



A Comprehensive Evaluation of Agile Maturity Self-assessment Surveys

Ozan Raşit Yürüm¹(✉), Onur Demirörs^{2,3}, and Fethi Rabhi³

¹ Informatics Institute, Middle East Technical University, Ankara, Turkey
oyurum@metu.edu.tr

² Department of Computer Engineering, İzmir Institute of Technology,
İzmir, Turkey
onurdemirors@iyte.edu.tr

³ School of Computer Science and Engineering,
University of New South Wales, Sydney, Australia
f.rabhi@unsw.edu.au

Abstract. Agile methodologies are adapted by growing number of software organizations. Agile maturity (also called agility) assessment is a way to ascertain the degree of this adoption and determine a course of action to improve agile maturity. There are a number of agile maturity assessment surveys in order to assess team or organization agility and many of them require no guidance. However, the usability of these surveys are not widely studied. The purpose of this study is to determine available agile maturity self-assessment surveys and evaluate their strengths and weaknesses for agile maturity assessment. An extensive case study is conducted to measure the sufficiency of 22 available agile maturity self-assessment surveys according to the seven expected features: comprehensiveness, fitness for purpose, discriminativeness, objectivity, conciseness, generalizability, and suitability for multiple assessment. Our case study results show that they do not satisfy all of the expected features fully but are helpful in some degree based on the purpose of usage.

Keywords: Agility assessment · Agility surveys
Agile maturity self-assessment · Agile maturity self-assessment surveys
Self-assessment

1 Introduction

As traditional development approaches did not produce the desired effects [1, 2] agile approaches became popular [3] specifically in largely growing SMEs [4]. This popularity has also increased the coverage and depth of agile methodologies in line with the agile manifesto [5]. Today, there are different agile software development methods, proposing different ways of achieving agile values and principles. The most popular

O. R. Yürüm—This study has been supported by Turkish Scientific and Technological Research Council of Turkey (TUBITAK), Project 113E528.

agile software development methods are Extreme Programming [6], Scrum [7], Feature Driven Development [8], Adaptive Software Development [9], Dynamic Software Development Method [10], Crystal [11], Rational Unified Process [12], Kanban [13], and Lean Software Development [14]. Each of these methods includes different practices and techniques to increase the agile maturity (e.g. agility) of an organization. Even though the underlying practices and techniques are quite different, all of these methods focus on agile values and principles and organizations frequently utilize a number of them to obtain desired benefits.

Proliferation of agile methods led to proliferation of assessment approaches for measuring agility of organizations. Assessment approaches in agile adoption frameworks [15–17] and agile maturity models [18–24] require expert judgement; therefore, professional assessors must perform assessments. It can also take a substantial amount of time depending on the size of the project or projects and detail level of the assessment required. There are also self-assessment techniques that can be used by teams or organization. It takes less time, it is cost effective and have the potential to provide much needed feedback directly to the team. Hence, agile maturity self-assessment surveys can play a crucial role for improving agile maturity. These self-assessment models are frequently called surveys and they are in the form of checklists, questionnaires, tests, or software tools.

In the existing literature, a limited number of surveys were examined in depth by independent researchers. In existing comparisons, generally two surveys which are Comparative Agility and Thoughtworks are discussed [25–28]. In addition, several outdated agile maturity assessment surveys such as Thoughtworks Agile Assessment survey [29] and Nokia Test [30] are compared. We have also compared a subset of these models by means of a case study [31]. Nevertheless, there is no study that examines all available surveys in a systematic way.

The purpose of our study presented in this paper is to extend our exploratory evaluation study [31] to all the surveys available in the literature. Our previous work has evaluated 8 self-assessment surveys while this paper includes results of 22 surveys. This study aims to provide two significant contributions. First, based on the results software organizations will be better equipped to be able to determine the most suitable survey for their needs. Secondly, the results will also depict potential areas of improvement. For this purpose, we applied all the surveys in a medium-sized software company and systematically evaluate the surveys in terms of the expected features identified in [31]: comprehensiveness, fitness for purpose, discriminativeness, objectivity, conciseness, generalizability, and suitability for multiple assessment.

The rest of this paper is organized as follows: In Sect. 2, we provide a review of literature. In Sect. 3, we describe the case study design and conduct. In Sect. 4, we present findings obtained during the case study in detailed way. Finally, in Sect. 5, we provide a conclusion and future work.

2 Related Work

The interest in agile methods led researchers or organizations to develop agile assessment approaches to assess and measure the adoption degree by evaluating a set of practices with respect to these methods. Agile assessment approaches can be categorized into three categories: (1) agile adoption frameworks, (2) agile maturity models, (3) agile self-assessment surveys.

Agile adoption frameworks were developed to guide organizations in order to adopt agile practices. Some of these frameworks [15–17] include assessment techniques as well. Assessment approaches of the adoption frameworks usually rely on expertise and cannot be performed quickly. As a result, as Jalali emphasizes that they are not used by any organization or team except their creators [27].

