Measuring Team Effectiveness in Scrum

Kars Beek[1,2(✉)], Gerard Wagenaar[1], Laura Kester[2], Sietse Overbeek[1], and Evert de Rooij[2]

[1] Faculty of Science, Department of Information and Computing Sciences, Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands
{g.wagenaar,s.j.overbeek}@uu.nl
[2] Info Support, Kruisboog 42, 3905 TE Veenendaal, The Netherlands
{kars.beek,laura.kester,evert.derooij}@infosupport.com

Abstract. Teams have become building blocks of organizations, leading to an exponential increase in team studies, including team effectiveness studies in Scrum software development. However, research on measuring Scrum team effectiveness based on objective measures, contrary to self-reporting with Likert scales, is absent. Through a design science research methodology with literature review, focus groups, interviews, and an expert panel, 29 objective measures were identified contributing to seven team effectiveness concepts. All measures can be quantified or directly derived from work management systems, such as Jira or Azure DevOps. Examples include the number of solved retrospective items after a new sprint, contributing to the team effectiveness concept 'Continuous Improvement', and the number of times a sprint goal has been achieved, contributing to both 'Team Morale' and 'Stakeholder Satisfaction'. In this way, the study offers proof of the benefits of agile, especially Scrum, software development through effective teams as well as providing practitioners a first insight in benchmarking their Scrum team effectiveness.

Keywords: Team Effectiveness · Scrum · Objective measures

1 Introduction

Scrum is one of the most popular methods in software development [14]. The Scrum framework consists of a framework that offers a way of team collaboration for solving complex problems [20]. As a result of the popularity of Scrum, research on the topic of teams has grown exponentially in the last decade, leading also to an increase in team effectiveness studies in Scrum [12]. However, observations show that all studies that address the topic of Scrum team effectiveness measure team effectiveness based on self-reporting. In other words, providing a personal opinion on a situation or question. A symptom of self-reporting is a Likert scale. A Likert scale is a rating scale that expresses the subjectivity of individuals [9]. Subjectivity in measures can bring limitations [8]. First, subjective measures are difficult to aggregate and interpret because they are often expressed on ordinal scales. Moreover, it has been noticed that these measures are not

J. M. Fernandes et al. (Eds.): QUATIC 2023, CCIS 1871, pp. 233–247, 2023.
https://doi.org/10.1007/978-3-031-43703-8_17

correlated with facts from the field. As a result, subjectivity in measures has limitations, according to research [8]. Therefore, the objective of this research is to find out:

RQ: To which extent can team effectiveness in Scrum be measured based on objective measures?

In this paper, measures that quantify team effectiveness in Scrum will be presented. Previous findings on the topic of team effectiveness will be discussed in Sect. 2. In Sect. 3, the research design is described. Section 4 presents the findings of the research. In Sects. 5 and 6 the findings will be discussed, and validity threats will be examined. Last, in Sect. 7, the conclusions of this paper will be presented.

2 Previous Findings

This section discusses previous findings on the definition of team effectiveness in the literature. In addition, related work will be used to compile information on the different methods to measure team effectiveness in other research areas and in Scrum.

2.1 A Definition of Team Effectiveness

There is a significant amount of ambiguity regarding the concept of team effectiveness [3]. This is mainly since different organizations have different views on what defines "effectiveness" [1].

Without giving a formal definition, Hackman [6] states that in addition to performance outcomes, such as speed to solution and the number of errors, other outcomes should also be taken into account, for example, group cohesiveness and member satisfaction, to determine the effectiveness of a team. A general observation was that literature addressed the above-mentioned criteria for team effectiveness. However, a general definition of Team Effectiveness is often lacking.

One of the few definitions of 'team effectiveness' found has been given by Fransen et al. [5], and defines team effectiveness as, "*the quality of team performance, as well as the perceived satisfaction with individual needs of team members*". This definition addresses team effectiveness at the team level (that is, performance) and the individual level (that is, satisfaction of team members). As a result, the definition of Fransen [5] can be applied in a broader context and has therefore been used during this research.

