Utrecht University

Measuring Team Effectiveness in Scrum

Author:
Kars Beek
1445308

1st Supervisor:
dr. Gerard Wagenaar
2nd Supervisor:
dr. Sietse Overbeek
Company Supervisors:
Evert de Rooij
Laura Kester

Universiteit Utrecht

Faculty of Science
Graduate School of Natural Sciences
Business Informatics

April 24, 2023

Abstract

Teams have become the building blocks of organizations, leading to an exponential increase in team studies, including team effectiveness studies in the scrum research area. However, no research has been done on the topic of measuring team effectiveness in scrum based on objective measures. The purpose of this study is to determine if team effectiveness in scrum can be measured using quantitative data values (objective measures). Through expert interviews, focus groups, and literature, a list of 30 objective measures was generated to measure scrum team effectiveness. This study contributes to existing research by offering a first insight into an additional method for measuring scrum team effectiveness.

1 Introduction

This section provides an introduction to the topic of scrum and team effectiveness in scrum. In addition, the motivation for this study will be examined. The motivation will be supported by the research aim and the research questions.

1.1 The idea behind scrum

Scrum is one of the most popular methods in software development [48]. The scrum methodology consists of a lightweight framework that helps people, teams, and organizations generate value through adaptive solutions to complex problems, most commonly used in software development [60]. The principles of scrum are formulated in the scrum guide, which is defined by Schwaber and Sutherland [60]. A scrum team is the fundamental part of the scrum method and consists of a product owner, a scrum master, and a number of developers. The product owner is responsible for maximizing the value of the final product that the scrum team produces. Establishing scrum according to the scrum guide's definition is the responsibility of the scrum master. Scrum masters want to achieve this by helping everyone on the scrum team and the organization understand the theory and practice of scrum. Developers are committed to creating any aspect of a usable feature of the final product. Teams that apply scrum operate in a cross-functional manner, which means that members have all the necessary skills to create value in their software development process.
Furthermore, scrum is made up of various events. The sprint is the first event and can be seen as the heart of scrum. A sprint consists of a fixed time period, and it consists of all the work necessary to achieve the product goal, including various other scrum events. Sprint planning is the second event. The work that will be performed during the sprint is organized through sprint planning. The third event, called the daily scrum, is intended to check on daily progress made toward the sprint goal. The sprint review, which assesses the results of a sprint and decides on future adaptations, is the fourth event. The fifth event is the sprint retrospective. The purpose of the sprint retrospective is to plan ways to increase quality and effectiveness.
In addition to events, scrum also includes several artifacts. The first artifact is the product backlog. The product backlog is an ordered, emergent list of what is needed to build or improve the product. The sprint backlog, the set of product backlog items selected for the sprint, is the second artifact. The third artifact is the increment. An increment is the sum of all items in the product backlog completed during a sprint and the value of the increments of all previous sprints. The last artifact is the definition of done. The definition of done is a formal description of the state of the increment when it meets the quality measures required for the product.

1.2 Team Effectiveness in Scrum

Teams have become the basic building blocks of current organizational designs [42]. Well-formed teams enable organizations the flexibility to compose and reconfigure their team memberships to align members' competencies with task demands [40]. As a result, research on the topic of teams has grown exponentially over the past decade, which has also led to an increase in team effectiveness studies in scrum [40]. These studies range from the effect of the involvement of the product owner in team effectiveness to the effect of scrum retrospectives on team effectiveness [39] [49]. The first studies to address the topic of team effectiveness in scrum were conducted by Moe et al. [45] [46]. These studies focused on various factors that affect teamwork in scrum teams and provide a well understanding of reflecting scrum teams based on teamwork models. However, teamwork is only one component of the whole picture of team effectiveness, according to Russo [65]. Russo conducted a study that surveyed more than 5000 developers and 2000 scrum teams, using Likert scales, and a literature study to understand which team-level factors determine scrum team effectiveness. A Likert scale is a response scale that is used mainly in questionnaires to obtain the preferences of participants or the degree of agreement with a statement or set of statements [2]. The method transforms individuals' subjectivity into objective reality, that is, transforming qualitative data values to quantitative data [33]. Russo concluded that seven factors contribute to scrum team effectiveness. These factors are Continuous improvement, Stakeholder Concern, Team Responsiveness, Management Support, Team Autonomy, Team Morale, and Stakeholder Satisfaction. Figure 1 provides an overview of the factors. Furthermore, the figure provides insight into how these factors are associated with each other. To be more precise, the arrows indicate that there is a positive association between the factors, in which the hypothesis, for instance H6a, is mentioned in the arrow,

The study by Russo [65] provides a solid foundation on team effectiveness in scrum, using likert scales. However, no studies have been conducted on measuring team effectiveness based on quantitative data values. Thus, without using individual subjectivity.

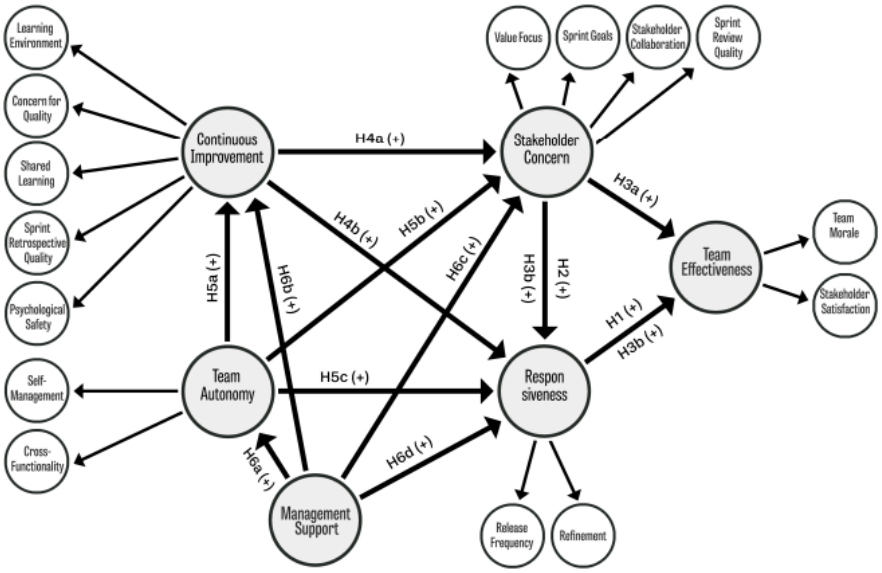

Fig. 1: Team effectiveness framework by Russo et al. [65]

1.3 Research aim and Research Questions

The study by Russo mentioned that scrum team effectiveness consists of seven concepts. These concepts have been measured on Likert scales, which are the subjectivity of individuals [33]. Subjectivity in measures can bring limitations, according to Jaheti et al. [31]. First, subjective measures are difficult to aggregate and interpret because they are often expressed in ordinal scales. Moreover, there has been noticed that these measures are uncorrelated with facts from the field. Also, there are no studies on measuring team effectiveness in scrum based on quantitative data values. Based on the research gap, and the limitations of subjectivity, the aim of the research is to measure team effectiveness supported by quantitative data values. It should be noted that the objective of this thesis is to complement, rather than replace, subjective measures.

Based on the research aim, the following research question has been formulated:

"To which extent can team effectiveness in scrum be measured based on quantitative data values?"

The following sub-questions have been defined to answer the main research question.

- *What is the definition of team effectiveness?*
  Team effectiveness will be the main subject of this study. According to research, there is still much ambiguity about the concept of team effectiveness [52]. Therefore, the concept of "team effectiveness" will be studied and a definition of the concept will be provided for consistency purposes and to define a scope to indicate which aspects are involved in team effectiveness.

- *Which concepts influence team effectiveness in scrum?*
  The main outcome of this research is to create a metric or overview to measure team effectiveness. Before the metric or overview can be created, several measurement concepts should be studied to indicate which concepts can be applied to the quantitative data measures. The seven concepts of team effectiveness by Russo [65] will form the basis of the concepts. In a later stage, focus groups will be held, involving product owners and scrum masters, to discuss the results of the concepts.

- *Which quantitative data values can be used to measure team effectiveness concepts in scrum?*
  After the concepts have been defined, the quantitative data values must be found. Exploring these values is expected to take place in work management systems such as Jira or Azure DevOps, and the use of focus groups and expert interviews. Ultimately, an overview of quantitative data values that can be quantified in different work management systems, and are related to team effectiveness will be provided.

The values of quantitative data have different characteristics, such as it can be measured and quantified and can be seen as objective. In other words, no personal meanings are involved in the data. In this study, the values are derived from work management systems such as Jira or Azure DevOps. The theoretical contribution of this study will provide information on to what extent team effectiveness can be measured based on quantitative data values. Furthermore, the outcome can also be applied as a benchmark for further studies in the area of team effectiveness in scrum. This benchmark is also applicable for practical purposes. In addition to being a benchmark for future studies on team effectiveness in scrum, the results can also be used as a benchmark for scrum teams. The outcome of this study will support scrum teams reflecting their team results and processes, which can help organizations adjust their current scrum teams to improve results. Therefore, it should be noted that the results will provide information only on how teams can be improved and not on how teams can be compared with each other.

2 Research Method

This section describes the research method. First, the chosen research method will be discussed. Afterwards, the different stages of the research will be elaborated, containing the activities performed at each stage.

The purpose of the study was to investigate whether team effectiveness can be measured using quantitative data values. The outcome of this study is an artifact that contains different quantitative values (objective measures) to measure team effectiveness. Therefore, this study can be indicated as a design science study since it aims to design an artifact by conducting research. According to Hevner, "design science creates and evaluates IT artifacts designed to solve identified organizational problems" [29] (p.77). Although the result of this study will not be an IT artifact, Hevner's study [29] can also be generalized to a broader context, as the methodology has also been applied in business process management, in a paper by Sonnenberg et al. [66] and human resource studies, which are discussed by Jennex et al. [32]. The organizational problem, mentioned in the definition of Hevner [29], can be defined as the lack of studies that discuss the question of measuring team effectiveness based on quantitative data values.

Within design science, several process models have been designed to guide design science methods. In this study, the design science process model of Pfeffers et al. [51] has been applied. The design science process model of Pfeffers. has been chosen because it provides a template that is consistent with prior literature. In addition, the template is robust and sufficiently complete to guide research in the area of design science. Lastly, the model has been widely used in Information Science research. Thus, the process model will provide a solid base for this study.

The process model has been divided into five different stages.

1. Problem Identification & Motivation
2. Objectives of a solution
3. Design & Development
4. Demonstration & Evaluation
5. Communication

A high-level overview of the research method is visualized in Figure 2.

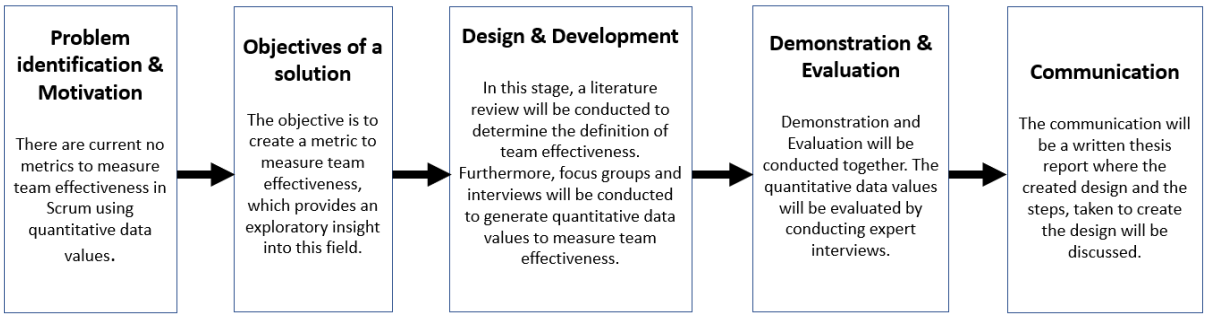

Fig. 2: Design Science Process Model adapted from Hevner [28]

2.1 Problem Identification & Motivation

The purpose of the Problem Identification & Motivation section was to define the specific research problem and justify the value of a solution [51]. Since the research problem has already been discussed in Sections 1.2 and 1.3, no further explanation of the Problem Identification & Motivation will be given.

2.2 Objectives of a solution

In the second stage, a description of how objectives should be rationally inferred from the problem specification will be provided [51]. The final objective of the research was to create a metric that can help measure team effectiveness in scrum based on quantitative data values. As mentioned in Section 1.2, research showed that no studies have been found discussing measuring team effectiveness in scrum, based on quantitative data values. Therefore, it was difficult to predict and state objectives, such as how many data values can be found and used in the final metric. As a result, it was still unknown whether a metric can be created and how comprehensive the metric will be. The goal of this research was to create a metric that provides the first exploratory information in the area of measuring team effectiveness based on quantitative data values.

2.3 Design & Development

Determining the desired functionality of the artifact and its architecture and then creating the actual artifact is the main objective of the Design & Development stage [51]. The Design & Development stage have been divided into two phases, the literature review phase and the research phase. In the literature review, the literature research protocol will be discussed and a brief introduction to the literature study will be provided. In the research phase, two focus groups will be examined and data extraction from work management systems will be discussed. For clarity purposes, an overview of the activities carried out in the Design & Development phase and the Demonstration & Validation phase will be provided in Figure 3.