Agile maturity models, also known as agile reference models are developed as baselines for guiding agile transformation. Similar to process based assessment models (e.g. CMMI (Capability Maturity Model Integrated) [32] and ISO/IEC 15504 [33]) they are used for improving and assessing agility of organizations. Agile maturity models developed based on these process assessment models are [18–24]. As Lappanen depicts these models are not mature and require further work to be usable in practice [34]. The latest and most complete model is AgilityMod [21]. It has been developed coherently with agile values and principles, based on the meta-model of ISO/IEC 15504 process assessment model and validated through a number of case studies.

Agile maturity self-assessment surveys, on the other hand, are used for assessing the health of team or organization in specific time range [35]. So, there are many surveys consisting of checklists, questionnaires and tests today such as Karlskrona Test, Nokia Test, 42-Point Test, and Scrum Master Checklist [36]. They are also attracted the attention of researchers who are interested in agility assessment. Chronis analyzed 4 of the surveys which are SAFE Team, Comparative Agility, 42-Point Test, and Thoughtworks [25]. The study concludes that they do not yield similar results and a measurement tool which satisfies the needs of one team may not be suitable for other teams. There is still work to be done in order to find a universal tool for measuring agility. Leppanen in a different study states that the most significant problem of the surveys is the predefined practice expectations and difficulty of adapting to various agile software development methods [34]. Some of the surveys on the other hand try to understand agility instead of measuring agility [37]. There are also studies focusing on features related to automation of the surveys [38, 39].

In summary we can state that although there are a number of agility self-assessment surveys and a few studies related with evaluation of surveys there are no studies in the literature that evaluates all the available surveys systematically by means of a case study in a software organization.

3 The Case Study

This section explains the design and conduct of the case study. Case study was selected as a research method in order to observe the usability of existing agile maturity self-assessment surveys. Case study enables us to examine a contemporary phenomenon

within its real-life context [40]. Thus, it suits best in evaluating surveys for agility assessment of software organizations.

In this study, there is one main research question having seven sub-questions regarding the expected features of a survey (C1: Comprehensiveness, F: Fitness for purpose, D: Discriminativeness, O: Objectivity, C2: Conciseness, G: Generalizability, S: Suitability for multiple assessment).

RQ 1: To what extent do surveys cover the features that are identified in our previous study [33]?

RQ 1.1: Do the surveys meet agile practices in AgilityMod? (C1)

RQ 1.2: Are the surveys fit for purpose? (F)

RQ 1.3: Are the surveys discriminative enough to determine the agility? (D)

RQ 1.4: Do the surveys have objective questions? (O)

RQ 1.5: Do the surveys have concise questions? (C2)

RQ 1.6: Can the survey be used for all kinds of agile methods? (G)

RQ 1.7: Are the surveys suitable for multiple assessment? (S)

3.1 Design of the Case Study

In the design of our case study, we adopted the following strategies:

Survey Selection Strategy

We planned to benefit from scientific papers and search engines in order to find agile maturity self-assessment surveys. Therefore, IEEE, ScienceDirect, Web of Science and Scopus were determined to find surveys in scientific papers. In addition to the scientific papers, commercial surveys were also planned to be investigated. Two research key sets which are {Agile Assessment, Agility Assessment} and {Survey, Test, Questionnaire} were determined to identify existing agile maturity self-assessment surveys. The criterion for selection from those identified surveys is the availability of agile maturity self-assessment surveys.

Case Selection Strategy

Our strategy was to select an organization having results about their agility assessment. So, we planned to perform assessment on an organization in which assessors had already assessed its agility. The reason for choosing such an organization was that we were able to access to evidences about weaknesses and strengths of the organization related to its agility. Therefore, an organization, which had already assessed by professional assessors according to AgilityMod reference model would be selected. In addition, we planned to select a different organization from the organization in our previous exploratory case study.

Data Collection Strategy

In order to record data, we use a spreadsheet consisting of aspects and practices of AgilityMod, a form consisting of 7 expected features and assessment reports. The assessor having experience on agility assessment based on AgilityMod reference model was expected to match each question with the practices in AgilityMod and fill the form during the assessment. At the end of the assessment, the assessment report is obtained

about the agility of the organization. We also planned to examine this assessment report in terms of expected features.

Data Analysis Strategy

Our plan was firstly to write the number of questions for each practice after finishing the assessment. Then, we aimed to calculate the number of covered practices according to existence or absence of a question for each practice in order to determine comprehensiveness. The following table shows an example analysis about the determination of comprehensiveness.

Table 1. Example analysis for comprehensiveness.

Aspect OR Aspect Attribute	Number of Question per Practice						Covered / Total				
	P1	P2	P3	P4	P5	P6					
Exploration (E.A.)	P1	P2	P3	P4	P5	P6	3/6				
	1	1	0	0	1	0					
Construction (C.A.)	P1		P2		P3		P4		2/4		
	1		2		0		0				
Transition (T.A.)	P1	P2		P3	P4		P5		P6		3/6
	0	1		0	1		1		0		
Management (M.A.)	P1	P2	P3	P4	P5	P6	P7	P8	7/8		
	2	3	0	1	4	2	5	3			
Iterative	GP 2.1.1				GP 2.1.2				2/2		
	9				9						
Simple	GP 2.2.1				GP 2.2.2				2/2		
	2				2						
Technically Excellent	GP 3.1.1				GP 3.1.2				1/2		
	5				0						
Learning	GP 3.2.1		GP 3.2.2		GP 3.2.3		GP 3.2.4		2/4		
	8		0		1		0				
Total Practice (over 34)								22/34			

AgilityMod [21] has 4 aspects and 4 aspect attributes. The aspects are Exploration, Construction, Transition and Management. The aspect attributes are Iterative, Simple, Technically Excellent, and Learning. Each aspect and aspect attribute have certain number of practices. There are totally 34 practices in AgilityMod. Therefore, Table 1 shows distribution of number of questions per practice and number of practices covered by a survey with questions.