2.2 Measuring Team Effectiveness

Since the authors found no studies on measuring team effectiveness in Scrum based on objective measures, other disciplines have been visited. In this case, an exploratory literature review has been done in the fields of healthcare and engineering. These two research areas comprised the majority of team effectiveness studies. Following the number of team effectiveness papers, it can be assumed that the papers in these research areas have a respectable level of team effectiveness maturity.

Healthcare. Two meta-review studies provide an overview of measuring team effectiveness within the healthcare discipline. One of the first studies in this research area [11],

reviewed 22 studies and concluded that all 22 studies applied objective measures and only 4 of the 22 studies also used subjective measures. A follow-up study [2] shows that this trend has not changed. Examples of objective measures can be categorized into patient outcomes (e.g., the functional status of a patient), and organizational outcomes (e.g., costs). Although no numbers are provided in studies that applied subjective or objective measurement, the study advised that research on the topic of team effectiveness in healthcare includes outcomes less frequently used, such as professional well-being, that is, staff satisfaction, and focuses on identifying possible deadly combinations between outcomes.

Engineering. For engineering, the studies done only contained subjective measures. Multiple studies have been conducted [25, 4, 7], and all of these studies applied Likert scales and therefore subjective measures to measure team effectiveness. Unfortunately, these papers do not elaborate on why they contain only subjective measures.

The general conclusion of this section is that there are a variety of methods of measuring team effectiveness in the areas of engineering and healthcare research. Although most healthcare research teams measure the effectiveness of the team based on objective measures, most effectiveness studies of engineering teams apply subjective measures.

2.3 Measuring Team Effectiveness in Scrum

The first study to address the topic of team effectiveness in Scrum was by Moe et al. [13]. In this study, to evaluate team effectiveness in Scrum, the "Big Five" teamwork [19] has been applied. Although this paper included a small case study and is already more than 15 years old, it provides a good understanding of the relationship between team effectiveness and Scrum. Furthermore, the findings of Moe et al. [13] form the basis for future research in the area of team effectiveness in Scrum, such as the paper by Verwijs [22]. Teamwork is only one component of the overall picture of team effectiveness [22]. Therefore, Verwijs concluded that seven factors contribute to team effectiveness in Scrum. These factors are continuous improvement, stakeholder concerns, team autonomy, responsiveness, management support, team morale, and stakeholder satisfaction. In the paper, these concepts have been measured using Likert scales. A Likert scale gives quantitative value to qualitative data [9]. Therefore, the study does not address objective measures to measure team effectiveness, which can be implied as a research gap. Delgado et al. [3] explain why most team effectiveness studies contain subjective measures instead of objective measures. He states that in most studies, subjective measures are used to measure performance effectiveness and behavioral outcomes, as data are often unavailable for objective measurement. As a result, it is difficult to make comparisons of the different characteristics of the team [3].

3 Research Design

In this study, the design science research methodology created by Peffers et al. [15] will be followed. Different phases are shown in Fig. 1. Each phase consists of activities.

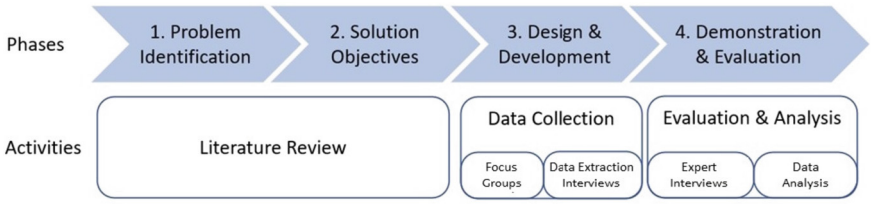

Fig. 1. Design Science Process Model adapted from Peffers et al. [15]