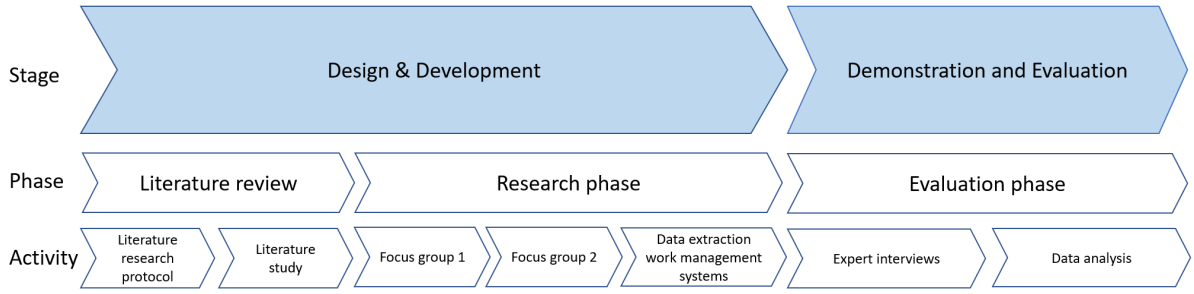

Fig. 3: Design & Development phase and Demonstration & Validation phase activities

Literature review

The first part of the Design & Development phase was a systematic literature review. The systematic literature review has been used for different purposes. It has been applied to understand the breadth and depth of the existing body of work and to identify gaps to explore. Furthermore, summarizing, analyzing, and synthesizing a group of related literature can help to test a specific hypothesis and/or develop new theories.
A systematic review of the literature helps to assess the validity and quality of existing work against a criterion to reveal weaknesses, inconsistencies, and contradictions [73]. For this study, the main objective of the literature review was to gather information about the concept of team effectiveness, which helps answer the subquestion:
What is the definition of Team Effectiveness?
Furthermore, the literature review was used to compile information about the different methods to measure team effectiveness.

To guide the literature review, a protocol has been assembled. This protocol was derived from Kitchenham et al. [35] and was tailored for this research. This method was chosen primarily because it is based on a well-known and widely cited publication in the field of information and computer science.
Different phases are associated with the literature research protocol. These phases are as follows:

1. Identification of research
2. Selection of primary studies
3. Quality Assessment
4. Snowballing Method
5. Data extraction & Synthesis

The first phase aimed to find as many primary studies related to the research question or problem statement as possible using an unbiased search strategy. The primary search area for collecting scientific papers was Google Scholar. In Google Scholar, different search terms have been applied to gather papers. These terms have been derived from the research question and the defined problem statement. 'Team effectiveness scrum' and 'productivity scrum' are the main terms derived from the research question and problem statement. In addition to the core terms, several other terms were applied, for example, to gather information on other measures of Team Effectiveness in other research areas.
The selection of primary studies was the second phase. The main purpose of this phase was to evaluate the relevance of the studies. The evaluation was carried out on the basis of inclusion and exclusion criteria. For this study, the following inclusion and exclusion criteria have been defined, which are based on the inclusion and exclusion applied by Shah et al. [62]. The exclusion criteria contain 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.

The inclusion criteria consisted of three subsequent levels.
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 from the first phase were analyzed. As a third step, the selected papers from the second 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.

In this study, the exclusion criteria were first applied. This was done because the inclusion criteria took more time to apply. This matter prevented papers from being selected based on the inclusion criteria, which is the most time-consuming task, and later removed based on the exclusion criteria.

The third stage was the quality assessment. In addition to the general inclusion-exclusion criteria, quality criteria have been defined. Quality criteria are needed to minimize bias and maximize internal and external validity [15]. Table 1 defines these quality assessment criteria, which are established by Protogerou et al. [53] and customized for this study. The reason for choosing these criteria is that they were evaluated using an expert consensus study that informed the development of the final set of checklist items and descriptions. Furthermore, the criteria have also been pilot tested in a case study. [53] The aim was that these quality criteria should be answered with a "Yes", otherwise the article was removed from the selection.

Quality Criteria Assessment	**Yes/No**
Were specific research questions or hypotheses stated?	
Was the problem or phenomenon under investigation defined, described, and justified?	
Were the data analysis techniques justified?	
Were the measures provided in the report (or in a supplement) in full?	
Is the publication date in the scope of 20 years?	
Was information provided about the context (e.g., place) of data collection?	
Were funding sources or conflicts of interest disclosed?	
Was evidence provided for the validity of all the measures (or instruments) used?	
Is it peer-reviewed?	

Table 1: Quality criteria assessment, derived from Protogerou et al. [53]

The phases, Identification of Research, Selection of primary studies, and Quality assessment of the research protocol have been visualized in Figure 4. The flow diagram consists of six phases. However, one phase has not been discussed, which is the article's retrieved randomly phase. This phase has been taken into account, since the first phase generated more than 20,000 papers. For the researcher, it was not possible to review all papers, solely due to time constraints. Therefore, a random selection of papers will be derived from the first phase. For this research, 200 papers were randomly derived from Google Scholar. To decide whether the number of randomly selected papers was sufficient, a benchmark based on a separate database of scientific articles has been established. There are a variety of scientific databases, but for this study, link.springer and the IEEE database serve as comparison databases. The same keywords from the first stage were applied to the link.springer and IEEE databases. The general rule for the selection of randomly retrieved papers was that the higher the number of generated papers in the link.springer and the IEEE databases from the final selection

of quality papers, the more reliable the number of randomly retrieved papers. Ultimately, the final selection of papers after the literature protocol was 24 of the 200 papers that were randomly retrieved. Of these 24 papers, 16 appeared in the link.springer and IEEE databases, which can be seen as the majority of the number of high-quality papers out of the 200 articles that were retrieved randomly. Therefore, it can be concluded that the 200 articles provide a good overview of high-quality papers for this research.

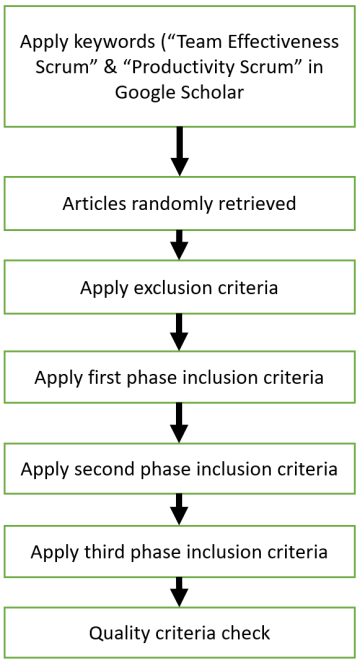

Fig. 4: Flow diagram containing the different phases to select the papers.

After the third stage of the research protocol, the final selection of papers was made. The snowballing method is the fourth phase, and this phase was added primarily due to the random extraction of papers. In the article's randomly retrieved phase, multiple papers had been left out of the literature-gathering process due to time constraints. By adding this phase, papers could be added to the reference list to gather more knowledge about the subject of team effectiveness in scrum.
The snowballing method is a study selection method based on a reference list [69]. This reference list contained the papers that were selected after the third phase of the research protocol, the quality criteria check phase. The backward and forward snowballing approaches are the basics of the snowballing method. While backward snowballing means using the references in the selected papers, forward snowballing refers to identifying new papers based on the citations of the paper. The snowballing method is visualized in Figure 5. Lastly, whenever an interesting paper has been found, the quality criteria, visualized in Table 1, will still be used to determine whether a paper should be added to the list.

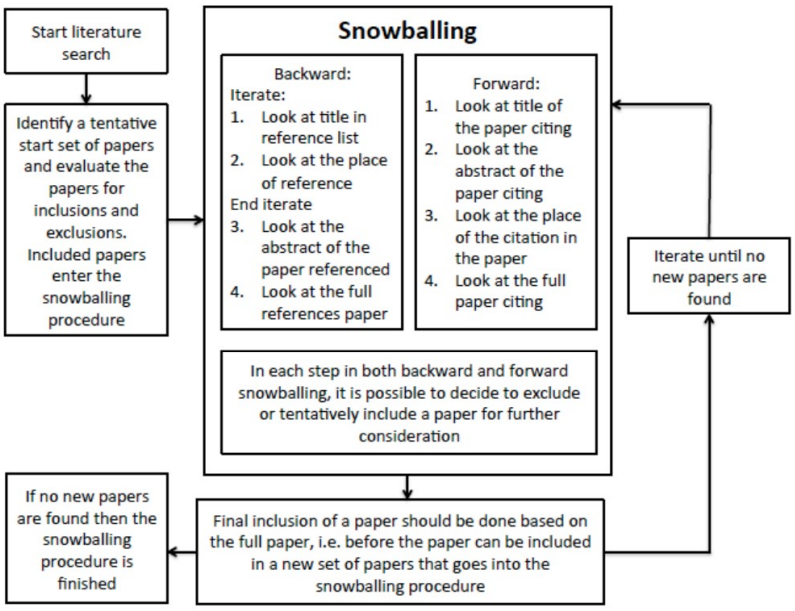

Fig. 5: The snowballing method, which is applied in the literature research protocol by Wohlin et al [69].

The data extraction and synthesis were the fifth phase. The objective of this phase was to take the relevant data from the selected studies. The strategy of this phase was to read all selected studies and summarize the most useful data. Information on team effectiveness in agile way of working, such as quotes, metrics, and percentages that are useful for this study, has been collected in a separate document. These data have been used to define the concept of team effectiveness. As a result, the main outcome of the literature study was to provide an answer to the sub-question "What is the definition of Team Effectiveness?" & "Which concepts should be studied to measure team effectiveness?

Research phase

After the literature review phase has been conducted, the second step in Design & Development is the research phase. The research phase aimed to answer the third sub-question "Which quantitative data values can be used to measure team effectiveness concepts in scrum?" This research question will be answered by conducting two focus groups. Focus groups have been conducted to generate objective measures. In addition, interviews will be conducted to identify the measures that can be quantified in a work management system.

This study aimed to measure team effectiveness based on objective measures. To guide the process of generating objective measures, two focus groups were conducted. Focus groups are defined as a group of individuals to discuss a particular issue or concern [68]. There are several reasons to conduct a focus group. The first reason is that focus groups are a suitable method to generate ideas [22]. Furthermore, it provides an organized setting for a discussion between participants, in which participants ask each other questions and explain themselves to each other [47]. The first focus group session aimed to gather objective measures that may be used in scrum teams or teams that follow the scrum principles, taking into consideration team effectiveness. As a result, by conducting focus groups and bringing more experts together, the prediction is made that more objective measures can be generated. Furthermore, discussions have a positive influence in indicating whether a measure is related to scrum. The second focus group aimed to link the objective measures, derived from the first focus group, with the seven concepts of team effectiveness, extracted from Russo [65]. For both focus groups, the objective was to gather participants who have at least five years of experience in scrum projects.

Focus Groups As mentioned, the first focus group session aimed to gather objective measures that can be used in scrum teams or teams that follow the scrum principles, taking into account team effectiveness. The methodology for conducting the focus group was derived from Simon et al. [63] and altered for this research.

First, the focus group participants were introduced with the study goal and the focus group goal mentioned. After the introduction, the group had been split up, and each member was asked to provide an individual response to the following question
'What can be measured and is applicable to a scrum team or teams using the scrum methodology, taking into account the definition of team effectiveness?' For example, the number of members of the scrum team, sprints, etc.
The following definition of team effectiveness was given to the participants:
"Team effectiveness includes the quality of team performance, as well as the perceived satisfaction of the individual needs of team members" [25].
The participants were asked to write the measures on post-its. After a certain period of time, the focus group was merged into two smaller groups to discuss the measures and encourage conversation. Both smaller groups were instructed to stick all of their post-its to a wall. This was done to create an overview of all post-its containing measures that might be applicable in scrum. Each group was asked to present its results in a brief presentation. The other group was asked to review their results and pose questions. In the end, the goal was to come to an agreement between the groups and to ensure that both groups were satisfied with the results. The focus group resulted in an overview of quantitative data values (objective measures) that can be derived from scrum.

After the first focus group, the measures obtained from the first focus group were analyzed. This study focuses solely on objective measures. Therefore, the researcher made a distinction between objective and subjective measures and was reviewed by the participants in the first focus group. The end result is a list that contains only objective measures that would be used in the second focus group.

The second focus group aimed to link the objective measures, derived from the first focus group, to the seven concepts of team effectiveness, extracted from Russo [65]. Russo proved that seven concepts influence and measure team effectiveness. Linking the measures to the concepts provided a brief idea of whether a measure can provide information on team effectiveness in scrum. Additionally, the second focus group served as an additional review group for the results of the first focus group.

