizing the optimization and continuous improvement activities. This is
usually perceived as a hidden risk, as the individuals may believe that the organization is on its
way to substantial optimization, without noticing that the optimization is directed towards a
wrong, or sub-optimal target.

Besides the aforementioned generic findings, below we also address our research questions
based on our findings.

**RQ1: Do the questions belonging to Human Centricity and Technical Excellence
principles cover the concepts related to the respective principles they represent?** It is
clearly indicated that the existing clusters sufficiently cover the principles they are positioned
under. This is a very important finding to ensure the reliable foundation of the structural
components of the overall assessment system, as the model components provide the foundation
for it.

**RQ2: Is the proposed mechanism for addressing the large scale product development
applicable from a questionnaire perspective?** As agile at large scale is one of the contem-

porary areas of research, and as organizational scale is one of the context defining components of our modular approach, we provided a mechanism for addressing the aspects of clusters that are only relevant for large scale organizations. Our mechanism for addressing the large scale aspects of agility has received very positive feedback and was found useful by the experts.

6.2. Human Centricity Specific Findings

Regarding human centricity, one of the essential comments was that the leadership, whether low, medium or high level, plays an important role in the organizations' ability to act autonomously. This was suggested as an aspect that would, once made explicit, increase the coverage of the Unit Autonomy and Empowerment. In the model structure, notion of leadership is implicitly considered under Unit Autonomy and Empowerment, yet, is expected by the interviewees to be more explicitly represented. Another aspect that was frequently discussed during the interviews was the contrast between enabling the members of the organization, and actively supporting them.

Under Effective Communication and Unit Collaboration, it was also discussed that the technology or best practice sharing as well as the community of practice like initiatives increase the cohesion among the organizational structures, and contributes to the organizational synergy. Regarding the unit collaboration, a systematic enabler for an organization was mentioned as establishing rewarding mechanisms and objectives that are supportive of the collaboration across units, and not inhibiting their information experience sharing. To us, it is also clear that, once organizations establish the rewarding mechanisms and objectives in a way that they do not conflict with other units, there exist higher chances for collaboration to foster organically.

It is evident that the new ways of work, along with the transformations COVID-19 brought into our professional lives, shifted the medium of work to online tools and platforms. This came up particularly under Personal Growth in relation to the training and learning topics. It was commented that, be it technical or personal, the importance and effectiveness of on-site trainings should not be overlooked. It was discussed from the trainers point of view that tracking the learning progress of an individual becomes very challenging and observation based evaluation opportunities are lost in an online setting.

One particular aspect related to the personal growth was the consideration of time and motivation as fundamental resources for individuals' growth journey. It was commented for many individuals, that it is not necessarily about whether there are enough training offerings or external resources for a learning task, rather it is about having the dedicated hours for a particular training, or about knowing the purpose and meaning behind the learning activities.

Additional discussion regarding human centricity was also in the direction of taking action, instead of just communicating the importance of a concept. With one of the experts, the necessity of making a distinction between communicating the importance of something and actively doing something was underlined, as this was perceived as an important aspect for capturing the actual state of agility being practiced and not staying only theoretical. This distinction can be made explicit by employing verbs such as enabling, realizing or instantiating.

6.3. Technical Excellence Specific Findings

The overall findings regarding the technical excellence principle were very positive and there were no major change requests.

From one of the senior experts from the testing domain, the assessment questionnaire was described as having a good structure for self assessments and positively commented on its ability to guide the individuals. Further, clusters were perceived as very valuable for representing the technical excellence, particularly in a software development organization.

In parallel, there was the discussion about the testing cluster's level of granularity and coherence in terms of representing testing related activities such as unit testing, integration testing, system testing, their automation and systematic advanced applications. The conclusion was distributing automated testing and advanced testing methods as an element of higher level capabilities, rather than having a standalone question for them.

For technical excellence, another main discussion line was on the infrastructure and environment orchestration. When it comes to technical excellence and the quality of the software products, it was highlighted that along with the dominance of new architectures and technologies such as microservices, containerized applications and automated pipelines, the technical excellence requires organizations to have a new perspective on their delivery pipelines. Particularly at large scale product development, delivery pipelines and their automation configurations and testing activities have become a critical player in the value streams. As a result of that, the importance of paying as much attention to environments and delivery pipelines as the source code itself was stressed. Representative example could be having a containerized, shippable product, yet having failures in the container application itself.

An additional highlight under technical excellence is that, although monitoring and data collection for systems have a great potential in providing valuable insights, organizations need to identify what and why they are willing to measure certain indicators. If not, may result in overwhelming amount of data that is of little value and do not provide any actionable feedback onto system.

Another discussion point with respect to addressing the large scale product development concerns was the following: while the alignment of terminology, practices and methods across multiple teams is beneficial, it should be kept in mind that enforcing certain best practices onto other teams may result in reduced productivity. One should pay utmost attention not to fall into the alignment trap where the effort towards ensuring alignment starts negatively affecting autonomy.

Additionally, when it comes to addressing the aspects regarding the large scale, it is important to remember that while large scale product development may require handling totally new artifacts and methods, it may as well be the case that certain team level artifacts and methods may remain relevant in a slightly different form. An example to this could be, while data collection and monitoring are relevant for both on a team as well as for product level, the data to be monitored and questions that need to be answered are different for one team and for a large scale product development process.

7. Threats to Validity

In this section, we discuss the validity threats and our attempts to ensure a high quality of research by keeping these threats minimal. The initial consideration of the threats to validity have been the four well known threats as described by Yin [15]: Construct Validity, External Validity, Internal Validity and Reliability. These threats to validity have been adapted to the domain of software engineering by Runeson et al. [16]. Within the scope of this paper, as our methodological focus is not the case study research, we address three most relevant threats for our methodology.