In order to find the objectivity and conciseness, we concentrated on distribution of concise and objective questions in each practice. Then, we would determine objectivity and conciseness of the survey according to number of “Largely Achieved”, and “Fully Achieved” practices. The following formula shows how objectivity is calculated for each survey.

$$Objectivity = (\# \text{ of FA Practices} + \# \text{ of LA Practices}) / (\# \text{ of Total Covered Practices})$$

Table 2 shows an example about the analysis of results in terms of objectivity.

Table 2. Example analysis for objectivity.

Rating	FA	LA	PA	NA	Objectivity (Total)	Objectivity (Percentage)
# of Practices	10	5	5	2	15/22	68.2

The same formula is used for determining the conciseness of the surveys.

$$Conciseness = (\# \text{ of FA Practices} + \# \text{ of LA Practices}) / (\# \text{ of Total Covered Practices})$$

Table 3 shows an example about the analysis of the results in terms of conciseness.

Table 3. Example analysis for conciseness.

Rating	FA	LA	PA	NA	Objectivity (Total)	Objectivity (Percentage)
# of Practices	18	4	0	0	22/22	100.0

In order to find sufficiency of other features, we decided to develop a case description for each case. We would evaluate the case description and form including seven features with content analysis. That is, we would match the findings with the criteria, and then rate each criterion for a survey according to 4-point scale.

Validation Strategy

After performing assessment with each survey, we planned to prepare assessment reports and discuss the results with an expert of both process improvement and agile software development methodologies.

3.2 Conduct of the Case Study

According to literature review, 22 available self-assessment surveys shown in Table 4 were determined for main case study. Table 4 shows available agile maturity self-assessment surveys with their name, owner, type, and number of questions they include.

The organization that we selected for the case study is one of the leading media companies in Turkey with its 17 million unique visitors on its various internet platforms. An ongoing online video platform project including 9 software developers, 2 graphical user interface designers, 2 business intelligence analysts, 1 tester and 8 content providers was assessed according to AgilityMod reference model. The evidence was collected from two project managers, a software team leader, and a graphical

Table 4. List of agile maturity self-assessment surveys.

Name	Survey owner	Type	Number of Questions/Items
42-Point Test [41]	Kelly Waters	Yes/No	42
Agile 3R Model of Maturity Assessment [42]	PhaniThimmapuram	5 point likert scale	11
Agile Assessment [43]	Piotr Nowinski	Yes/No	66
Agile Karlskrona Test [44]	Mark Seuffert	Multiple Choice	11
Agile Maturity Self-assessment [45]	Bryan Campbell &Robbie Mac Iver	Multiple Choice (2 Option)	6
Agile Maturity Self-Assessment Survey [46]	Eduardo Ribeiro	Yes/No	26
Agile Team Evaluation [47]	Eric Gunnerson	Yes/No	17
AgileTest [48]	ACM	5 point likert scale	14
Agility Questionnaire [49]	Marcel Britsch	6 point likert scale	60
Borland Agile Assessment [50]	Borland	7 point likert scale	12
Cargo Cult Agile Checklist [51]	Stefan Wolpers	Yes/No	25
Comparative Agility [52]	Mike Cohn and Kenny Rubin	5 point likert scale	65
Corporate Agile 10-point Checklist [53]	Elena Yatzeck	5 point likert scale	10
Depth of Kanban [54]	Christophe Achouiantz	5 point likert scale	69
IBM's Scaled Agile Framework © (SAFe™) Team Self-Assessment [55]	IBM	Yes/No, 5 point scale, Multiple Choice, Open Question	38
Maturity Assessment Model for Scrum Teams [56]	MarmamulaPrashanth Kumar	5 point likert scale	15
SAFe Team Self Assessment [57]	Scaled Agile	5 point likert scale	25
Scrum Checklist [58]	Henrik Kniberg	5 point likert scale	80
Scrum Master Checklist [59]	Michael James	5 point likert scale	42
Scrum Self Assessment [60]	Cape Project Management	Yes/No	60
Team Barometer [61]	Jimmy Janlén	Yes/No	16
The Art of Agile [62]	James Shore	Yes/No	46

user interface designer via interviews. The assessment results based on AgilityMod reference model showed that exploration and construction aspects of the organization were in first levels while its transition and management aspects were not implemented.

During this study, agility assessment was performed again with each available agility assessment survey in the light of assessment results and the evidence collected from the organization. Since we reached 22 agility assessment surveys, we performed 22 different assessments with the same organization in different time periods.