The method by which the focus group was conducted will be explained. Similarly to the first focus group, participants were first briefed on the purpose of the study and the objective of the focus group. The purpose of the focus group was to link the objective measures, derived from the first focus group, with the seven concepts of team effectiveness by Russo [65]. The seven concepts were Continuous Improvement, Stakeholder Concern, Responsiveness, Management Support, Team Autonomy, Team Morale, and Stakeholder Satisfaction. Additionally, an explanation of the concept and a definition of each objective measure were provided to reduce the debate over the meaning of a measure or the meaning of a concept. In addition to the objective measures of the first focus group, par-

ticipants could also add new objective measures to the list. Additionally, the definition of team effectiveness was displayed to ensure that all participants were on the same line. After the briefing, the focus group was split into two groups of three participants. Large sheets of paper containing the name of the concept of team effectiveness were spread throughout the room. Each group was told to evaluate which of the objective measures is related to the concept of team effectiveness. A measure was added to a concept if the team was certain that the measure and the concept were linked. This process was repeated until all seven concepts had been covered by both groups. Both teams were also instructed to reflect the results of the other team when discussing the measures. Whenever a team questioned the outcome of the other team, side notes were added to discuss the measure at a later stage of the focus group. The final phase of the focus group was the discussion. In this phase, both teams were required to review the measures of the other teams. Also, an explanation of the added side notes and why the team felt that it did not fit the concept was given. The opposing team could respond to this critical note. In the end, both teams had to agree on whether the measure fit the concept. Fortunately, a consensus was reached in all discussions, which was approved by both teams. The final result was an overview of the concepts and objective measures that the two teams agreed on.

Data extraction work management systems The result of the second focus group was a list containing objective measures related to the seven concepts of Team Effectiveness. The last step in the research phase is to extract data from the different work management systems. The objective of this phase was to obtain information on the possibility of measuring the objective measures, obtained from the second focus group, in work management systems such as Jira or Azure DevOps. In other words, to indicate whether the measures are or can be applied in a practical setting. The goal was achieved through interviews with members of the scrum team. During the interviews, each measure was examined, and the following questions were asked for each measure.
-Has the measure already been quantified within the work management system of the scrum team?
If the answer to this question was NO, the following follow-up was asked:
Would it be possible to measure the measure in the work management system?
Both questions indicated whether a measure can be quantified in a work management system. It has been assumed that the measure cannot be quantified if both questions are answered negatively, according to the interviewee. The end result is an overview of all measures, providing information on whether it is possible to measure a measure in a work management system, such as Jira or Azure DevOps.

2.4 Demonstration & Evaluation

Observing and measuring how well the artifact supports a solution to the problem is the main objective of the Demonstration & Evaluation phase [51]. In other words, this phase can be interpreted as the validation phase of the created artifact. As indicated in the previous section, the end product of the Design & Development phase is an overview of the objective measures, indicating whether it is possible to measure them in a work management system such as Jira or AzureDevOps. During the focus groups in this phase, measures were collected and linked to the concepts of team effectiveness. However, during this phase, the exact influence of this measure on team effectiveness has not been defined. Therefore, expert interviews were conducted. The objective of the expert interviews was to gather information on how a certain measure influences team effectiveness. Furthermore, these experts should have at least five years of experience in scrum.

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 [56]. For each expert interview, the following question was asked:
How does this measure influence team effectiveness, taking into account the definition of team effectiveness.
The definition used is the definition of Fransen et al. [25]. "Team effectiveness includes the quality of team performance, as well as the perceived satisfaction of individual needs of team members. " This question has been repeated for the 40 measures derived from the second focus group. Ultimately, an overview was created on how each measure is related to team effectiveness, according to experts.
The last step in the Demonstration & Evaluation phase was an analysis of the interviews. Each measure contained four 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 individual measure. The end result will be an overview that includes the expert's perspective on each measure. Furthermore, a color-coded analysis was also performed to indicate whether experts were on the same line on the influence of each measure on team effectiveness.

2.5 Communication

In the last phase of the design science process model, the problem and its importance, the artifact, its utility and novelty, the rigor of its design and its effectiveness to researchers and other relevant audiences, such as practicing professionals, will be communicated [51]. For this study, all steps taken and all the results of the research have been documented in a written thesis. In addition, a presentation will be presented to present the results and provide a further explanation of the research carried out.

3 Literature review

In the literature review, the definition of team effectiveness will be discussed to answer the following sub-question: What is the definition of Team Effectiveness? This section elaborates on the general definitions of team effectiveness and the definitions that are applied in scrum research. Furthermore, the literature review is used to compile information on the different methods to measure team effectiveness in other research areas, Section 3.2, and in scrum, Section 3.3. The last part also helps to answer the second sub-question: Which concepts influence team effectiveness in scrum? The papers for the literature review have been selected based on the literature research protocol, described in Section 2.3. Ultimately, 24 articles met the criteria stated in the literature research protocol. These articles form the basis for the literature review.

3.1 A definition of Team Effectiveness

There is still a significant amount of ambiguity regarding the concept of team effectiveness [52]. This is mainly due to the fact that different organizations have different views on what defines "effectiveness" [6]. Fransen et al. [25] underline that there is much ambiguity around the concept of "Team Effectiveness". Therefore, in this research, a definition of "Team Effectiveness" will be provided for consistency purposes and to define the scope of the thesis.

Hackman et al. [27] have proposed one of the first papers to discuss the topic of team effectiveness. Although Hackman [27] does not define a formal definition, the article states that in addition to performance outcomes, such as speed to solution and 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. Mathieu et al. [41] underline the rationality of Hackman [27] and specify two outcomes for team effectiveness. Tangible products of team interaction and influences on people in a team. Furthermore, Mathieu et al. [41] classify tangible outcomes into three types: efficiency, productivity, and quality. The numerical output of some units indicates the productivity of a team. An example of a numerical output might be sales logged, clients served, or engagements completed. Efficiency is a similar concept, but it is defined in terms of quantitative counts of units produced compared to some standard or benchmark, for example, products relative to raw materials consumed, the time required to reach a decision versus time allocated, and sales relative to quotas. The last type is quality, which represents an assessment of the value or value of outputs such as product rejection rates, decision quality, customer satisfaction, and safety rates. The second general category of team outcomes can be defined in terms of influence on members, which can be defined as collective or individualistic outcomes. The collective level of analysis includes shared experiences, such as cohesion or psychological safety. These outcomes are often experienced in a similar way by all members. On the contrary, individual-level outcomes refer to attitudes, reactions, learning, and behaviors of individuals that may vary not

only between teams, but also within teams. The papers of Hackman [27], and Mathieu [41] describe the results of team effectiveness and provide a solid foundation for team effectiveness; however, team effectiveness has not been defined in these papers. The lack of stated definitions of team effectiveness is a general observation when reading team effectiveness studies. Although many articles have been written on team effectiveness addressing topics such as improving team effectiveness in teams [9] [14], team effectiveness in a virtual environment [54], and broader research on team effectiveness in organizational contexts [52] [26]; no definitions of team effectiveness have been provided. A general observation is that these articles address criteria or factors that influence team effectiveness or provide the desired outcomes of team effectiveness. However, a general definition of Team Effectiveness is often left out.

One of the few definitions of the concept of team effectiveness found has been defined by Fransen et al. [25]. In this paper, team effectiveness has been defined as the quality of team performance, as well as the perceived satisfaction with individual needs of team members. Wu et al. [72] provides a different definition of team effectiveness, defining team effectiveness as the extent to which teams meet the expectations of organizations. In the previous section, Hackman et al. [27], and Mathieu et al. [41] underline the importance of individual aspects such as, team members' satisfaction. This aspect has not been defined by Wu et al, and therefore this definition has not been applied in this paper.
Another definition has been defined by Walters et al. [67] and defines team effectiveness as the sum of satisfaction and perceived performance. Although this definition is similar to the definition by Fransen et al. [25], it is not elaborated further in the paper. As a result, the definition of Fransen et al. has been applied in this paper.
Fransen et al. [25] addresses team effectiveness at the team level (that is, performance) and the individual level (that is, satisfaction of team members). Effectiveness at the team level is indicated through performance quality, which includes both process and product quality. The quality of the product and whether a deadline has been met are frequently discussed when discussing product quality. Process quality refers to efficiency. According to Fransen [25], efficiency is the balance between time and materials invested versus the results achieved as a result of this balance. In addition, performance also refers to the quality of collaboration, which is the effective use of the expertise and capacity of a team, along with smooth processes of negotiation, decision-making, and performance monitoring in the team.
Whereas performance has been divided into three parts, team member satisfaction has not been further elaborated. Therefore, other studies have been consulted to discuss the concept of team member satisfaction. Team member satisfaction has a significant impact on the overall well-being and the performance of a team [44]. In Acuna et al. [1], the concept of satisfaction indicates how much a team member agrees and conforms with his team members about the work method, the atmosphere generated, the achievement of goals, etc. Furthermore,

research shows that a variety of factors influence team member satisfaction such as the presence of communication and cooperation between groups [11], the level of communication and cooperation within the work groups [12], and team leadership behavior [44]. As a result, team member satisfaction is a very broad concept in which many factors influence the outcome of the concept.
In addition to the definition of Fransen et al. [25] which can be understood in a wider context, the definitions of "Team Effectiveness" have also been defined in the Scrum research area. Moe et al. [45] conducted one of the first studies on team effectiveness in scrum. However, in their paper, no definition of scrum team effectiveness is mentioned. A definition of scrum team effectiveness has been given by Russo et al. [65]. In this paper, the effectiveness of the scrum team has been defined as: 'the effectiveness of the scrum team as the satisfaction of the team members with their work process and the satisfaction of the stakeholders with the results of that process' (p.2). Although this definition is in line with the definition of Fransen et al. [25] and corresponds to the Team Effectiveness outcomes defined by Hackman [27], an important difference should be mentioned. In the definition by Fransen et al. [25] and the desired outcomes of Team Effectiveness by Hackman [27], it is mentioned that the quality of team outcomes is an important factor in Team Effectiveness. According to the definition by Russo [65], stakeholder satisfaction indicates the quality outcomes of Team Effectiveness. Although this can be used to measure the tangible outcomes of the team, the question is: Is it valid to measure the outcome of a team solely on stakeholder satisfaction? In addition to quality, efficiency and productivity should also be taken into account as a tangible outcome of team effectiveness, according to Mathieu [41]. Therefore, the definition by Russo [65] can be indicated as too narrow for this research. As a result, the definition of Fransen [25] will be applied in this research, since it can be applied in a broader context. In conclusion, this study's definition of team effectiveness is "Team effectiveness includes the quality of the team's performance, as well as the perceived satisfaction of individual needs of team members" [25] (p.1108).

3.2 Measuring Team Effectiveness

Team effectiveness has been studied in all kinds of research areas. In this section, team effectiveness will be discussed in the research areas of healthcare and engineering. Based on literature research, these two research areas comprised the majority of team effectiveness studies. Following the number of team effectiveness papers, there can be assumed that the papers in these research areas have a respectable level of team effectiveness maturity. As a result, these research areas have functioned as a benchmark for this study. The focus of this section will be on the various methods to measure team effectiveness in these research areas.

Healthcare While researching team effectiveness in other research areas, an observation was that most team effectiveness studies have been conducted in the Healthcare sector. The main reason for this trend is that team effectiveness has become a vital role in healthcare, which can be underlined in the following case. Research shows that 3–4% of patients hospitalized in the United States were harmed by the care received and 44,000 to 98,000 patients died as a result of medical errors. The study concluded that effective teamwork and better communication between caregivers could have prevented half of them [17]. As a result, much research has been done, resulting in a wide variety of papers. Literature studies have been conducted to summarize these papers and discuss the topic of team effectiveness measurement in healthcare. One of the literature reviews of Healthcare team effectiveness has been written by Lemieux et al. [37]. Although this paper is more than 15 years old, it provides the first insight into team effectiveness measurement methods in healthcare considering the 10,224 citations. Furthermore, this paper can be used as a benchmark for follow-up studies. The paper by Lemieux et al. [37] reviewed 22 studies. An observation that has been made relating to team effectiveness measurement is that all studies applied objective measures of patient and/or organizational outcomes. Only 4 of 22 studies also examined subjective outcomes such as staff satisfaction or perceived team effectiveness [37]. The study mentions several objective outcomes that have been applied in healthcare. Objective outcomes include patient outcomes (e.g., functional status, satisfaction), organizational outcomes (e.g., efficiency, costs), staff behavior (e.g., absenteeism, prescribing patterns), and patient behavior (e.g., adherence to medical advice). This shows that team effectiveness measurement in healthcare is different from other research areas since Pina et al. [52] concluded that most team effectiveness studies focused on subjective matters rather than objective matters. A follow-up study of the literature by Buljac-Samardzic et al. [8] shows that this observation has not been changed. 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 [8]. The papers of Lemieux et al. [37] and Buljac et al. [8] do not explain why objective measures are more popular than subjective measures in

team effectiveness studies. To conclude, research on team effectiveness measurement methods in Healthcare mostly involves objective measurement methods, despite research [52] showing that most team effectiveness measurement methods involve subjective measurement methods.