3.1 Literature Review Protocol

Before the different stages will be discussed, a literature review protocol will be elaborated to guide the literature review. This protocol was derived from Kitchenham et al. [10] and was tailored for this research. The first step of this protocol was to apply search terms in different search engines. These terms have been derived from the research question and the defined problem statement. 'Team Effectiveness Scrum' and 'Productivity Scrum' are the main search terms derived from the research question and problem statement. This step led to a selection of papers in which inclusion and exclusion were applied. The exclusion criteria contained three elements. Studies written in another language than English were excluded. Textbooks and papers that include student experiments were also excluded. Papers were excluded if they have been published at conferences that are grouped into categories less than C based on the core conference ranking. After the exclusion criteria, the inclusion criteria were applied. These contained also of three steps. First, the titles were being screened. The papers were selected if the title contained 'team effectiveness' and 'Scrum' or 'productivity' and 'Scrum'. Second, the abstracts of the papers that were selected after the first inclusion criteria step were analyzed. As a third step, the selected papers from the second inclusion criteria step were thoroughly read. Only papers were included that describe/discuss at least one of the following elements: Team effectiveness in Scrum, Productivity in Scrum, Method to calculate team effectiveness productivity, or team effectiveness productivity metrics in Scrum teams.

Ultimately 24 papers were derived from the protocol that can been considered suitable for this paper.

3.2 Problem Identification and Solution Objectives

The main part of the Problem Identification & Solution Objectives is to identify the problem and research gap and formulate objectives for the final created artifact. The problem identification of this research is described in Sect. 1, and the Solution Objectives phase has been elaborated in Sect. 2.

3.3 Design and Development

The Design & Development phase consists of the data collection process, which is elaborated in Sect. 4. There were several activities involved in the data collection process. First, two focus groups were organized to generate objective measures related to team

effectiveness and Scrum. In the first focus group, five Scrum experts were asked to generate measures with objective measures related to Scrum and team effectiveness, taking into account the definition of team effectiveness by Fransen et al. [5]. *"the quality of team performance, as well as the perceived satisfaction with individual needs of team members"*.

In the second focus group, the aim was to link the objective measures generated in the first focus group with the seven concepts of team effectiveness, defined by Verwijs et al. [22]. Verwijs proved that these seven concepts form the basis of Scrum team effectiveness. Linking them to a certain team effectiveness concept indicated that the measures can be applied in measuring Scrum team effectiveness. Furthermore, the second focus group was also applied to review the measures that were generated from the first focus group. The last phase of data collection was to indicate whether the measures derived from the focus groups could be measured in a practical setting. For this, work management system interviews were conducted with Scrum masters to indicate whether a measure can be quantified in work management systems, such as Jira or Azure DevOps.

3.4 Demonstration and Evaluation

During focus groups in the previous phase, measures were collected and linked to the seven concepts of team effectiveness. Expert interviews were conducted to define the exact influence of a certain measure on team effectiveness, since this has not been specified yet. As a result, information was collected on how a certain measure influences team effectiveness, which is described in Sect. 5. In addition, these experts had at least five years of experience in Scrum projects. The structured interview method was used during these interviews. This method involves scheduling questions in which the researcher will ask each respondent the same questions in a similar way [17]. For each expert interview, the following question was asked, taking into account the definition of team effectiveness by Fransen et al. [5]:

How does this measure influence team effectiveness?

The last step of this phase is an analysis of the interviews. Each measure contains several opinions on whether the measure affects team effectiveness. The purpose of the analysis was to obtain information on whether the four opinions were on the same line. Therefore, a coding scheme was applied in which a specific color was applied to each measure. The end result was an overview that includes the expert's perspective on each measure. Additionally, a color-coded analysis was also performed to indicate whether experts were on the same line about the influence of each measure on team effectiveness.

4 Generating Objective Measures

In total, two focus groups have been conducted to generate objective measures. In addition, six interviews have been conducted to indicate whether the measures generated from the focus groups, can be measured in a work management system.

4.1 Focus Group 1

The first focus group consisted of five participants, in which each participant had at least five years of experience in Scrum projects. In total, 54 measures were derived from the first focus group session. After removing the duplicate measures, 54 measures have been reduced to 39 measures. In addition to removing duplicates, the measures needed to be divided into objective and subjective measures. Although participants were asked to mention the objective measures down, the evaluation showed that there were also subjective measures involved. As a result, a check was needed to remove the subjective measures. This process was carried out by the researcher and validated by the focus group participants. In this process, the 39 measures have been reduced to 30 measures. Ultimately, 30 objective measures were collected after the first focus group.