As mentioned in data collection part, we took notes in a form according to seven evaluation criteria and examined each question to match them with the practices in AgilityMod. After finishing the assessment, we obtained an assessment result based on related survey if the survey supports to generate assessment result.

We analyzed the forms, the spreadsheets, the assessment results, and case descriptions as mentioned in data analysis part in case study design section. Then, we rated each agility assessment surveys using the same rating approach [63] defined in ISO/IEC 15504 in terms of each expected feature. After rating, the results were presented to an expert who has more than 10 years' experience on process assessment and 2 years' experience on agile software development.

3.3 Validity Threats

The survey-based assessments were performed by one of the authors of this paper. He has assessment experience based on AgilityMod [21], CMMI [32] and ISO/IEC 15504 [33] since 2014. However, he had not taken active role on the previous assessment process of the selected case based on AgilityMod reference model. Therefore, there was a possibility to assess agility of the case incorrectly. In order to avoid this possibility, he examined the expert-based case study results and evidences in a detailed way before starting the case study with each agile maturity survey. Since he has enough experience on agility assessment process, it was easy to understand and adopt the expert-based case study results. In addition, after performing survey-based assessments, the results were discussed with an expert of both process improvement and agile software development methodologies in order to eliminate any bias of the assessor. We selected all possible agile maturity assessment surveys since our aim was a comprehensive evaluation. Although the evaluation is performed based a single case as the properties are related with the surveys not with the case it is unlikely to find further insight through replication. In terms of construct validity, which refers to the degree of measuring what is expected to measure, we selected an organization having enough number of indicators for agility assessment so that we can evaluate the measurement capability of the surveys.

3.4 Limitations

This study is limited to the agile maturity self-assessment surveys published on the time where the case study was performed. After that time, it is possible that new versions of the surveys have been developed by their owners. In addition, new agile maturity self-assessment could have been published during this study. However, we only take into account the surveys obtained at the end of literature review.

4 Results

The assessment results obtained at the end of the case study for each agile maturity self-assessment survey according to seven features are shown in the following Table 5.

Comprehensiveness

Assessment results show that there is no survey that can be called fully comprehensive. In other words, the surveys do not focus on all agility aspects. Many of them concentrate on management aspects only and exploration, construction, and transition

Table 5. Case study results.

Survey/Criteria Legend FA: Fully achieved LA: Largely achieved PA: Partially achieved NA: Not achieved	Comprehensiveness	Fitness for purpose	Discriminateness	Objectivity	Conciseness	Generalizability	Suitability for multiple assessment
42-Point Test [41]	PA	PA	PA	FA	LA	FA	LA
Agile 3R Model of Maturity Assessment [42]	PA	PA	LA	PA	LA	FA	FA
Agile Assessment [43]	LA	PA	LA	LA	LA	FA	FA
Agile Karlskrona Test [44]	PA	PA	LA	FA	FA	FA	NA
Agile Maturity Self-assessment [45]	PA	PA	LA	PA	PA	FA	NA
Agile Maturity Self-Assessment Survey [46]	PA	PA	NA	PA	FA	FA	NA
Agile Team Evaluation [47]	PA	PA	NA	PA	LA	FA	NA
AgileTest [48]	PA	PA	NA	LA	FA	FA	NA
Agility Questionnaire [49]	PA	PA	LA	LA	LA	FA	NA
Borland Agile Assessment [50]	PA	PA	NA	PA	LA	FA	NA
Cargo Cult Agile Checklist [51]	PA	PA	LA	FA	LA	PA	NA
Comparative Agility [52]	LA	LA	LA	LA	FA	FA	NA
Corporate Agile 10-point Checklist [53]	PA	NA	NA	PA	NA	FA	NA
Depth of Kanban [54]	PA	LA	FA	FA	FA	PA	NA
IBM's Scaled Agile Framework © (SAFe™) Team Self-Assessment [55]	PA	PA	PA	LA	NA	FA	FA
Maturity Assessment Model for Scrum Teams [56]	PA	PA	LA	PA	LA	FA	FA
SAFe Team Self Assessment [57]	PA	PA	PA	LA	PA	FA	NA
Scrum Checklist [58]	LA	PA	NA	FA	FA	PA	NA
Scrum Master Checklist [59]	PA	PA	NA	LA	PA	PA	NA
Scrum Self Assessment [60]	LA	PA	LA	FA	FA	PA	NA
Team Barometer [61]	NA	PA	PA	NA	NA	FA	FA
The Art of Agile [62]	PA	LA	FA	LA	PA	PA	NA

aspects are not well covered by the surveys. In addition, the surveys generally focus on specific aspect attributes such as iterative and learning whereas they disregard the attributes related to simplicity and technically excellence. There are four surveys that we call largely comprehensive. These surveys are Comparative Agility, Scrum Self-Assessment, Scrum Checklist, and Agile Assessment. While Comparative Agility and Agile Assessment include Likert type questions, Scrum Self-Assessment and Scrum Checklist include true/false questions. They include sufficient number of questions for more practices when compared to others. Team Barometer is the least comprehensive survey as most of the questions focus on same practice that is “support collaborative work and shared responsibility” of the learning aspect.