Engineering The second research area that has been analyzed is the engineering research area. Multiple studies have been conducted in the area of team effectiveness in engineering [74] [18] [30]. Although there is no literature review on the topic of team effectiveness in engineering, the three most cited papers, according to Google Scholar, will be used to provide insight into engineering team effectiveness. The paper by Yang et al. [74] studies the difference in team effectiveness between distributed and co-located engineering teams. Ten characteristics of team effectiveness have been used to measure team effectiveness. These characteristics are given below.

1. Goals & Objectives
2. Utilization of Resources
3. Trust & Conflict
4. Leadership
5. Control & Procedures
6. Interpersonal communication
7. Problem-Solving
8. Experimentation
9. Evaluation
10. Cohesion

Ultimately, each team member had to complete a questionnaire to rate their team according to the characteristics mentioned above on a scale of 1 (low) to 7 (high). Unfortunately, no questionnaire is provided to analyze the questions. However, it is based on the method described in this section. The effectiveness measurement found in the team has been examined by applying a quantitative data method using qualitative data values. In the second paper by Doolen et al. [18] 16 engineering has been studied to investigate the role of organizational context on Team Effectiveness. Two surveys have been designed to assess organizational variables, and the second survey was provided to measure team effectiveness. In this context of assessing team effectiveness, the focus will be on the latter survey. The survey conducted included 12 items and the items were evaluated using a six-point likert scale [18]. In addition, three examples of survey questions were provided in the paper. These questions were as follows.

- "This team can be depended on to meet their goals."
- "This team is successful in meeting their objectives."
- "I view this team as successful."

The last study analyzed was written by Imbrie et al. [30]. Although this study has also been conducted in the engineering research area, the main difference

compared to the other studies is that the participants were students instead of company employees. In this study, engineering students were asked to determine if they perceived their team as effective. The questionnaire contained 24 questions divided into four sub-scales; Interdependency, Learning, Potency, and Goal Setting. The responses to the questions were recorded on a Likert scale (1=strongly disagree to 5=strongly agree). Even though only three studies have been analyzed, which makes it difficult to generalize these observations, they give a small insight into team effectiveness Engineering studies. Based on the three most cited engineering team effectiveness studies, it can be concluded that the use of Likert scales, qualitative data values, is an important measurement method for team effectiveness in engineering studies.

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 the majority of healthcare research team measure the effectiveness of the team based on objective measures, most of the effectiveness studies of engineering teams apply subjective measures. Although only two research areas have been discussed, the main objective of this section was to gain more insight into the research areas that have conducted studies on measuring team effectiveness. Lastly, this section shows that it is possible to measure team effectiveness based on objective measures, which have been investigated in the healthcare sector.

3.3 Measuring Team Effectiveness in Scrum

The first study to address the topic of team effectiveness in scrum was introduced by Moe et al. [45]. The main purpose of this study is to discuss the relationship between the general literature on teams and, in particular, team effectiveness in scrum [45]. To evaluate team effectiveness in scrum, the "Big Five" teamwork model of Salas et al. [59] has been applied. Based on the model of Salas et al. [59], the paper concluded that there are several mechanisms of the model that support team effectiveness in scrum, such as adaptability and team orientation. However, the paper also discovered that several mechanisms are not supported. These factors were lack of mutual trust, handling of problems, and long-term planning. 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. [45] form the basis for future research in the area of team effectiveness in scrum, such as the paper by Russo [65]. As stated in Section 1.2 Russo concluded that seven factors contribute to the effectiveness of the scrum team in his study. 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 by likert scales. A likert scale gives quantitative value to qualitative data [33]. Therefore, the study does not address quantitative data values to measure team effectiveness, which can be implied as a research gap. Pina et al. [52] 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, since data are often unavailable for objective measurement. As a result, it is difficult to make comparisons of the different characteristics of the team [52].

Another observation that can be made is that the subject of productivity has not been taken into account as an indicator of team effectiveness. According to Moe et al. [45], there is an important difference between team productivity and team effectiveness, since productivity is based on external factors in some cases. However, team effectiveness and team productivity are also related, as team effectiveness has been considered a perceived factor that influences productivity [20]. Therefore, the method of measuring productivity in the scrum has been considered. Although research shows that Scrum provides a productivity improvement [13], there is still much discussion about how to measure productivity in Scrum. A review of the literature by Shah et al. [62] confirms this statement. In the study, 13 papers have been discovered reporting on productivity measurement in agile software development. According to their study, lines of code were the most widely used metric to measure productivity. See Figure 6 for an overview of productivity measurement methods.

Study	Productivity Metrics	Knowledge Worker Force
J1	Lines of executable code / staff day	Team
J1	Function Points / staff month	Team
J2	Lines of code / person-hour	Team
J3	Lines of code / hours	Team
J4	Average number of unadjusted function points completed per unit of time	Development team of 2 developers
J5	Resolved issues / month	Per developer
C1	Lines of code / person-hours	Team
C2	Lines of code / hour	Team
C3	Lines of code	Team
C4	Lines of code	Team
C5	Lines of code / hour	Development team of 4 developers
C6	Functional size / effort	Team (scrum)
C7	Function points / months	Per developer

Fig. 6: Overview Productivity metrics derived from Shah et al. [62].

However, refactoring is an upcoming trend in software development [58]. According to Fowler [24], refactoring is the process of altering the software system to improve the internal structure, without changing the external behavior of the code. Thus, this method frequently results in fewer lines of code [64], proving that having more lines of code does not necessarily translate into higher productivity [3]. As a result, the study concluded that current productivity measures are not efficient enough to satisfy the requirements to define productivity in agile software development. This provides a research gap to address the question of measuring productivity in team effectiveness based on quantitative data. However, an important remark by Forsgren et al. [23] must be taken into account when researching quantitative data values. The paper states that one productivity metric cannot tell us everything. Salas et al. [59] underlines this statement and describes that many external factors may contribute to the success (or failure) of the team and therefore, in some cases, team performance measures may be inadequate to understand a team. Therefore, this aspect should be taken into account during the design and development phase, discussed in Section 2.3, of the research.
The main outcome of this subsection is that there are still many research opportunities in the team effectiveness and productivity fields of agile, and to be more specific, scrum. Moreover, this section also discusses that it is quite hard to compare teams based on team effectiveness due to team characteristics and external factors.

The general conclusion of Section 3 relates to answering sub-question 1 & 2. The first sub-question aims to define team effectiveness. For this study, the following definition has been applied: "Team effectiveness includes the quality of team performance, as well as the perceived satisfaction of individual needs of team members". For the second sub-question, the purpose was to identify which concepts influence team effectiveness in scrum. Based on the literature review, the following concepts influence team effectiveness: continuous improvement, stakeholder concern, responsiveness, team autonomy, management support, team morale, and stakeholder satisfaction.

4 Results

The results section shows the outcomes of both focus groups executed, elaborated in Section 2.3. Furthermore, the section provides information on the results of interviews with scrum masters and developers, on whether measures can be quantified in a work management system, also discussed in Section 2.3.

4.1 Focus Group 1

The first focus group aimed to gain insight into different measures in scrum teams, taking into account team effectiveness. Additionally, the goal was to gather participants, each of whom had at least five years of experience working in scrum projects. In total, five participants took part in the first focus group. Table 2 lists the job function and years of experience in scrum projects for each focus group participant. Additionally, each participant is an IT consultant with experience in a variety of industries, including mobility, healthcare, and finance.

Function	**Years of scrum project experience**
Scrum Master	6 years
Scrum Master	9.5 years
Software Engineer	6 years
Scrum Master and Scrum Coach	10 years
Scrum Master	12 years

Table 2: Overview participants focus group 1

Each participant was asked to come up with as many objective measures as possible that are related to both scrum and team effectiveness. Taking into account the definition of team effectiveness, defined by Fransen et al. [25]. In total, 54 measures were derived from the first focus group session. After processing all 54 measures, the total number of measures had been reduced to 39. The reason for this reduction are duplicate measures. In addition to removing duplicates, the measures needed to be divided into objective and subjective measures. Although participants were asked to write only objective measures down, a check was needed to evaluate all measures. This process was carried out by the researcher and validated by the focus group participants. Ultimately, 30 objective measures were collected after the first focus group. Table 3 provides a part of the overview of the objectives established. This table includes the name and description of the measure. Figure 7 illustrates a collection of sticky notes that contain measures from the first focus group. Additionally, Appendix A contains an overview of the 30 objective measures.

Objective Measure	Description
Response time of stakeholders	The time it takes for stakeholders to respond to a question from the scrum team
The number of sprint retrospectives with 1 or 2 retrospective items	The number of sprint retrospectives that resulted in 1 or 2 sprint retrospective items
Velocity	Velocity represents the amount of work accomplished in each sprint expressed in story points. Mahnic, V., & Zabkar, N. (2012). Measuring progress of scrum-based software projects. Elektronika ir elektrotechnika, 18(8), 73-76.
Average velocity on the previous number of sprints	The average velocity over a period of time
The finished user stories compared to the predicted number of user stories that need to be fulfilled in the sprint	The number of solved user stories/tasks compared to the number user stories/tasks that have been scheduled
The number of defects/bugs	The number of bugs/defects within a sprint
Scrum team size	The number of team members in a scrum team

Table 3: Part of the objective measures collected in the first focus group.

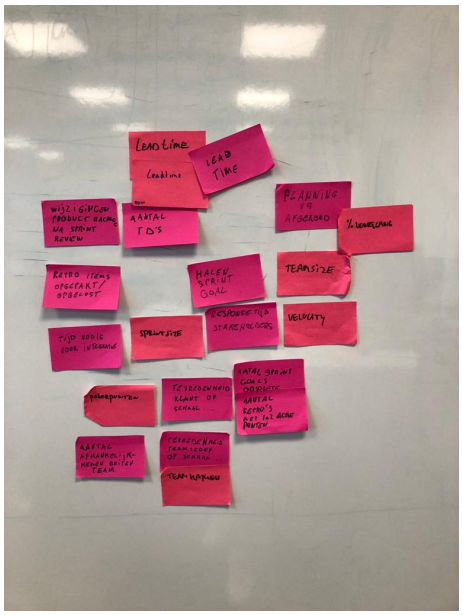

Fig. 7: Picture of the sticky notes that contain part of the measures gathered in the first focus group.

4.2 Focus Group 2

In the second focus group, the objective was to validate the results of the first focus group and adjust the objective measures to the seven concepts of team effectiveness of Russo [65]. The seven concepts were Continuous Improvement, Stakeholder Concern, Responsiveness, Management Support, Team Autonomy, Team Morale, and Stakeholder Satisfaction. As a result, this focus group showed which measures could help measure a particular concept. In total, the 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. Table 4 contains the job function and years of experience in scrum projects of each focus group participant. In addition, the last column indicates whether a participant participated in the first focus group. This column shows that only one participant from the first focus group was also available for the second focus group due to scheduling conflicts. Similar to the first focus group, each participant is an IT consultant with experience in a variety of industries, including mobility, healthcare, and finance.

Function	**Years of scrum project experience**	**Participated in focus group 1**
Scrum Master	10 years	No
Agile project manager	12 years	No
Product owner/Scrum Master	8 Years	No
Product Owner	0.5 Years	No
Scrum Master/Coach	12 Years	No
Scrum Master	6 Years	Yes

Table 4: Overview of participants of focus group 2

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. Additionally, participants could add objective measures to the concepts, as most of the participants did not participate in the first focus group. As a result, 50 objective measures were distributed in the seven concepts, of which 10 measures were used in more than one concept. This means that 40 unique measures were applied. Table 5 provides an overview of the measures related to the concept of Team Autonomy. The table consists of four columns. The first column states to which team effectiveness concept the measures are linked. In this case, Team Autonomy is the key concept of the table. The second column provides the name of the measure. The third column provides a definition of the measure, and the last column specifies whether the measure is derived from the first focus group. In total, 15 new measures were added in the second focus group. This means that 25 measures were derived from the first focus group. Ultimately, an overview of all concepts and the associated measures is included in Appendix B.

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

Table 5: Objective measures linked to the Team Autonomy team effectiveness concept.

4.3 Data extraction work management systems

The last part of the results section dives deeper into the objective measures that are currently quantified in work management systems such as Jira or DevOps. As mentioned in Section 2.3, expert interviews have been conducted to gather the results. In total, six interviews were conducted with four scrum masters, a software engineer, and a delivery manager; see Table 6. Four participants work in a project base environment, that is, building software projects for external clients. The other two interviewees work on a monitoring and maintenance team, maintaining software products for internal and external clients, and occasionally adding software features.

Function	**Years of scrum project experience**
Scrum Master	6 Years
Scrum Master & Software Engineer	10 years
Software Engineer	5 years
Delivery Manager & Solutions Architect	11 Years
Scrum Master & Software Engineer	7 Years
Scrum Master & Software Engineer	9 Years

Table 6: Overview of participants data extraction work management system

The purpose of the interviews was to review the measures and determine whether they can be quantified 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. Therefore, categories were needed to classify the measures. Ultimately, five categories have been created. The five categories are listed below. Each category also has a certain color. These colors were applied for clarity purposes and named after the definition of the measure.