4.2 Focus Group 2

The second focus group consisted of six participants. Similarly to the first focus group, the objective was to gather participants who have more than five years of Scrum experience. However, one participant had less than five years of Scrum experience. Participants in the second focus group were asked to relate 30 objective measures, from the first focus group, to the seven concepts of team effectiveness. As mentioned in Sect. 2.3, the seven concepts were continuous improvement, stakeholder concern, responsiveness, management support, team autonomy, team morale, and stakeholder satisfaction. In addition, participants could add objective measures to the concepts, as most of the participants did not participate in the first focus group. In total, 10 measures were added, which resulted that the number of measures grew from 30 to 40 unique measures. However, there was noted that measures were applicable to multiple concepts. In total, 10 of the 40 measures were applied to two concepts. As a result, 50 measures, of which 10 duplicate measures, were distributed across the seven concepts. Table 1 provides an overview of the Team Autonomy team effectiveness concept and the measures related to the concept. The first column shows the team effectiveness concept that is related to the measures, and the second and third columns show the measure and the definition of the measure. The fourth column indicates whether the measures were derived from the first focus group, or were added in the second focus group. An overview of all team effectiveness concepts and their associated measures will be provided in Table 3 in Sect. 5.2.

4.3 Data Extraction Work Management Systems

The last part of the results section dives deeper into the objective measures that are currently measured in work management systems such as Jira or DevOps. In total, six interviews were conducted with four Scrum masters, a software engineer, and a delivery manager. The purpose of the interviews was to review the measures and determine whether they can be measured in work management systems. During interviews, it became evident that it was not always that straightforward whether a measure can be quantified in a work management system. As a result, five categories have been created, in which each category contains a color.

Table 1. Objective measures linked to the Team Autonomy team effectiveness concept.

Team Effectiveness Concept	Objective measure	Definition	Derived from 1st focus group session
Team Autonomy	The amount of technical debt in a sprint/release	The number of trade-offs during a sprint. Technical debt is the consequence that software projects face when they make trade-offs to implement lower quality, less complete solutions to meet budget and schedule constraints imposed by business realities	Yes
	The number of scrum teams working together on the same product	The number of teams working together on the same product	Yes
	The number of reviews/acceptance tests executed by external parties	The number of reviews and acceptances given by external people outside the scrum team	No
	The number of software releases	The number of releases of a scrum team within a certain period/sprint	Yes
	The number of user stories/items executed by a minimum of 2 scrum team members	The number of tasks that are executed by at least 2 scrum team members	No

- The measure can be quantified in a work management system. (GREEN)
- The measure can be quantified, but not in Jira or Azure DevOps. (BLUE)
- The measure cannot be directly derived from the work management system. However, data points are available in the system. (YELLOW)
- The measure can be counted manually and put into the work management system. (ORANGE)
- The measure can neither be quantified nor visualized in a work management system. (RED)

The first three categories contain measures that can already be computed in work management systems or the data available to compute the measure. The last two categories are measures in which a large adjustment has to be made to the system to compute the measure, or it is not possible to compute the measure. The number of measures and the percentage of the total number of measures, which is 40, attached to a certain category are shown in Table 2. The first column indicates the category. Columns 2 to 7 show the number of measures and the percentage related to the category, according to the interviewee. Column 8 provides insight into the average percentage per category. Table 2 shows that a variety of responses have been provided to determine whether a measure can be quantified. Since the knowledge of the work management system differs

Table 2. The number of measures related to a category.