Fitness for Purpose

Even though the surveys claim that they are developed with the aim of assessing agility of a team or an organization, there is no survey that meets fully the feature “fitness for purpose”. They do not generate assessment results that include the improvement opportunities and suggestions. The purpose of the assessment is to guide the organization for continuous improvement as well as identifying problems. Almost all of the surveys focus to identify problems only. Therefore, many of the surveys are partly suitable for assessment purpose. There are only three surveys that largely achieve fitness for purpose. While The Art of Agile does not give the details about the improvement opportunities, Comparative Agility and The Depth of Kanban do not provide improvement suggestions to increase the agility level of the organization. The common positive property of these three surveys is to support reporting of the assessment. Corporate Agile 10-point Checklist does not meet the feature “fitness for purpose” at least partially. It does not include assessable items or questions. Furthermore, it does not support showing any improvement opportunity or improvement suggestion.

Discriminativeness

Assessment results show that Depth of Kanban and The Art of Agile surveys achieve discriminativeness fully. In these surveys, there are defined agility levels that show the agility degree of an organization based on the specific scoring range. Both surveys include four agility levels. In The Art of Agile, three of them are defined with colors: Red, Yellow, Green while one of them indicating 100 percent agile is not defined with any color since color also shows risk level of the organization. In Depth of Kanban, the level names are different. These levels are “Necessary for Sustainable Improvements”, “Improving Sustainably”, “Excellence” and “Lean”. All of these are indicated with colors in radar chart. In addition, both survey results include areas or features that yield more specific results about the agility of the organization. In Depth of Kanban, these are related to 7 properties of Kanban: Visualize Effects, Improve, Feedback Loops, Explicit Policies, Manage Flow and Limit WIP. In The Art of Agile, these are based on phases of Extreme Programming: Thinking, Collaborating, Releasing, Planning, and Developing. There are also some practices to increase the agility in both surveys. Apart from these surveys, less than half of the surveys are largely discriminative. Comparative Agility, Agile Karlskrona Test, Agile Maturity Self-Assessment, Cargo Cult Agile Checklist, Agile 3R Model of Maturity Assessment, Maturity Assessment Model for Scrum Teams, Scrum Self-Assessment, Agility Questionnaire, and Agile Assessment are achieved largely in terms of discriminativeness. They do not include either agility level for the organization or scoring for areas/aspects of agility. The remaining are either partially achieved or not achieved. The general reason of not meeting discriminativeness fully is the missing comprehensive assessment methods in the surveys.

Objectivity

Approximately quarter of the surveys, include measurable questions that lead to objectivity: 42-Point Test, Agile Karlskrona Test, Cargo Cult Agile Checklist, Depth of Kanban, Scrum Self-Assessment, Scrum Checklist. The scaling type of most of these surveys is true/false. They are generally checklists or marked as true or false. Only Agile Karlskrona Test has multiple-choice questions. From other surveys, Comparative

Agility, Scrum Master Checklist, The Art of Agile, Agility Questionnaire, SAFe Team Self-Assessment, IBM's Scaled Agile Framework © (SAFe™) Team Self-Assessment, Agile Test, Agile Assessment are largely objective. Many of them include questions that are subjective such as “the team is more productive” or “the team produces higher quality products”. Other surveys except Team Barometer are partially objective. They use words leading to subjective answers such as “good”, “better”, and “well”. Team Barometer does not include sufficient questions related to practices that it covers in term of objectivity since it focuses on the ideas of team members about their team's agility.

Conciseness

There are eight surveys that meet conciseness fully. These surveys are Comparative Agility, Agile Karlskrona Test, Depth of Kanban, Scrum Self-Assessment, Scrum Checklist, Agile Maturity Self-Assessment Survey, Agile Test, and Agile Assessment. They are asking one-question at a time. In addition, the lengths of the questions are short enough to be easily comprehended. While 42-Point Test, Borland, Cargo Cult Agile Checklist, Agile 3R Model of Maturity Assessment, Maturity Assessment Model for Scrum Teams, Team Agile Evaluation, Agility Questionnaire, and SAFe Team Self-Assessment are achieved largely in terms of conciseness, Agile Maturity Self-Assessment, Scrum Master Checklist, The Art of Agile are partially achieved. Most of these include questions that ask two different things at once. An example from 42-Point Test is the item “Software is tested and working at the end of each sprint/iteration.” Testing software and delivering working software are two different things that need to be considered separately. The other example is “Team members trust each other and are motivated to deliver sprint deliverables” from Agile 3R Model of Maturity Assessment. Trusting each other and being motivating are two different things. Their general property is that they are asked at once with a conjunction such as and, but etc. The remaining three surveys which are Team Barometer, IBM's Scaled Agile Framework © (SAFe™) Team Self-Assessment and Corporate Agile 10-point Checklist have not achieved conciseness even partially. They usually include more than one question for each item.