- *The measure can be quantified in a work management system* (GREEN)
  The first category indicates that the measure is already being used by the scrum team or that it is available, but not being used. For example, the number of user stories in a sprint or the product backlog.
- *The measure can be quantified, but not in Jira or Azure DevOps.* (BLUE)
  The second category suggests that the measure can be quantified. However, another system has to be applied to get insight into the measure. For example, SonarQube captures the software quality.
- *The measure can not be directly derived from the work management system. However, data points are available in the system.* (YELLOW)
  The third category group measures that could be indirectly extracted from a work management system. The work management system provides the data points for the measure. However, the measure itself is not visualized in the system. For example, a measure describes the number of days of a

certain activity. The system shows both the start date and the end date of the activity. However, the number of days, which can be calculated by distracting the end date from the start date, is not visible. Therefore, the measure can be indirectly derived from the data.

– *The measure can be counted manually and put into the work management system.* (ORANGE)
Measures that could be manually counted and entered into a work management system are included in the fourth category. These measures cannot be derived from a work management system, but have to be put in manually. For example, the number of stakeholders attending a sprint review meeting.
– *The measure can be neither quantified nor visualized in a work management system.* (RED)
The last category contains measures that cannot be quantified in Jira or Azure DevOps due to the complexity of the measure or due to the limitations of the work management system according to the scrum masters. For instance, the number of acceptance tests that is first time right.

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 had to be made to the system to compute the measure, or it is not possible to compute the measure. The results of the six interviews are shown in Table 7. The first column indicates the category. Columns 2 to 7 show the number of measures and the percentage of the total number of measures, which is 40, attached to a certain category.

	Interview					
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%)
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%)
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%)
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%)
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%)

Table 7: The number of measures related to a category.

Table 7 generates an overview of the number of quantitative data values that can be measured. However, another aspect that is interesting to cover is the number of measures per team effectiveness concept that has been given.

Therefore, an overview of the values that can be measured per team effectiveness concept. Table 8 provides an overview of the measures that can be quantified of the concept of Team Autonomy. In addition, colors have been applied to indicate whether the measure is measurable according to the interviews. In addition, the I with an appropriate number, at the top of the table, represents the number of interviews. An overview of all concepts is included in Appendix C.

		Interview					
Team Effectiveness Concept	Measure	I1	I2	I3	I4	I5	I6
Team Automony	The amount of technical debt in a sprint/release						
	The number of scrum teams working together on the same product						
	The number of reviews/acceptance tests executed by external parties						
	The number of software releases						
	The number of user stories/items executed by a minimum of 2 scrum team members						

Table 8: Measures and the color codes that concern the Team Autonomy concept

5 Discussion & Evaluation

This section consists of four parts. The measures generated in Section 4 will first be evaluated and discussed determining whether a measure is related to team effectiveness. Second, an overview of whether a certain measure contributes to measuring team effectiveness will be provided. In the third part, an analysis will be given on the overview that was generated in the previous section. In the last part, the validity threats of this study will be discussed.

5.1 Evaluation of the measures

In total, four expert interviews were conducted. Table 9 provides an overview that contains the title of the job and years of experience in scrum projects.

Function	**Years of scrum project experience**
Product Owner	10 years
Scrum Master	5 years
Product Owner & Scrum Master	6 Years
Product Owner & Scrum Master	10 Years

Table 9: Overview participants of the expert evaluation phase.

This study aimed to find out whether quantitative data values (objective measures) can be used to measure team effectiveness. In Section 4, an overview of the results of the focus groups and the extraction of data from the work management systems has already been given. The results in Section 4 provide information on the generated measures and whether these measures can be quantified in a work management system. This section will build on these results by providing a detailed evaluation of how these are related to team effectiveness. Therefore, this section will serve as an evaluation of all the measures collected from the focus groups. A measure only contributes to measuring team effectiveness if the measure is related to team effectiveness and if the measure can be quantified. The objective of section 5.2 is to determine the measures that contribute to team effectiveness. After the detailed evaluation of the measures has been completed, a link will be made to Section 4.3, with the aim of finding out if the measures that provide information on team effectiveness are measurable in a work management system.

As mentioned in Section 2.4, four expert interviews were conducted to evaluate whether a measure provides information on Team Effectiveness. Tables 10 & 11 give insight into two measures that have been discussed. Each measure contains four expert opinions. In addition to the answers of the experts, a link has been made to the scientific literature to find out if the influence of a certain measure on team effectiveness has already been investigated. However, scientific

literature could not be found for all measures. Table 11 provides an example of a measure that cannot be linked to scientific literature. An overview of all measures, containing the four expert opinions, can be found in Appendix D.

Measure	Interview 1	Interview 2	Interview 3
Response time stakeholders	This measure can be seen as a bottleneck. Without the input of stakeholders, features could in some cases not be finished unless the stakeholders provide answers or explanations. A slow response time has a negative influence on team performance and team member satisfaction. However, if stakeholders have a fast response time, productivity will increase.	The theory has been built on completing features/user stories in small iterations (a few weeks). Scrum teams are unable to complete their sprints when stakeholders take a long time to respond to demands. This will harm the team's morale and therefore team effectiveness.	If a stakeholder does not respond to requests or the response time is high, the experience is that these stakeholders are avoided in the future, which would negatively influence stakeholder satisfaction. A lack of response time will also have a negative influence on team member satisfaction since team members could not complete their user stories.
Measure	**Interview 4**	**Literature**	
Response time stakeholders	The effectiveness of the team is significantly influenced by response time. If the team does not reply to stakeholders, they are unable to continue producing features, which causes delays. In the end, productivity and team member satisfaction will decrease.	The response time of stakeholders can be seen as a lack of engagement of stakeholders. The lack of stakeholder engagement and support can negatively impact on the team performance in projects. Bahadorestani, A., Naderpajouh, N., & Sadiq, R. (2020). Planning for sustainable stakeholder engagement based on the assessment of conflicting interests in projects. Journal of Cleaner Production, 242, 118402.	

Table 10: Four expert opinions and literature on the "response time stakeholders" measure

Measure	Interview 1	Interview 2
The number of retrospective items solved after a new sprint	This measure functions as a scrum feedback loop and supports team improvement. By solving retrospective items, efficiency, production, and team relationships will all be strengthened. Team Effectiveness helps in increasing both team effectiveness and perceived team member satisfaction. The drawback may be that an excessive number of retrospective items will harm team members' satisfaction.	Retrospective items can be viewed as a proposal for team improvement. Retrospective items that are resolved show that a team is improving. This could result in a more effective process and more satisfied team members.
Measure	**Interview 3**	**Interview 4**
The number of retrospective items solved after a new sprint	It depends on the quality of the retrospective items. Whenever a team is not critical, it does not matter how many retrospective items the team solves. It will not improve team member satisfaction. Team member satisfaction will be stable whenever critically formulated retrospective items are not solved. However, it will negatively influence team member satisfaction whenever critically formulated retrospective items are not solved.	It does not necessarily influence team effectiveness. The amount of solved retrospective items do not necessarily represent how effective a team is, even though it does show that the team is developing. This is related to the next measure, the number of bottlenecks identified in a value stream map.

Table 11: Four expert opinions on "the number of retrospective items solved after a new sprint" measure

The analysis of the opinions on each measure showed that there is still much discussion among experts about whether a measure influences team effectiveness. Therefore, a color code scheme has been applied to indicate what the influence of a certain measure is on team effectiveness.

- The color *GREEN* has been used if all experts agree that the measure has an influence on team effectiveness.
- If the color is BLUE, this means that an expert disagreed, and three experts agreed on whether the measure influences team effectiveness.
- If two experts agree that the measure influences team effectiveness, YELLOW is applied.
- 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 said that measures that contain green and blue colors strongly influence team effectiveness. It could be argued that there is too much debate on whether the measure impacts team effectiveness for the colors yellow, orange, and red. Due to this discussion, it has been decided that these

colors indicate that there is no direct relationship between the measure and team effectiveness.

The results of the evaluation of each measure are shown in Table 12. Table 12 shows that 35 of the 40 measures influence or strongly influence team effectiveness. However, several remarks are made about the opinions provided. These remarks are related to the interpretability, similarity and literature linked to the measures.

Measure	Color Code
The number of retrospective items solved after a new sprint	
The number of bottlenecks in a scrum visualized by a value stream map.	
Software quality (SonarQube)	
The number of bugs/defects within a sprint	
Test time	
Built time	
Release time	
The number of changes to the product backlog after a sprint	
Response time stakeholders to requests	
Business value	
The number of stakeholders attending sprint meetings	
The number of acceptance tests 'first time right'	
The number of times the same feedback is addressed by stakeholders	
The difference between the items stories that are created in this sprint compared to the previous sprints.	
The time it takes to execute an integration	
Done work	
Review time	
Lead time release/story	
The number of software releases	
Cycle time	
The ratio between the working hours and meeting hours	
The number of managers during a review meeting	
Response time of management to requests	
Availability and recognizability of management	
Resources (euros)	
The amount of technical debt during a sprint/release	
The number of scrum teams working together on the same product	
The number of reviews/acceptance tests executed by external parties	
The number of user stories/items executed by a minimum of 2 scrum team members	
User Story age	
The number of backlog items	
The number of times the sprint goal is achieved	
The number of scrum team formation changes	
The number of releases to production without bugs	
The number of team events at least two members are present	
Average velocity previous X sprints	
The number of uncommitted features delivered within a release/sprint	
The finished user stories compared to the predicted number of user stories that need to be fulfilled in the sprint	
Downtime	
The lead time of a feature compared to the expected delivery time of a feature	

Table 12: Color coding of the measures to indicate influence on team effectiveness.

Interpretability of the measures As mentioned, Table 12 shows that most measures influence or strongly influence the effectiveness of the team. However, it should be noted that the measures can be interpreted in many ways. An example of this can be provided by the measure "The number of retrospective items solved after a new sprint", shown in Table 11. An expert mentioned the following about this measure: *"It depends on the quality of the retrospective items. Whenever a team is not critical, it does not matter how many retrospective items the team solves. It will not improve team member satisfaction. Team member satisfaction will be stable whenever critically formulated retrospective items are not solved. However, it will negatively influence team member satisfaction whenever critically formulated retrospective items are not solved"*. This underlines the statement above. In other words, the measure provides information on team effectiveness if certain criteria of this measure are met. As a result, it was difficult to determine whether a measure influences team effectiveness. Therefore, the decision has been made that whenever an expert mentioned that a measure affects team effectiveness, it can be assumed that there is a connection between the measure and the concept of team effectiveness. Even though criteria have to be met first, as shown in the above-mentioned example. Taking into account the expert opinions, measures with certain criteria were observed in 12 of the 160 opinions, given that each of the 40 measures has four opinions. This indicates that in 7.5% of the opinions, this scenario occurred.

Similarities between the measures Furthermore, experts pointed out that the measures had similarities. For example, the number of stakeholders who attended a sprint review meeting and the number of managers who attended a sprint review meeting. Two experts mentioned *"Managers can be seen as similar to the number of stakeholders attending a sprint review meeting measure"* and *"Managers can be seen as a stakeholder"*. A discussion could be held in which one of the above-mentioned measures could be removed since they can be considered similar. However, during the second focus group, participants decided that each measure contributes to the complete picture of measuring team effectiveness. Therefore, it has been decided that if the four experts unanimously agree that two measures are similar to each other, one measure would be removed. However, this was not the case, so no measures have been removed from the list.

Literature on the measures Finally, for 19 measures, related literature could be found. The literature included in the study can be divided into two categories. The first category consists of measures in which the research is directly related to the measure. In this case, the measure has a direct influence on team effectiveness. An example is the literature on the measure of "the response time of stakeholders", shown in Table 10. Research indicates that a lack of stakeholder engagement can negatively impact the performance of a team [5]. The second category consists of measures in which the literature can be found based

on expert responses. This could be explained by the "The finished user stories compared to the predicted number of user stories that need to be fulfilled in the sprint" measure. The four experts unanimously agreed that this measure is an important factor for the predictability of the team. Although no literature on this could be found, which was directly related to the name of the measure, information was available on the impact of predictability on senior management and members of the project team [36]. Both categories are used for data triangulation. In other words, to analyze whether the opinions of experts are in line with the literature. A comparison of the literature and expert opinions showed that no significant differences were observed. Therefore, no follow-up research was carried out.

The general conclusion of this subsection is that, based on the opinion of four experts, 35 of the 40 measures derived from Chapter 4 can influence or highly influence team effectiveness. The subsections on interpretability, similarities, and the literature of measures show that these aspects have little or no impact on how well measures can influence team effectiveness.

5.2 Relating the validation of the results to the data extraction

The aim of this section is to indicate whether a measure provides information on team effectiveness and can be quantified in a work management system. A link will be made between the evaluation of the measures and the extraction of data from the work management systems. Section 5.1 provides information on what measures can be related to measuring team effectiveness. Section 4.3 describes the measures that can be quantified in work management systems such as Jira or Azure DevOps. The outcome of this section was an overview of the measures and whether the measure could be quantified in a work management system.