Category	Interview						Avg. % Category
	1	2	3	4	5	6	
The measure can be measured in a work management system	12 (30%)	10 (25%)	19 (47.5%)	20 (50%)	10 (25%)	8 (20%)	32.9%
The measure can be computed, but not in Jira or Azure DevOps	7 (17.5%)	1 (2.5%)	1 (2.5%)	1 (2.5%)	1 (2.5%)	2 (5%)	5.4%
The measure can not be directly derived from the work management system. However, the data points for indicating the measure are available in the system	11 (27.5%)	6 (15%)	2 (5%)	5 (12.5%)	2 (5%)	13 (32.5%)	16.3%
The measure can be counted manually and put into the work management system	6 (15%)	4 (10%)	5 (12.5%)	5 (12.5%)	3 (7.5%)	3 (7.5%)	10.8%
The measure can be neither measured nor visualized in a work management system	4 (10%)	19 (47.5%)	13 (32.5%)	9 (22.5%)	24 (60%)	14 (35%)	34.6%

among the interviewees, the most optimistic scenario has been chosen. This means that whenever an expert mentions that the measure can be quantified, it is assumed that the measure can be quantified. Table 3, one page 11 provides a more detailed overview of whether a measure can be quantified taking into account the most optimistic scenario mentioned above.

Table 3. Overview of each measure and color codes from the evaluation and data extraction research phases.

Measure	Color Code Evaluation	Color Code WMS	Contributes to TE?	TE Concept
The number of retrospective items solved after a new sprint			Yes	CI
The number of bottlenecks visualized by a value stream map			Yes	CI
Measuring software quality with SonarQube			Yes	CI
The number of bugs/defects within a sprint			Yes	CI
Test time			Yes	CI/R
Built time			Yes	CI/R
Release time			Yes	CI/R
The number of changes to the product/sprint backlog after a sprint review meeting			Yes	CI/SC
Response time stakeholders to requests			Yes	SC/R
Business value			Yes	SC/SS
The number of stakeholders attending sprint meetings			No	SC/TM
The number of acceptance tests 'first time right'			No	SC
The number of times the same feedback is addressed by stakeholders			Yes	SC
The difference between the items/stories that are created in this sprint compared to the previous sprints			No	SC
The time it takes to execute an integration			Yes	R
Done work			Yes	R
Review time			Yes	R
Lead time release/story			Yes	R
The number of software releases			Yes	R/TA
Cycle time			Yes	R
The ratio between the working hours and meeting hours			No	R
The number of managers attending a sprint review meeting			No	MS
Response time of management to requests			No	MS
Availability and recognizability of management			No	MS
Resources (euros)			No	MS
The amount of technical debt during a sprint/release			Yes	TA
The number of scrum teams working on the same product			Yes	TA
The number of reviews/acceptance tests executed by external parties			No	TA
The number of user stories/items executed by a minimum of 2 scrum team members			Yes	TA
User story age			Yes	R/TM
The number of backlog items			Yes	TM
The number of times the sprint goal has been achieved			Yes	TM/SS
The number of scrum team formation changes			Yes	TM
The number of releases to production without bugs			No	TM
The number of team events at least two members are present			No	TM
Average velocity previous X sprints			Yes	SS
The number of uncommitted features delivered within a release/sprint			Yes	SS
The finished user stories compared to the predicted number of user stories			Yes	TM/SS
Downtime			Yes	SS
The lead time of a feature compared to the expected delivery time of a feature			Yes	SS

5 Evaluation of the Team Effectiveness Measures

In this section, the evaluation of the findings will be discussed. Additionally, measures that contribute to measuring team effectiveness will be presented.

5.1 Expert Interviews

In total, four expert interviews were conducted to evaluate whether a measure provides information on team effectiveness. In addition to the answers of the experts, a link has been made to scientific literature to find out if the influence of a certain measure on team effectiveness has already been investigated. However, not all measures could be linked to scientific literature.

The analysis of the opinions on each measure showed that there is still much discussion among experts on whether a measure influences team effectiveness. Therefore, categories have been assembled to distinguish measures.

- The color *GREEN* has been used if all experts agree that the measure has an effect on team effectiveness.
- If the color is *BLUE*, this means that an expert disagreed and three experts agreed that the measurement had an impact on team effectiveness.
- If two experts agree, *YELLOW* is used if the measure influences team effectiveness.
- The color *ORANGE* was applied if three experts disagreed and one expert agreed on whether a measure influences team effectiveness.
- *RED* has been used if all experts state that the measure does not influence team effectiveness.