Generalizability

Most of the surveys focus on generic methods rather than focusing on one specific method. Apart from six surveys that are partially generalizable, others are suitable for assessment in the context of variety of agile methodologies. From the surveys, six of them focus on only one agile software development method. These surveys are Cargo Cult Agile Checklist, Depth of Kanban, Scrum Master Checklist, Scrum Self-Assessment, The Art of Agile, and Scrum Checklist. That is, they are not generalizable to all agile software development methodologies. Four of them, which are Cargo Cult Agile Checklist, Scrum Master Checklist, Scrum Self-Assessment, and Scrum Checklist, are based on Scrum, while Depth of Kanban is based on Kanban and The Art of Agile is based on Extreme Programming. In other words, these surveys are developed to be compatible with only one specific agile software development method. They are partially applicable to other methods since all agile software development methodologies share same principles.

Suitability for Multiple Assessment

Five of the surveys are fully suitable for multiple assessment. That is, they give each member a chance to rate each item. According to ratings of each team member, average rating value is determined for each item. Then, this value is used in determining the agility of the organization for the specific area. In addition, the analysis of multiple assessment is performed and depicted in detail in the assessment result. The fully achieved surveys are Team Barometer, Agile 3R Model of Maturity Assessment, Maturity Assessment Model for Scrum Teams, IBM's Scaled Agile Framework © (SAFe™) Team Self-Assessment, and Agile Assessment. 42-Point Test survey is largely suitable for multiple assessment. It allows each team member of an agile team to fill the survey and it gives superficial information about the result. However, it is not possible to get detailed analysis from the survey. Other surveys do not support multiple assessment. They are suitable for single agility assessment. In other words, only one person can perform self-assessment. with most of the surveys.

5 Conclusion

In this study, twenty-two available agile maturity self-assessment surveys are evaluated by means of a case study in terms of Comprehensiveness, Fitness for Purpose, Discriminativeness, Objectivity, Conciseness, Generalizability, and Suitability for Multiple Assessment. The case study results support the results of our previous study in the way that none of the agile maturity self-assessment surveys has fully satisfied the expected features. We found that, comprehensiveness, and fitness for purpose are the most problematic features that are not fully achieved by any of the surveys. While four surveys are largely comprehensive, three surveys are largely fit for the purpose. Only Comparative Agility meets largely both features. We also found that there are a number of surveys that are generic enough to be used by a variety of agile methodologies and there are also some surveys that enable multiple assessment.

From twenty-two surveys, Comparative Agility and Agile Assessment had six features which are largely or fully achieved. However, both surveys also have significant improvement opportunities. Comparative Agility meets only two features, which are conciseness and generalizability completely. This survey is largely comprehensive, fit for the purpose, discriminative, and objective. Nevertheless, it need to focus on more practices, establish ways to suggest practices for improvement, should have agility level definitions and more measurable questions. In addition, it needs to be improved to support multiple assessments. Like Comparative Agility, Agile Assessment also has significant improvement opportunities. Especially, it has serious deficiencies about fitness for purpose.

When all surveys are examined, it is seen that almost all of them have at least one fully achieved feature depicting how this feature can be implemented. For example, IBM's Scaled Agile Framework © (SAFe™) Team Self-Assessment is a good example for multiple assessment. Or some surveys can be used by organizations who adopt specific agile methodology. For example, while Depth of Kanban is more useful for organizations implementing Kanban, Scrum Self-Assessment is more useful for organizations implementing Scrum when compared to others. Comparative Agility is

suitable for organizations to measure their agility from generic perspective. In other words, organizations can select the most suitable survey according to their priorities based on the results of this study. Organizations should consider that each survey has certain limitations even though they have some good features.

As a further research, we are planning to increase the number of case studies. In addition, we are planning to measure the effects of self-assessment surveys on success of organization's agility by using some of them repetitively over a long term period. We hope this study will establish a baseline for improving the usability of available surveys and lead to the development of new surveys.

References

1. Uskarcı, A., Demirörs, O.: Do staged maturity models result in organization-wide continuous process improvement? insight from employees. *Comput. Stand. Interfaces* **52**, 25–40 (2017)
2. Uskarcı, A., Demirörs, O.: Causes of continuity and participation problems in process improvement with staged maturity models. *Commun. Comput. Inf. Sci.* **526**, 177–178 (2015)
3. Garousi, V., Coşkunçay, A., Betin-Can, A., Demirörs, O.: A survey of software engineering practices in Turkey. *J. Syst. Softw.* **108**, 148–177 (2015)
4. Garousi, V., Coşkunçay, A., Demirörs, O., Yazici, A.: Cross-factor analysis of software engineering practices versus practitioner demographics: an exploratory study in Turkey. *J. Syst. Softw.* **111**, 49–73 (2016)
5. Beck, K., et al.: Manifesto for Agile Software Development (2001). <http://www.agilemanifesto.org/>
6. Beck, K.: *Extreme Programming Explained: Embrace Change*, 1st edn. (1999)
7. Schwaber, K., Sutherland, J.: *The Scrum Guide*, July 2014
8. Palmer, S.R., Felsing, J.M.: *A Practical Guide to Feature-driven Development*. Prentice Hall PTR, So Paulo (2002)
9. Highsmith, J.: *Adaptive Software Development: A Collaborative Approach to Managing Complex Systems*. Dorset House Publishing, New York (2000)
10. Wells, T.: *Dynamic Software Development: Managing Projects in Flux*. CRC Press LLC, Florida (2003)
11. Cockburn, A.: *Crystal Clear: A Human-Powered Methodology for Small Teams*. Addison-Wesley Professional, Boston (2004)
12. Kruchten, P.: *The Rational Unified Process: An Introduction*, 3rd edn. Addison-Wesley Professional, Boston (2004)
13. Stellman, A., Greene, J.: *Learning Agile: Understanding Scrum, XP, Lean, and Kanban*. O'Reilly Media Inc., Sebastopol (2014)
14. Poppendieck, M., Poppendieck, T.: *Lean Software Development: An Agile Toolkit*. Addison-Wesley Professional, Boston (2003)
15. Qumer, A., Henderson-Sellers, B.: A framework to support the evaluation, adoption and improvement of agile methods in practice. *J. Syst. Softw.* **81**(11), 1899–1919 (2008)
16. Sidky, A., Arthur, J.: A disciplined approach to adopting agile practices: the agile adoption framework. *Innov. Syst. Softw. Eng.* **3**(3), 203–216 (2007)