A situation that could occur is that a measure can be quantified in a work management system but does not provide information on team effectiveness. This situation could also be reversed, in case a measure provides information on team effectiveness, but cannot be used in a work management system. The measure applicable to either scenario does not contribute to the final objective of this study, which is to determine the measures that can be used to measure team effectiveness. Therefore, the results of Sections 4.3 and 5.1 are compared to indicate whether a measure contributes to measuring team effectiveness.

Before making the comparison, an aspect needs to be discussed. In the Data Extraction Work Management Systems section, six scrum masters have been interviewed and the data of the scrum master teams have been analyzed. As a result, a variety of answers have been provided on whether a certain measure can be quantified in a work management system. Therefore, a decision has to be made on how these different opinions are handled. For this study, the decision was made that the most optimistic scenario was used to quantify a measure. This choice has been made, since knowledge about work management systems could differ among scrum masters. This means that the color that provides the most

optimistic result to quantify the measure in a work management system. In other words, whenever five experts mention that the measure cannot be quantified, or the experts state that it is unknown if the measure can be quantified, and one expert states that it is possible, the most optimistic scenario has been selected. In this case, the experience of the single scrum master has been followed, which results in the fact that the matter can be quantified. Table 14 provides the name of the measure, the evaluation color, the color to indicate if the measure can be quantified in a Work Management System (WMS), whether the measure contributes to measuring team effectiveness (TE), and the Team Effectiveness (TE) concept(s) the measure is linked to, based on the second focus group. In the last column, acronyms are used to define the team effectiveness concept. Table 13 contains the team effectiveness concept and the acronym of the concept.

Team Effectiveness Concept	**Acronym**
Continuous Improvement	CI
Stakeholder Concern	SC
Responsiveness	R
Team Autonomy	TA
Management Support	MS
Team Morale	TM
Stakeholder Satisfaction	SS

Table 13: Acronyms Team Effectiveness Concept

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
Software Quality (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
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

Table 14: Overview of each measure and color codes from the evaluation and data extraction research phases.

Each color represents a certain category. The color code evaluation column is explained in section 5.1 and section 4.3 elaborates the color code work management system (WMS) column. In summary, the green and blue colors in the evaluation column state that the measure influences or strongly influences team effectiveness. The yellow, blue, and green colors in the work management system column indicate that a measure is measurable in a work management system. Whenever a measure contains the above-mentioned colors, the measure provides information on team effectiveness, and the measure can be quantified in a work management system. In other words, this objective measure contributes in measuring team effectiveness in scrum. In total, 29 of the 40 measures contribute to measuring scrum team effectiveness. The next section will provide a further analysis of whether measures contribute to measuring team effectiveness.

5.3 Analyzing the evaluation results and further observations

In this section, the results of Section 5.2 will be examined at the team effectiveness concept level. Furthermore, the observations made in the focus groups will further discuss the results of the previous section.

Concept level analysis As described in section 5.2, 29 of the 40 measures 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 there is a difference in measuring a team effectiveness concept. Russo [65] 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. This process was carried out in the second focus group. An overview of the measures related to the seven concepts can be seen in section 4.2.
Table 14 shows the ten measures that did not contribute to measuring team effectiveness. These measures and the team effectiveness concept linked to the measure are shown in Table 15.

Team Effectiveness Concept	**Measure**
Stakeholder concern	The number of stakeholders attending a sprint review meeting
Stakeholder concern	The number of acceptance tests 'first time right'
Stakeholder concern	The difference between the items\stories that are created in this sprint compared to the previous sprints.
Management Support	The number of managers attending a sprint review meeting
Management Support	Response time of management to requests
Management Support	Availability and recognizability of management
Management Support	Resources (euros)
Team Autonomy	The number of reviews/acceptance tests executed by external parties
Team Morale	The number of releases to production without bugs
Team Morale	The number of team events at least two members are present
Responsiveness	The ratio between the working hours and meeting hours

Table 15: Overview of measures that do not contribute to measuring team effectiveness, linked to a team effectiveness concept.

Table 15 shows that the 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. Although Table 14 indicates that most measures benefit from measuring team effectiveness, there is still a great difference at the concept level. Especially the Stakeholder Concern and Management Support concepts, in which the Management Support measure cannot be measured at all. Since Russo [65] 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.

Further observations In addition to the difference between concepts, more interesting notes on measures can be derived from the second focus group. In this focus group, several discussions were held among the participants. Two main topics emerged during these discussions.

The first topic 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 [61] [71]. 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, this is called benchmarking. In Raymond [57] (p786), benchmarking is defined as "enabling and motivating one to determine how well one's current practices compare to other practices". In this study, benchmarking can be specified as internal benchmarking, which is benchmarking against internal operations or standards [10]. 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 [55].

Besides providing meaning to a number, consistency in scrum teams is another topic that has been discussed in the second focus group. According to the focus group participants, scrum teams must aim to minimize variety in both the environment and within the team to increase consistency. Consistency is an important factor for the team to be effective. As a result, consistent team and project variables positively affect the effectiveness and performance of a scrum team. These statements have been researched, and the literature shows that there is a scientific foundation for these statements. Peterson [50] underscores

the importance of consistency and describes that consistency in the direction of the project is required to avoid rework, additional costs, and conflicts. Furthermore, Appelbaum [4] showed that inconsistency in teams led to a lower perceived cohesion of the team, perceived team effectiveness, and psychological safety. Although the topic of consistency had a minor influence on the end results, they gave additional insight into the participant's thinking process. Furthermore, it provides additional information on team effectiveness in scrum teams based on the practical knowledge of the focus group participants.

In general, Sections 5.2 and 5.3 show 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.

5.4 Threats to validity

A critical element of any research study is the analysis and mitigation of threats to the validity of the results [21]. Validity threats are concerned with the question of how conclusions might be wrong, i.e., the relationship between conclusions and reality [43]. For this study, four validity threats of Wohlin et al. [70] are used to ensure rigor in this research. The reason for applying these validity threats is that the threats are well known and highly applied in the fields of computer science and information science, which resulted in a high number of citations. As mentioned above, Wohlin has described four validity threats [70]. The four types are internal, construct, conclusion, and external validity threats.

Internal validity The internal validity focuses on how sure we can be that the treatment actually caused the outcome. In other words, investigate whether there is a risk that the investigated factor is also affected by a third factor [70]. For interviews and focus groups, selection criteria have been assembled to select participants. For example, participants should have at least five years of experience in scrum projects. However, a participant did not meet these criteria, as the participant was selected based on convenience sampling. Convenience sampling is a non-probability sampling method in which units are selected for inclusion in the sample because they are the easiest to access for the researcher [19]. This limits the internal validity. However, all other participants met the selection criteria, which has a positive effect on internal validity. Furthermore, this study used focus groups. A limitation of a focus group is that participants can be influenced in their reasoning by responses from other participants [7]. This caused bias and thus negatively impact internal validity. It would have been better to go through the results of the focus group individually with the participants to indicate whether the participants were satisfied with the results. This resulted primarily from the researcher's experience in leading focus groups.

Construct validity This aspect of validity reflects to what extent the operational measures that are studied represent what the research has in mind and what is investigated according to the research questions. [70]. For example, whether interview questions are interpreted in the same way by all participants. An important part of the study was focus groups and interviews. In these interviews and focus groups, all questions were related to a definition of team effectiveness. To ensure that each participant in the interviews and focus groups was on the same page, the same definition was applied during this research. Additional information was also provided on the definition when there were uncertainties. This limits the space for one's own interpretation regarding the meaning of team effectiveness. Furthermore, data triangulation has been applied to validate the statements made by participants in the evaluation phase and focus groups. In the evaluation phase, statements about whether a certain measure influences team effectiveness have been validated by literature. For focus groups, statements that considered the meaning of a number and which factors are important in team effectiveness have also been validated by expert interviews.

Validation of results helped limit the threat of construct validity. However, since this study is exploratory in the scrum research area, not all statements could be validated by the literature. Lastly, the interpretation by the researcher of the evaluation results is also a construct validity threat. In the evaluation, experts were asked whether a certain measure influences team effectiveness. The explanation of the experts differed from each other. As a result, the researcher interpreted whether the answers were on the same line, which can leave room for discussion. Ultimately, this has a negative influence on the construct validity.

Conclusion validity The conclusion validity is concerned with the relationship between the treatment and the outcome. Thus, to ensure that there is a statistical relationship [70]. For this study, four expert interviews were used to evaluate the measures. The sample size of four can be considered rather small to create a full picture of whether a certain measure influences team effectiveness. Therefore, it is unclear whether these opinions offer a complete picture of how to evaluate the measures. Furthermore, all teams that have been analyzed apply scrum. However, it is unknown how these teams have adapted the scrum, since research shows that only 50% apply the 'pure' scrum as originally described [16]. This threatened the conclusion validity, since the data of scrum teams have been analyzed with the mind set that all teams apply pure scrum. However, it is unknown which scrum principles each team applies and how mature each team is in applying scrum. Next time, an instrument should be applied to indicate the scrum maturity of a team.

External validity The threat of external validity is related to the extent to which it is possible to generalize the findings [70]. Although all participants can be considered domain experts, these experts are all from the same company. As a result, there has to be some caution in how the conclusions are formulated, since all data and the perspective on the topic are based on a single organization. Furthermore, scrum methodology is also being introduced in other areas such as construction [38], and education [34]. The measures collected in this study focused solely on scrum software development practices.

Although several validity threats are discussed in this section, it is expected that most threats only have a small impact on the results. The main reason for that is that criteria have been formulated to minimize the threats, such as defining selection criteria for participants and applying data triangulation. The only aspects that could influence the results are the sample size of the participants and the generalizability. The sample size aspect has arisen mainly due to time and resource constraints in this research. The generalizability aspect can partly be solved due to the fact that all participants are consultants who also worked with or at other companies. In addition, the participants worked in different industries, such as mobility, healthcare, and finance. As a result, experiences derived from other organizations and different industries were also indirectly taken into account.

6 Conclusions

This research intends to investigate whether it is possible to measure team effectiveness based on quantitative data values (objective measures). The process model of Pfeffers et al. [51] has been applied to structure this research. The problem investigation and the literature study were the first part of this research. These parts helped to find answers for the first and second sub-questions of this research. For the third sub-question, focus groups have been conducted to generate objective measures to measure team effectiveness. In addition, interviews have been carried out to determine whether a measure can be quantified in a work management system. Finally, the measures have been evaluated through expert interviews to indicate whether a measure influences team effectiveness. The results show that the majority of the generated measures contribute scrum team effectiveness. The conclusions of the three sub-questions and the main research questions have been formulated and are stated below.

6.1 Sub-Question 1

-What is the definition of Team Effectiveness?
For this research, the following definition has been applied, *"Team effectiveness includes the quality of the team's performance, as well as the perceived satisfaction of individual needs of team members"* (p. 1108) by Fransen et al. [25]. Studies show that team effectiveness is a widely studied concept. These studies address criteria or factors that influence team effectiveness. Although many papers have discussed the topic of team effectiveness, a general observation was that a definition of team effectiveness is often left out. A definition of team effectiveness has been found in Fransen et al. [25]. This definition consists of two parts, team performance and perceived satisfaction of individual needs of stakeholders. Team performance consists of three parts. Product quality, which indicates whether stakeholders are satisfied and, for instance, deadlines are met; process quality, which concerns the efficiency of team; and lastly, the quality of collaboration, which related to the team's expertise and capacity. The second part of the definition of team effectiveness, the perceived satisfaction of individual needs of team members, related to the happiness of team members. The happiness of the team members plays an important role in the overall well-being and performance of the team [44]. Since the definition of Fransen [25] covers multiple aspects and can be applied as a broader concept, this definition has been used in this study.

6.2 Sub-Question 2

-Which concepts influence Team Effectiveness in Scrum?
Although the concept of team effectiveness has been intensively researched, there are currently few studies that address team effectiveness in scrum. [65]. One of the first studies to discuss scrum team effectiveness has been written by Moe et al. [45]. These studies focused on various factors that affect teamwork in scrum

teams. Although this research offers a comprehensive understanding of the reflection of scrum teams based on teamwork models, teamwork is only one element of the whole picture of team effectiveness, according to Russo[65]. Russo concluded that seven concepts impact team effectiveness in scrum. These concepts are Continuous Improvement, Stakeholder Concern, Responsiveness, Team Autonomy, Management Support, Team Morale, and Stakeholder Satisfaction.

6.3 Sub-Question 3

-Which quantitative data values can be used to measure team effectiveness in scrum?
In total, 29 measures contribute to measuring team effectiveness in Scrum. Several steps have been taken to come to this number of measures. At first, focus groups produced a set of 40 objective measures that are related to scrum. Based on data from scrum teams, 32 measures can be quantified in a work management system, i.e., they can be quantified in a practical setting. Third, the set of 40 measures from the focus groups was evaluated through four expert interviews. The purpose of the expert interviews was to evaluate whether a measure is related to team effectiveness. As a result, 36 measures were identified to be related to team effectiveness. The overviews of both whether a measure can be quantified in a work management system and whether a measure is related to team effectiveness were compared. This overview indicates if a measure contributes to team effectiveness and whether it is measurable. In total, this was the case for 30 measures. As a result, most of the measures generated from the focus groups contribute to measuring team effectiveness.