In general, it can be concluded that measures in the first and second categories, the colors Green & Blue, strongly influence team effectiveness. It could be argued that there is too much debate on whether the measure impacts team effectiveness for the measures that contain, the Yellow, Orange & Red colors. Due to this discussion, it has been assumed that there is no direct relationship between the measure and team effectiveness.

Ultimately, 35 of the 40 measures were assigned to the first and second categories and thus influence or strongly influence team effectiveness.

5.2 Relating the Evaluation Phase to the Work Management Data Extraction Phase

The aim of this section is to indicate whether a measure provides information on team effectiveness and can be measured in a work management system. The previous section provided information on what measures can be related to measuring team effectiveness. In this section, a link will be made between the evaluation of the measures and the extraction of data from the work management systems.

The colors Green & Blue in Sect. 5.1 state that the measure influences or strongly influences team effectiveness. For Sect. 4.3, Data extraction Work Management Systems, the color Green, Blue & Yellow indicate that a measure is measurable in a work management system. Whenever a measure is attached to both categories mentioned above, the measure provides information on team effectiveness, and the measure can be quantified in a work management system. In other words, the objective measure contributes to measuring team effectiveness in scrum. In the previous section, there was described that 35 measures out of the 40 measures influence or strongly influence team effectiveness based on the Green and 'Blue colours. These 35 measures will be taken to the expert interviews column. This column shows that out of the 35 measures, 29 measures also have the Green, Blue or Yellow colour. This means that, according to the expert interviews, and whether a measure can be quantified in a work management system, 29 measures contribute to measuring Team Effectiveness. These measures contain a Yes, in the Contributes to TE? column. A complete overview of whether a measure contributes to measuring scrum team effectiveness is shown in Table 3, which provides the name of the measure, the evaluation color, the color to indicate if the measure can be quantified in a Work Management System[1], whether the measure contributes to measuring Team Effectiveness[2] and the concept(s), Continuous Improvement[3], Stakeholder Concern[4], Responsiveness [5], Team Autonomy[6], Management Support[7], Team Morale[8], Stakeholder Satisfaction[9], of team effectiveness to which the measure is linked, based on the second focus group.

6 Discussion

In this section, the 29 measures that contribute to team effectiveness will be examined at the concept level, which is described in Sect. 2.3. Furthermore, observations during the second focus group will further discuss the generated measures.

[1] WMS
[2] TE
[3] CE
[4] SC
[5] R
[6] TA
[7] MS
[8] TM
[9] SS

The second focus group showed that 40 unique measures were related to the seven concepts of team effectiveness. The evaluation phase showed that 29 of the 40 measures ultimately contribute to measuring team effectiveness in scrum. This shows that these measures could provide a broad overview of measuring team effectiveness as a whole. However, further analysis shows that not all concept can be fully measured. Verwijs [22] states that there are seven concepts important in scrum team effectiveness. To ensure that all concepts are covered in measuring team effectiveness, objective measures are linked to the seven concepts of team effectiveness.

An analysis was performed to identify differences in the measurability of concepts. In other words, whether a concept can be quantified. The analysis shows that two concepts, stakeholder concern and management support, are the most represented. Three of the seven measures (42.9%) related to stakeholder concerns do not contribute to measuring team effectiveness. Four of the four measures (100%) related to management support do not help measure scrum team effectiveness. For team morale, team autonomy, and responsiveness these percentages are 25%, 20%, and 8.3% respectively. This indicates that there is still a great difference in the measurability at the concept level. Verwijs [22] concluded that all these concepts influence team effectiveness. It is important to note that not all seven concepts can be fully measured on the basis of objective measures. Therefore, this should be taken into account when measuring team effectiveness based on these measures. Table 4 provides an overview of the measurability of each concept.

Table 4. Measurability overview concepts

Team Effectiveness Concept Measurability	(%)
Continuous Improvement	100%
Stakeholder Satisfaction	100%
Responsiveness	91.7%
Team Autonomy	80%
Team Morale	75%
Stakeholder Concern	57.1%
Management Support	0%

In addition to the difference between concepts, another interesting note on measures can be derived from discussions in the second focus group.