17. Sureshchandra, K., Shrinivasavadhani, J.: Adopting agile in distributed development. In: Proceedings of the 2008 3rd IEEE International Conference on Global Software Engineering ICGSE 2008, pp. 217–221 (2008)
18. Packlick, J.: The agile maturity map a goal oriented approach to agile improvement. In: Proceedings of the Agil 2007, pp. 266–271 (2007)
19. Patel, C., Ramachandran, M.: Agile maturity model (AMM): a software process improvement framework for agile software development practices. *Int. J. Softw. Eng.* **2**(1), 3–28 (2009)
20. Ambler, S.: The agile scaling model (ASM): adapting agile methods for complex environments. *Environments* 1–35 (2009)
21. Ozcan-Top, O., Demirörs, O.: A reference model for software agility assessment: AgilityMod. In: Rout, T., O'Connor, R.V., Dorling, A. (eds.) *SPICE 2015*. CCIS, vol. 526, pp. 145–158. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-19860-6_12
22. Türetken, O., Stojanov, I., Trienekens, J.J.M.: Assessing the adoption level of scaled agile development: a maturity model for scaled agile framework. *J. Softw. Evol. Process* **29**(6), e1796 (2016)
23. Yin, A.P.G.: *Scrum Maturity Model*. Technical University of Lisbon (2011)
24. Nawrocki, J., Walter, B., Wojciechowski, A.: Toward maturity model for eXtreme programming. In: Proceedings of the 27th EUROMICRO Conference, pp. 233–239 (2001)
25. Chronis, K.: *Measuring Agility: A Validity Study on Tools Measuring The Agility Level of Software Development Teams*. University of Gothenburg (2015)
26. Gandomani, T.J., Nafchi, M.Z.: Agility assessment model to measure Agility degree of Agile software companies. *Indian J. Sci. Technol.* **7**(7), 955–959 (2014)
27. Jalali, S., Wohlin, C., Angelis, L.: Investigating the applicability of agility assessment surveys: a case study. *J. Syst. Softw.* **98**, 172–190 (2014)
28. Nafchi, M.Z., Zulzalil, H., Gandomani, T.J.: On the current agile assessment methods and approaches. In: 2014 8th Malaysian Software Engineering Conference MySEC 2014, pp. 251–254 (2014)
29. Thoughtworks: *Agile Assessment* (2010). <https://www.agileassessments.com>
30. Sutherland, J.: *ScrumButt Test aka the Nokia Test* (2009). <http://jeffsutherland.com/ScrumButtTest.pdf>
31. Yürüm, O.R., Demirörs, O.: Agile maturity self-assessment surveys: a case study. In: 43th Euromicro Conference on Software Engineering and Advanced Applications (SEAA) (2017)
32. CMMI Product Team: *CMMI® for Development, Version 1.3 CMMI-DEV, V1.3*. Software Engineering Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania (2010)
33. ISO/IEC: *ISO 15504-5 Information Technology - Process Assessment - Part 5: An exemplar Process Assessment Model* (2006)
34. Leppanen, M.: A comparative analysis of agile maturity models. In: Pooley, R., Coady, J., Schneider, C., Linger, H., Barry, C., Lang, M. (eds.) *Information Systems Development: Reflections, Challenges and New Directions*, Information Systems Development, vol. 1, pp. 329–343. Springer, New York (2013). https://doi.org/10.1007/978-1-4614-4951-5_27
35. Diebold, P., Ostberg, J.-P., Wagner, S., Zandler, U.: What do practitioners vary in using scrum? In: Lassenius, C., Dingsøyr, T., Paasivaara, M. (eds.) *XP 2015*. LNBP, vol. 212, pp. 40–51. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-18612-2_4
36. Nebe, K., Baloni, S.: Agile human-centred design: a conformance checklist. In: Yamamoto, S. (ed.) *HIMI 2016*. LNCS, vol. 9734, pp. 442–453. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-40349-6_42