6.4 Main Research Question

-To which extent can team effectiveness in scrum be measured based on quantitative data values?
The answers of the sub-questions have led to the answer to the main research question. In general, it can be concluded that 29 measures contribute to measuring team effectiveness. To answer the main research question, this result indicates that team effectiveness can be quantified to a large extent based on quantitative data values. However, a few notes should be taken into account. First, a number on its own of a certain measure has no meaning. Therefore, benchmarking or a trend in numbers have 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 quantified in a work management system; for another concept, all linked measures are related and can be quantified in a work management system. As a result, it is hard to state that the complete picture of team effectiveness can be quantified, since the seven concepts of team effectiveness are important for scrum team effectiveness. Although taking into account the mentioned notes, this research can be considered as an exploratory study on the topic of measuring team effectiveness based on

quantitative data values. The results provide a first insight into this topic which can be built on.

6.5 Future research

This section explores the possibilities for future research. These possibilities are based on limitations, validity threats, and research opportunities due to the time constraints of this research. Some future research possibilities are described below.

- To increase reliability, it would be interesting to expand this research to other software companies that apply the scrum principles. This research only interviewed members of scrum teams and analyzed data from scrum teams that belong to the same company. Therefore, it is unknown whether the results can be generalized to other companies. Moreover, conducting focus groups with other companies could be of interest in order to find out whether similar measures would be generated. Therefore, expanding this research to other organizations could provide new information on this topic.

- Another technique to improve reliability is to validate results with statistical proof. Expert interviews were conducted to indicate whether a measure contributes to team effectiveness. Future research could be conducted to validate the results based on statistical proof. These tests can be executed by surveying participants on these measures, and apply statistical tests to this data to prove whether these results are significant. This provides an additional layer of evidence and reliability for the results of this research.

- The result of this study was an overview of measures that help measure team effectiveness in scrum. However, due to time constraints, the measures have not been applied in practice. Future research could be done to apply these measures in a scrum team over a period of time. As a result, feedback can be collected for new measures or current measures can be reexamined.

- The participants in the focus group consisted mainly of scrum masters and product owners. As a result, the measured results focused mainly on the scrum process, rather than on the products generated through scrum. For example, the continuous integration and continuous deployment pipeline is often left out. Therefore, it would be interesting to involve more developers and software engineers in the process of generating measures.

References

1. Acuña, S.T., Gómez, M., Juristo, N.: How do personality, team processes and task characteristics relate to job satisfaction and software quality? Information and Software Technology **51**(3), 627–639 (2009)
2. Albaum, G.: The likert scale revisited. Market Research Society. Journal. **39**(2), 1–21 (1997)
3. Ambler, S., et al.: Roles on agile teams: From small to large teams. vol. 22, p. 2016 (2012)
4. Appelbaum, N.P., Lockeman, K.S., Orr, S., Huff, T.A., Hogan, C.J., Queen, B.A., Dow, A.W.: Perceived influence of power distance, psychological safety, and team cohesion on team effectiveness. Journal of interprofessional care **34**(1), 20–26 (2020)
5. Bahadorestani, A., Naderpajouh, N., Sadiq, R.: Planning for sustainable stakeholder engagement based on the assessment of conflicting interests in projects. Journal of Cleaner Production **242**, 118402 (2020)
6. Benders, J., Van Hootegem, G.: Teams and their context: moving the team discussion beyond existing dichotomies. Journal of Management Studies **36**(5), 609–628 (1999)
7. Bogner, A., Littig, B., Menz, W.: Interviewing experts. Springer (2009)
8. 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. Human resources for health **18**(1), 1–42 (2020)
9. Buljac-Samardzic, M., Dekker-van Doorn, C.M., van Wijngaarden, J.D., van Wijk, K.P.: Interventions to improve team effectiveness: a systematic review. Health policy **94**(3), 183–195 (2010)
10. Camp, R.C., Camp Robert, C.: Benchmarking: the search for industry best practices that lead to superior performance. Quality Press Milwaukee, WI (1989)
11. Campion, M.A., Medsker, G.J., Higgs, A.C.: Relations between work group characteristics and effectiveness: Implications for designing effective work groups. Personnel psychology **46**(4), 823–847 (1993)
12. Campion, M.A., Papper, E.M., Medsker, G.J.: Relations between work team characteristics and effectiveness: A replication and extension. Personnel psychology **49**(2), 429–452 (1996)
13. Cardozo, E.S., Araújo Neto, J.B.F., Barza, A., França, A.C.C., da Silva, F.Q.: Scrum and productivity in software projects: a systematic literature review. In: 14th International Conference on Evaluation and Assessment in Software Engineering (EASE). pp. 1–4 (2010)
14. Chiu, C.Y., Balkundi, P., Owens, B.P., Tesluk, P.E.: Shaping positive and negative ties to improve team effectiveness: The roles of leader humility and team helping norms. Human relations **75**(3), 502–531 (2022)
15. Clarke, M., Oxman, A.: Cochrane reviewers' handbook, version 4.2. 0. Oxford: Cochrane Collaboration (2003)
16. Diebold, P., Ostberg, J.P., Wagner, S., Zendler, U.: What do practitioners vary in using scrum? In: International Conference on Agile Software Development. pp. 40–51. Springer (2015)
17. Donaldson, M.S., Corrigan, J.M., Kohn, L.T., et al.: To err is human: building a safer health system (2000)
18. Doolen, T.L., Hacker, M.E., Van Aken, E.: Managing organizational context for engineering team effectiveness. Team Performance Management: An International Journal (2006)

19. Etikan, I., Musa, S.A., Alkassim, R.S., et al.: Comparison of convenience sampling and purposive sampling. American journal of theoretical and applied statistics **5**(1), 1–4 (2016)
20. Fatema, I., Sakib, K.: Factors influencing productivity of agile software development teamwork: A qualitative system dynamics approach. In: 2017 24th Asia-Pacific Software Engineering Conference (APSEC). pp. 737–742. IEEE (2017)
21. Feldt, R., Magazinius, A.: Validity threats in empirical software engineering research-an initial survey. In: Seke. pp. 374–379 (2010)
22. Fern, E.F.: The use of focus groups for idea generation: the effects of group size, acquaintanceship, and moderator on response quantity and quality. Journal of marketing Research **19**(1), 1–13 (1982)
23. Forsgren, N., Storey, M.A., Maddila, C., Zimmermann, T., Houck, B., Butler, J.: The space of developer productivity: There's more to it than you think. vol. 19, pp. 20–48. ACM New York, NY, USA (2021)
24. Fowler, M.: Refactoring: improving the design of existing code. Addison-Wesley Professional (2018)
25. Fransen, J., Kirschner, P.A., Erkens, G.: Mediating team effectiveness in the context of collaborative learning: The importance of team and task awareness. Computers in human Behavior **27**(3), 1103–1113 (2011)
26. Gil, F., Alcover, C.M., Peiró, J.M.: Work team effectiveness in organizational contexts: Recent research and applications in spain and portugal. Journal of Managerial Psychology (2005)
27. Hackman, J.R., Morris, C.G.: Group tasks, group interaction process, and group performance effectiveness: A review and proposed integration. Advances in experimental social psychology **8**, 45–99 (1975)
28. Hevner, A.R.: A three cycle view of design science research. Scandinavian journal of information systems **19**(2), 4 (2007)
29. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. MIS quarterly pp. 75–105 (2004)
30. Imbrie, P., Maller, S., Immekus, J.: Assessing team effectiveness. In: 2005 Annual Conference. pp. 10–229 (2005)
31. Jahedi, S., Méndez, F.: On the advantages and disadvantages of subjective measures. Journal of Economic Behavior & Organization **98**, 97–114 (2014)
32. Jennex, M.E.: A proposed method for assessing knowledge loss risk with departing personnel. VINE: The journal of information and knowledge management systems (2014)
33. Joshi, A., Kale, S., Chandel, S., Pal, D.K.: Likert scale: Explored and explained. vol. 7, p. 396. SCIENCEDOMAIN International (2015)
34. Jurado-Navas, A., Munoz-Luna, R.: Scrum methodology in higher education: Innovation in teaching, learning and assessment. International Journal of Higher Education **6**(6), 1–18 (2017)
35. Kitchenham, B.: Procedures for performing systematic reviews. vol. 33, pp. 1–26 (2004)
36. Lander, M.C., Purvis, R.L., McCray, G.E., Leigh, W.: Trust-building mechanisms utilized in outsourced is development projects: a case study. Information & Management **41**(4), 509–528 (2004)
37. Lemieux-Charles, L., McGuire, W.L.: What do we know about health care team effectiveness? a review of the literature. Medical care research and review **63**(3), 263–300 (2006)
38. Liu, Y.: Scrum in construction industry to improve project performance in design phase (2018)

39. Marshburn, D.: Scrum retrospectives: Measuring and improving effectiveness. In: Proceedings of the southern Association for Information systems conference (2018)
40. Mathieu, J.E., Gallagher, P.T., Domingo, M.A., Klock, E.A.: Embracing complexity: Reviewing the past decade of team effectiveness research. Annual Review of Organizational Psychology and Organizational Behavior **6**, 17–46 (2019)
41. Mathieu, J.E., Gilson, L.L.: Criteria issues and team effectiveness (2012)
42. Mathieu, J.E., Wolfson, M.A., Park, S.: The evolution of work team research since hawthorne. American Psychologist **73**(4), 308 (2018)
43. Maxwell, J.A.: Qualitative research design: An interactive approach. Sage publications (2012)
44. Miles, S.J., Mangold, G.: The impact of team leader performance on team member satisfaction: The subordinate's perspective. Team Performance Management: An International Journal (2002)
45. Moe, N.B., Dingsøyr, T.: Scrum and team effectiveness: Theory and practice. In: International conference on agile processes and extreme programming in software engineering. pp. 11–20. Springer (2008)
46. Moe, N.B., Dingsøyr, T., Dybå, T.: A teamwork model for understanding an agile team: A case study of a scrum project. vol. 52, pp. 480–491. Elsevier (2010)
47. Morgan, D.L.: Focus groups. Annual review of sociology **22**(1), 129–152 (1996)
48. One, V.: 14th annual state of agile development survey. Version One, INC. (2020), `https://stateofagile.com/#ufh-c-7027494-state-of-agile`
49. Oomen, S., De Waal, B., Albertin, A., Ravesteyn, P.: How can scrum be succesful? competences of the scrum product owner (2017)
50. Peterson, T.M.: Motivation: How to increase project team performance. Project management journal **38**(4), 60–69 (2007)
51. Pfeffers, K., Tuunanen, T., Gengler, C.E., Rossi, M., Hui, W., Virtanen, V., Bragge, J.: The design science research process: A model for producing and presenting information systems research. In: Proceedings of the First International Conference on Design Science Research in Information Systems and Technology (DESRIST 2006), Claremont, CA, USA. pp. 83–106 (2006)
52. Piña, M.I.D., Martínez, A.M.R., Martínez, L.G.: Teams in organizations: a review on team effectiveness. Team Performance Management: An International Journal (2008)
53. Protogerou, C., Hagger, M.S.: A checklist to assess the quality of survey studies in psychology. Methods in Psychology **3**, 100031 (2020)
54. Purvanova, R.K., Kenda, R.: The impact of virtuality on team effectiveness in organizational and non-organizational teams: A meta-analysis. Applied Psychology **71**(3), 1082–1131 (2022)
55. Purwanto, A.: Benefit of benchmarking methods in several industries: A systematic literature review. Sys Rev Pharm **11**(8), 508–518 (2020)
56. Rashidi, M.N., Begum, R.A., Mokhtar, M., Pereira, J.: The conduct of structured interviews as research implementation method. Journal of Advanced Research Design **1**(1), 28–34 (2014)
57. Raymond, J.: Benchmarking in public procurement. Benchmarking: An International Journal (2008)
58. Rizvi, S., Khanam, Z.: A methodology for refactoring legacy code. In: 2011 3rd International Conference on Electronics Computer Technology. vol. 6, pp. 198–200. IEEE (2011)
59. Salas, E., Sims, D.E., Burke, C.S.: Is there a "big five" in teamwork? vol. 36, pp. 555–599. Sage Publications Sage CA: Thousand Oaks, CA (2005)

60. Schwaber, K., Sutherland, J.: The scrum guide. vol. 21, p. 1 (2011)
61. Scissons, E.H.: All numbers are not created equal: Measurement issues in assessing board governance. Corporate Governance: The international journal of business in society (2002)
62. Shah, S.M.A., Papatheocharous, E., Nyfjord, J.: Measuring productivity in agile software development process: a scoping study. In: Proceedings of the 2015 International Conference on Software and System Process. pp. 102–106 (2015)
63. Simon, J.S.: How to conduct focus groups. Nonprofit World **17**, 40–43 (1999)
64. Solla, M., Patel, A., Wills, C.: New metric for measuring programmer productivity. In: 2011 IEEE Symposium on Computers & Informatics. pp. 177–182. IEEE (2011)
65. Verwijs, C., Russo, D.: A theory of scrum team effectiveness (2021)
66. Vom Brocke, J., Schmiedel, T., Recker, J., Trkman, P., Mertens, W., Viaene, S.: Ten principles of good business process management. Business process management journal (2014)
67. Walters, K.K.G.: A study of the relationship between trust and perceived effectiveness in virtual teams. Ph.D. thesis, Capella University (2004)
68. Wilson, V.: Research methods: Focus groups. Evidence based library and information practice **7**(1), 129–131 (2012)
69. Wohlin, C.: Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: Proceedings of the 18th international conference on evaluation and assessment in software engineering. pp. 1–10 (2014)
70. Wohlin, C., Runeson, P., Höst, M., Ohlsson, M.C., Regnell, B., Wesslén, A.: Experimentation in software engineering. Springer Science & Business Media (2012)
71. Wright, D.B., Gaskell, G.D., O'Muircheartaigh, C.A.: How much is 'quite a bit'? mapping between numerical values and vague quantifiers. Applied Cognitive Psychology **8**(5), 479–496 (1994)
72. Wu, Q., Cormican, K.: Shared leadership and team effectiveness: An investigation of whether and when in engineering design teams. Frontiers in Psychology **11**, 569198 (2021)
73. Xiao, Y., Watson, M.: Guidance on conducting a systematic literature review. Journal of planning education and research **39**(1), 93–112 (2019)
74. Yang, M.C., Jin, Y.: An examination of team effectiveness in distributed and colocated engineering teams. International Journal of Engineering Education **24**(2), 400 (2008)

Appendix A

This appendix provides an overview of all measures collected in the first focus group. The measures are divided into two tables. Each table contains the name of the measure and the description of the measure.