The discussion dealt with the idea that numbers alone do not mean anything. In other words, if a measure provides a number, what does this number mean? Several studies discuss the importance of providing meaning to a number [21, 24]. First, these studies concluded that the meaning of vague quantifiers and numerical values can vary greatly. Also, the problem with people is that each individual has his or her own internal scale to make judgments. As a result, numbers can be interpreted differently and can create confusion. During the focus group, a solution was already suggested. According to a focus group participant, *"to determine whether a given number is high or low, a comparison should be made to, for instance, a predefined goal or a certain trend"*, in

other studies, benchmarking is defined as 'enabling and motivating one to determine how well one's current practices compare to other practices [18, p 786]'. By applying benchmarking, a number of a measure can be understood and helps to understand what the number means for certain standards or for a trend [16]. To apply this matter to this study. As the authors point out, the 29 metrics identified may be related to team effectiveness, but as they are just numbers they are not of use. In order to assess team effectiveness and make comparisons, these measures will have to be converted into indicators, with thresholds that determine when the team has a low, medium or high level, for example.

In general, Sect. 5.2 shows that 29 of the 40 measures help measure team effectiveness. Although this is the majority of the measures generated from the focus group, it does not mean that mapping these measures provides a complete picture of measuring team effectiveness. There is still a great difference in the ability to measure the seven concepts of team effectiveness. Furthermore, the measures have to be seen in series or in a trend, or a certain benchmark has to be applied to provide context to a number of a measure.

6.1 Validity Threats

In this section, threats to validity will be discussed. This concerns internal, conclusion, and external validity threats [23].

The internal validity threat is related to the expertise of the participants and interviewees in this research. However, for all interviews and focus groups, the participants had multiple years of Scrum project experience. However, different groups formations were used during both focus groups. As a result, most participants of the first focus group did not participate in the second focus group. This harms the internal validity.

The conclusion validity threat relates to the sample size of the focus group and the number of expert interviews. This can be seen as rather small.

The external validity threat relates to the generalizability of this research. This research took place at one organization. Although this threatens generalizability, all participants were consultants who also worked with or at other companies. As a result, experiences derived from other organizations were also indirectly taken into account. Therefore, this limits the external validity threat.

7 Conclusions and Future Research

In this research, results were presented on whether it is possible to measure team effectiveness based on objective measures. Focus groups and interviews were conducted to answer research questions:

"To which extent can team effectiveness in Scrum be measured based on objective measure?"

We identified 29 measures that contribute to measuring team effectiveness in Scrum. This signals that team effectiveness can indeed be measured to a large extent on the basis of objective measures. However, a few notes have to be taken into account. First, a number on its own of a certain measure has no meaning. Therefore, benchmarking or a trend in numbers has to be applied to provide meaning or context to a number of

a measure. Furthermore, there is a variety in the measurability of team effectiveness concepts. For example, for one concept, all linked measures do not provide meaning to team effectiveness or cannot be measured in a work management system; for another concept, all linked measures are related and can be measured in a work management system. Although not all concepts can be fully measured, five of the seven concepts can be measured for at least 75%, providing a solid basis for measuring team effectiveness. Last, this research can be considered an exploratory study on the topic of measuring team effectiveness based on objective measures. The outcomes provide a first insight into this topic, which can be built on.

Future works should focus on extending this research to other software companies that apply scrum principles. This can result in new insights and new measures that can be applied to measure team effectiveness. Furthermore, it would be interesting to apply the measures in practice. As a result, feedback can be collected for new measures or current measures can be reexamined.