7.1. Construct Validity

Construct validity refers to the mechanism which entitles researchers to address the mutual understanding of the concepts being discussed and vocabulary being used. To address threats on construct validity, we have provided clear examples in our interview guidelines to establish the boundaries for discussion. Further, exactly the same examples and artifacts were used across each of the interviews. We also asked interviewees if the provided examples were helpful for them to have clarity of discussion. All experts validated that providing concrete examples within the questionnaire was helpful for them.

7.2. External Validity

External validity refers to the generalizability of the generated results by the research activities. To address external validity threats, we consulted experts from different divisions, and put effort in gaining viewpoints based on geographical diversity. By doing so, our aim was to have a sample that is representative of different business strategies, work cultures and disciplines, yet, having seven experts partnering in our interviews remains as a threat to external validity. External validity is also affected by the fact that, although it is a very diverse conglomerate, all experts are within the Siemens organization, and may be expected to share common values and thinking patterns of the organization to a certain degree.

7.3. Reliability

Reliability refers to the reproducibility of the research results. If the research have been conducted by different researchers, the same results should be derived. To ensure the reliability of our research, we included additional questions and objective interview guidelines for receiving feedback on the complexity of the interview. The interview results have been provided to the experts for an additional approval step.

8. Conclusion

8.1. Summary

This study follows on the previous two steps of our research [3], [4] that we have published in SEAA'20 and XP'21 respectively. We share the results of the interview study that examines the

questionnaires we have developed with respect to the *Human Centricity* and *Technical Excellence* principles.

Our aim in questionnaire development is to reach maximum coverage with minimal set of questions, by reducing complexity. The results of our interview study shows that not only the structural components falling under human centricity and technical excellence provide sufficient coverage, the provided questionnaires both for human centricity and technical excellence have been evaluated as providing sufficient coverage for a practical evaluation as well. Overall, only one expert disagreed to use one of the two questionnaires. Final and perhaps the most essential remark for us has been that 5 out of 7 interviewees agreed that they would be willing to use one of the questionnaires in real life while 4 out of 7 stated that they would be willing to use both of the questionnaires. As already mentioned in Section 2.1.3, the development of the questionnaire was carried out on the basis of guidelines we developed beforehand. The validation of the clusters of the two principles is a good indicator for us now, that the guidelines used for questionnaire development proved to be useful and will be used for the development of the questions for the remaining clusters and principles.

8.2. Future Work

In the process of preparing this publication, we have completed the questions for the remaining clusters. First, we will be integrating the interview results in the questionnaire and implementing minor improvements as part of the intervention step of our action design research. Minor improvements regarding the example practices, and certain component name alterations will also be continuously considered. We will be finalizing our aggregation mechanism, as having an aggregation mechanism is important for being able to provide quantitative results out of the collected assessment data. Having this done we will apply the model in a pilot project in order to validate the overall suitability of the assessment model and method in an industrial setting by means of a case study.

References

[1] K. Beck, et al., Manifesto for agile software development, 2001. URL: http://www.agilemanifesto.org/.

[2] P. Rodríguez, J. Markkula, M. Oivo, K. Turula, Survey on agile and lean usage in finnish software industry, in: Proceedings of the 2012 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, IEEE, 2012, pp. 139–148.

[3] D. Tuncel, C. Körner, R. Plösch, Comparison of agile maturity models: Reflecting the real needs, in: 2020 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), IEEE, 2020, pp. 51–58.

[4] D. Tuncel, C. Körner, R. Plösch, Setting the scope for a new agile assessment model: Results of an empirical study, in: International Conference on Agile Software Development, Springer, 2021, pp. 55–70.

[5] S. Freudenberg, H. Sharp, The top 10 burning research questions from practitioners, Ieee Software 27 (2010) 8–9.

[6] T. Dingsøyr, T. E. Fægri, J. Itkonen, What is large in large-scale? a taxonomy of scale for agile software development, in: International Conference on Product-Focused Software Process Improvement, Springer, 2014, pp. 273–276.

[7] A. Sidky, J. Arthur, S. Bohner, A disciplined approach to adopting agile practices: the agile adoption framework, Innovations in systems and software engineering 3 (2007) 203–216.

[8] T. P. Rout, K. El Emam, M. Fusani, D. Goldenson, H.-W. Jung, Spice in retrospect: Developing a standard for process assessment, Journal of Systems and Software 80 (2007) 1483–1493.

[9] C. Institute, Cmmi model v2.0 in retrospect: Developing a standard for process assessment (2018).

[10] A. R. Hevner, S. T. March, J. Park, S. Ram, Design science in information systems research, MIS quarterly (2004) 75–105.

[11] M. K. Sein, O. Henfridsson, S. Purao, M. Rossi, R. Lindgren, Action design research, MIS quarterly (2011) 37–56.

[12] A. Qumer, B. Henderson-Sellers, A framework to support the evaluation, adoption and improvement of agile methods in practice, Journal of systems and software 81 (2008) 1899–1919.

[13] R. M. Fontana, V. Meyer Jr, S. Reinehr, A. Malucelli, Progressive outcomes: A framework for maturing in agile software development, Journal of Systems and Software 102 (2015) 88–108.

[14] P. Mayring, Qualitative content analysis: a step-by-step guide, SAGE, 2021.

[15] R. K. Yin, Case study research and applications, 1st Edition from 1984, Sage, 2018.

[16] P. Runeson, M. Host, A. Rainer, B. Regnell, Case study research in software engineering: Guidelines and examples, John Wiley & Sons, 2012.