37. Guillot, I., Paulmani, G., Kumar, V., Fraser, S.N.: Case studies of industry-academia research collaborations for software development with agile. In: Gutwin, C., Ochoa, S.F., Vassileva, J., Inoue, T. (eds.) CRIWG 2017. LNCS, vol. 10391, pp. 196–212. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-63874-4_15
38. Adalı, O.E., Özcan-Top, Ö., Demirörs, O.: Evaluation of agility assessment tools: a multiple case study. In: Clarke, P.M., O'Connor, R.V., Rout, T., Dorling, A. (eds.) SPICE 2016. CCIS, vol. 609, pp. 135–149. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-38980-6_11
39. Yürüm, O.R., Top, Ö.Ö., Demirörs, O.: Assessing software processes over a new generic software process assessment tool. *Coll. Econ. Anal.* (43), 1–30 (2017)
40. Yin, R.K., Case study research design and methods. **26**(1) (2003)
41. Waters, K.: How Agile Are You? (Take This 42 Point Test) (2008)
42. Thimmapuram, P.: Agile 3R Model of Maturity Assessment (2015). <https://www.scrumalliance.org/community/articles/2015/march/agile-3r-model-maturity-assessment>
43. Nowinski, P.: Agile Assessment (2016). <https://nowinskiopiotr.wordpress.com/2016/04/29/agile-assessment/>
44. Marberg Consulting: Karlskrona Test Online (2008). <http://mayberg.se/learning/karlskrona-test-online>
45. Campbell, B., Mac Iver, R.: Agile Maturity Self-Assessment. http://www.robbiemaciver.com/documents/presentations/A2010-Agile_Maturity_Self-Assessment.pdf
46. Ribeiro, E.: Agile Maturity Self-Assessment Survey (2015). <https://www.scrumalliance.org/community/articles/2015/december/agile-maturity-self-assessment-survey>
47. Gunnerson, E.: Agile Team Evaluation (2015). <https://blogs.msdn.microsoft.com/ericgu/2015/10/05/agile-team-evaluation/>
48. ACM: Agile Test (2016). <http://www.acm-software.com/agile-test/>
49. Britsch, M.: Assessing your Client's Agility - An Agility Questionnaire (2014). <http://www.thedigitalbusinessanalyst.co.uk/2014/07/Agile-Questionnaire.html>
50. Borland: Borland Agile Assessment 2009 (2009). http://borland.typepad.com/agile_transformation/2009/03/borland-agile-assessment-2009.html
51. Wolpers, S.: Cargo Cult Agile: The 'State of Agile' Checklist for Your Organization. <https://age-of-product.com/cargo-cult-agile-state-agile-checklist-organization/>
52. Cohn, M., Rubin, K.: Comparative Agility (2008). <https://comparativeagility.com/>
53. Yatzeck, E.: A Corporate Agile 10-point Checklist (2012). <http://pagilista.blogspot.com.tr/2012/12/a-corporate-agile-10-point-checklist.html>
54. Achouiantz, C.: Depth of Kanban - A Good Coaching Tool (2013). <http://leanagileprojects.blogspot.com.tr/2013/03/depth-of-kanban-good-coaching-tool.html>
55. DevOps Community: IBM's Scaled Agile Framework © (SAFe™) Team Self-Assessment (2015). [https://www.ibm.com/developerworks/community/wikis/home?lang=es#!/wiki/W54ecb028c53d_48b0_9d5e_4584a00489d3/page/IBM%E2%80%99s%20Scaled%20Agile%20Framework%20C2%AE%20\(SAFe\)%20Related%20Webcasts](https://www.ibm.com/developerworks/community/wikis/home?lang=es#!/wiki/W54ecb028c53d_48b0_9d5e_4584a00489d3/page/IBM%E2%80%99s%20Scaled%20Agile%20Framework%20C2%AE%20(SAFe)%20Related%20Webcasts)
56. Kumar, M.P.: Maturity Assessment Model for Scrum Teams (2014). <https://www.scrumalliance.org/community/articles/2014/july/maturity-assessment-model-for-the-scrum-teams>
57. Scaled Agile: Team Self-Assessment - Scaled Agile Framework (2008). <http://www.scaledagileframework.com/?wpdmact=process&did=NjEuaG90bGluaw==>
58. Kniberg, H.: The Unofficial Scrum Checklist (2010). <https://www.crisp.se/wp-content/uploads/2012/05/Scrum-checklist.pdf>
59. James, M.: An Example Checklist for ScrumMasters (2007). http://scrummasterchecklist.org/pdf/ScrumMaster_Checklist_12_unbranded.pdf

60. Cape Project Management: Scrum Self Assessment (2015). <http://www.agileprojectmanagementtraining.com/Agile-Self-Assessment/quiz.html>
61. Janlén, J.: Team Barometer (Self-Evaluation Tool) (2014). <http://blog.crisp.se/2014/01/30/jimmyjanlen/team-barometer-self-evaluation-tool>
62. Shore, J., Warden, S.: The Art of Agile Development. O'Reilly Media Inc., Beijing (2008)
63. ISO/IEC: ISO 15504-2 Information Technology - Process Assessment - Part 2: Performing an Assessment (2003)