References

1. Benders, J., Van Hootegem, G.: Teams and their context: moving the team discussion beyond existing dichotomies. J. Manage. Stud. **36**(5), 609–628 (1999)
2. Buljac-Samardzic, M., Doekhie, K.D., van Wijngaarden, J.D.: Interventions to improve team effectiveness within health care: a systematic review of the past decade. Hum. Resour. Health **18**(1), 1–42 (2020)
3. Delgado Pinã, M.I., María Romero Martínez, A., Gómez Martínez, L.: Teams in organizations: a review on team effectiveness. Team Perform. Manage. Int. J. **14**(1/2), 7–21 (2008)
4. Doolen, T.L., Hacker, M.E., Van Aken, E.: Managing organizational context for engineering team effectiveness. Team Perform. Manage. Int. J. **12**(5/6), 138–154 (2006)
5. Fransen, J., Kirschner, P.A., Erkens, G.: Mediating team effectiveness in the context of collaborative learning: the importance of team and task awareness. Comput. Hum. Behav. **27**(3), 1103–1113 (2011)
6. Hackman, J.R., Morris, C.G.: Group tasks, group interaction process, and group performance effectiveness: a review and proposed integration. Adv. Exp. Soc. Psychol. **8**, 45–99 (1975)
7. Imbrie, P., Maller, S., Immekus, J.: Assessing team effectiveness. In: 2005 Annual Conference, pp. 10–229 (2005)
8. Jahedi, S., Méndez, F.: On the advantages and disadvantages of subjective measures. J. Econ. Behav. Org. **98**, 97–114 (2014)
9. Joshi, A., Kale, S., Chandel, S., Pal, D.K.: Likert scale: explored and explained. Br. J. Appl. Sci. Technol. **7**(4), 396 (2015)
10. Kitchenham, B.: Procedures for performing systematic reviews **33**, 1–26 (2004)
11. Lemieux-Charles, L., McGuire, W.L.: What do we know about health care team effectiveness? A review of the literature. Med. Care Res. Rev. **63**(3), 263–300 (2006)
12. Mathieu, J.E., Gallagher, P.T., Domingo, M.A., Klock, E.A.: Embracing complexity: reviewing the past decade of team effectiveness research. Annu. Rev. Organ. Psych. Organ. Behav. **6**, 17–46 (2019)
13. Moe, N.B., Dingsøyr, T.: Scrum and team effectiveness: theory and practice. In: Abrahamsson, P., Baskerville, R., Conboy, K., Fitzgerald, B., Morgan, L., Wang, X. (eds.) XP 2008. LNBIP, vol. 9, pp. 11–20. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-68255-4_2
14. One, V.: 14th annual state of agile development survey. Version One, INC. (2020). https://stateofagile.com/#ufh-c-7027494-state-of-agile

15. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A design science research methodology for information systems research. J. Manag. Inf. Syst. **24**(3), 45–77 (2007)
16. Purwanto, A.: Benefit of benchmarking methods in several industries: a systematic literature review. Syst. Rev. Pharm. **11**(8), 508–518 (2020)
17. Rashidi, M.N., Begum, R.A., Mokhtar, M., Pereira, J.: The conduct of structured interviews as research implementation method. J. Adv. Res. Des. **1**(1), 28–34 (2014)
18. Raymond, J.: Benchmarking in public procurement. Benchmarking Int. J. **15**(6), 782–793 (2008)
19. Salas, E., Sims, D.E., Burke, C.S.: Is there a "big five" in teamwork? Small Group Res. **36**(5), 555–599 (2005)
20. Schwaber, K., Sutherland, J.: The Scrum Guide, vol. 21, p. 1 (2011)
21. Scissons, E.H.: All numbers are not created equal: measurement issues in assessing board governance. Corp. Gov. Int. J. Bus. Soc. **2**(2), 20–26 (2002)
22. Verwijs, C., Russo, D.: A theory of scrum team effectiveness. ACM Trans. Softw. Eng. Methodol. **32**, 1–51 (2023)
23. Wohlin, C., Runeson, P., Höst, M., Ohlsson, M., Regnell, B., Wesslén, A.: Experimentation in Software Engineering-An Introduction. Kluwer Academic Publishers, Doedrecht (2000)
24. Wright, D.B., Gaskell, G.D., O'Muircheartaigh, C.A.: How much is 'quite a bit'? Mapping between numerical values and vague quantifiers. Appl. Cogn. Psychol. **8**(5), 479–496 (1994)
25. Yang, M.C., Jin, Y.: An examination of team effectiveness in distributed and colocated engineering teams. Int. J. Eng. Educ. **24**(2), 400 (2008)