Objective Measure	**Description**
The number of times the team complete their sprint goal	The number of times a sprint goal has been achieved over a certain period of time
The number of findings in code review or testing of the code	The number of comments during a code review or testing of the code
Lead time release/story	Lead time is the measurement of how much time passes between task creation and when the work is completed. Bluemel, A.D., (2022). What is Lead Time and Why Should Agile Teams Care?
Cycle time	Cycle time is how long a project takes from starting the work to completion —when the project is ready for delivery. ADOBE COMMUNICATIONS TEAM., (2022). Cycle Time
Built time	The time the developer is researching or building the user story
Release time	The time it takes to execute a release
Test time	The time it takes to test a user story
Review time	The number of minutes/hours for reviewing a user story/release
Downtime	The number of minutes/hours that a computer or IT system is unavailable for use
The number of changes to the product backlog after a sprint review	The number of changes made to the product or sprint backlog after a sprint review
The number of teams working together on the same product	The number of teams working together on the same product

Fig. 8: Objective measures collected in the first focus group (part 1)

Objective Measure	Description
The amount of technical debt in a sprint/release	Technical debt is: the consequences that software projects face when they make trade-offs to implement a lower quality, less complete solutions in order to meet budget and schedule constraints imposed by business realities. Lim, E. (2012). Technical debt: what software practitioners have to say (Doctoral dissertation, University of British Columbia).
The number of retrospective items solved after a new sprint	The number of retrospective items solved after a new sprint
The number of min/hours it takes to do an integration of user story	The number of min/hours it takes to do an integration of user story/release
Story points per user story	Story Point is a measure for relatively expressing the overall size of a user Story or a feature. The value of the Story Point is dependent on the development complexity, effort involved, and the inherent risk and so on. Coelho, E., & Basu, A. (2012). Effort estimation in agile software development using story points. International Journal of Applied Information Systems (IJAIS), 3(7).
The total number of story points in a sprint	The number of user stories in a sprint
The number of releases of a scrum team within a certain period/sprint	The number of releases of a scrum team within a certain period/sprint
Business value	The economic value of a user story or release.
The number of stakeholders attending sprint meetings	The number of stakeholders attending a review meeting
The number of user stories that are released without causing problems	The number of user stories that are released without causing problems for both software quality and for the stakeholders
Done work	The number of user stories/items that are completed after a sprint
The number of scrum team meetings	The number of scrum team meetings during a sprint
User Story age	The age of user story/time that has not been solved yet

Fig. 9: Objective measures collected in the first focus group (part 2)

Appendix B

This appendix provides an overview of all measures linked to the seven concepts of team effectiveness. The table consists of four columns. The first column indicates the team effectiveness concepts. The second column provides the name of the measure. The third column gives an explanation of the measure, and the last column states if the measure is derived from the first focus group or if the measure has been generated in the second focus group.

Team Effectiveness Concept	Objective measure	Definition	Derived from 1st focus group session
Continuous Improvement	The number of retrospective items solved after a new sprint	The number of retrospective items solved after a new sprint	Yes
	The number of bottlenecks in a scrum visualized by a value stream map.	The number of bottlenecks of a scrum team visualized by a value stream map	No
	Software quality (SonarQube)	The quality of software, reproduced by e.g. SonarQube	No
	The number of bugs/defects within a sprint	The number of bugs/defects within a sprint	Yes
	Test time	The time it takes to test a user story	Yes
	Built time	The time the developer is researching or building the user story	Yes
	Release time	The time it takes to execute a release	Yes

Fig. 10: Objective measures related to the Continuous Improvement concept

Team Effectiveness Concept	Objective measure	Definition	Derived from 1st focus group session
Stakeholder Concern	The number of changes to the product backlog after a sprint review	The number of changes made to the product or sprint backlog after a sprint review	Yes
	Response time stakeholders to requests	The time it takes for stakeholders to respond to a question from the scrum team	Yes
	Business value	The economic value of a user story or release. This measure helps stakeholders and scrum teams to prioritize user stories	Yes
	The number of stakeholders attending a sprint review meeting	The number of stakeholders attending a review meeting	Yes
	The number of acceptance tests 'first time right'	The number of acceptance tests that were executed without causing problems for both software quality and for the stakeholders.	Yes
	The number of times the same feedback is addressed by stakeholders	The number of times the same feedback is addressed by stakeholders	No
	The difference between the items\ stories that are created in this sprint compared to the previous sprints.	The difference between the items\ stories that are created in this sprint compared to the previous sprints.	No

Fig. 11: Objective measures linked to the Stakeholder Concern concept

Team Effectiveness Concept	Objective measure	Definition	Derived from 1st focus group session
Responsiveness	Response time stakeholders to requests	The time it takes for stakeholders to respond to a question from the scrum team	Yes
	The time it takes to execute an integration	The number of min/hours it takes before a user story/release can be integrated	Yes
	Done work	The number of user stories/items that are completed after a sprint	Yes
	Built time	The time the developer is researching or building the user story	Yes
	Release time	The number of minutes/hours to implement a new release	Yes
	Test time	The time it takes to test a user story	Yes
	Review time	The number of minutes/hours for reviewing a user story/release	Yes
	Lead time release/story	Lead time is the measurement of how much time passes between task creation and when the work is completed.	Yes
	User Story age	The age of user story/time that has not been solved yet	Yes
	The number of software releases	The number of releases of a scrum team within a certain period/sprint	Yes
	Cycle time	Cycle time is how long a project takes from starting the work to completion—when the project is ready for delivery.	Yes
	The ratio between the working hours and meeting hours	The ratio of working hours and hours of meetings compared	No

Fig. 12: Objective measures linked to the Responsiveness concept

Team Effectiveness Concept	Objective measure	Definition	Derived from 1st focus group session
Management Support	The number of managers attending a sprint review meeting	The number of managers attending a sprint review meeting	Yes
	Management's response time to requests.	The number of hours/minutes that management responds to a request of the scrum team.	No
	Availability and recognizability of management. (in hours available)	The number of hours that management is available for a scrum team.	No
	Resources (euros)	The amount in euros that the scrum team will receive.	No

Fig. 13: Objective measures linked to the Management Support concept

Team Effectiveness Concept	**Objective measure**	**Definition**	**Derived from 1st focus group session**
Team Morale	User Story age	The age of user story/time that has not been solved yet	Yes
	The number of backlog items	The number of backlog items	No
	The number of times the sprint goal is achieved	The number of times that the sprint goal has been achieved over a certain period of time	Yes
	The number of scrum team formation changes	The number of changes within a scrum team	No
	The number of changes to the product backlog after a sprint	The number of changes made to the product or sprint backlog after a sprint review	Yes
	The number of retrospective items solved after a new sprint	The number of retrospective items solved after a new sprint	Yes
	The number of releases to production without bugs	The number of releases to production without bugs	No
	The number of team events at least two members are present	The number of team events at least two members are present	No

Fig. 14: Objective measures related to the Team Morale concept

Team Effectiveness Concept	Objective measure	Definition	Derived from 1st focus group session
Stakeholder Satisfaction	Average velocity previous X sprints.	The average velocity over a certain number of sprints. Velocity represents the amount of work accomplished in each sprint expressed in story points.	Yes
	The number of uncommitted features delivered within a release/sprint.	The number of features delivered after a sprint or release that were not planned or not communicated to the stakeholders.	No
	The finished user stories compared to the predicted number of user stories that need to be fulfilled in the sprint.	The number of solved user stories/tasks compared to the number user stories/tasks that have been scheduled.	Yes
	The number of times the sprint goal has been achieved.	The number of times a sprint goal has been achieved over a certain period of time.	Yes
	Downtime	The number of minutes/hours that a computer or IT system is unavailable for use.	Yes
	Business value	The economic value of a user story or release. This measure helps stakeholders and scrum teams to prioritize user stories.	Yes
	The lead time of a feature compared to the expected delivery time of a feature.	The lead time of a feature compared to the expected delivery time of a feature.	No

Fig. 15: Objective measures linked to the Stakeholder Satisfaction concept

Measure	Interview 1	Interview 2
The number of bugs/defects within a sprint	Fewer bugs suggest that the developers are satisfied with their work. For this measure, the same explanation will be given as the software quality measure. This leads to fewer unexpected costs for a stakeholder and fewer bugs, which improves the happiness of scrum teams. As a result, code quality has an impact on both the perceived satisfaction of the team members and team performance. For this measure, the same explanation will be given as the software quality measure.	The more bugs, the less efficient the team process, since developers will spend more time on bugs than on building features. A few bugs are fine and can be seen as a threshold. A high number of bugs has an impact on team effectiveness. Stakeholder satisfaction is primarily decreased, while team morale could be slightly impacted.
Measure	**Interview 3**	**Interview 4**
The number of bugs/defects within a sprint	Is a consequence of bad software quality, and has the same effects as the software quality measure.	A large number of bugs shows information about the team's consistency. When multiple bugs are found, it indicates that the team and the software quality are inconsistent. This could make the team less effective and have a negative impact on team effectiveness.

Table 24: Four expert opinions on the "number of bugs/defects within a sprint" measure

Appendix C

This appendix provides an overview of all measures related to the seven concepts of team effectiveness. The table consists of eight columns. The first column indicates the team effectiveness concepts. The second column provides the names of the measures. Columns three to eight represent the interview number, and the color indicates whether the measure is measurable in a work management system.

Continuous Improvement Table 16 provides an overview of the Continuous Improvement concept.

		Interview					
Team Effectiveness Concept	**Measure**	I1	I2	I3	I4	I5	I6
Continuous Improvement	The number of retrospective items solved after a new sprint						
	The number of bottlenecks visualized by a value stream map						
	Software Quality (SonarQube)						
	The number of bugs/defects within a sprint						
	Test time						
	Built time						
	Release time						

Table 16: Measures that concern the Continuous Improvement Concept

Stakeholder Concern Table 17 provides an overview of the Stakeholder Concern concept.

		Interview					
Team Effectiveness Concept	**Measure**	I1	I2	I3	I4	I5	I6
Stakeholder Concern	The number of changes to the product backlog after a sprint review						
	Response time stakeholders to requests						
	Business value						
	The number of stakeholders attending a sprint review meeting						
	The number of acceptance tests 'first time right'						
	The number of times the same feedback is addressed by stakeholders						
	The difference between the items\stories that are created in this sprint compared to the previous sprints.						

Table 17: Measures that concern the Stakeholder Concern Concept

Responsiveness Table 18 provides an overview of the responsiveness concept.

		Interview					
Team Effectiveness Concept	**Measure**	I1	I2	I3	I4	I5	I6
Responsiveness	Response time stakeholders to requests						
	The time it takes to execute an integration						
	Done Work						
	Built time						
	Release time						
	Test time						
	Review time						
	Lead time release/story						
	User Story age						
	The number of software releases						
	Cycle time						
	The ratio between the working hours and meeting hours % tijd						

Table 18: Measures that concern the Responsiveness Concept

Management Support Table 19 provides an overview of the Management Support concept.

		Interview					
Team Effectiveness Concept	**Measure**	I1	I2	I3	I4	I5	I6
Management Support	The number of managers attending a sprint review meeting						
	Management's response time to requests						
	Availability and recognizability of management (in hours available)						
	Resources (euros)						

Table 19: Measures that concern the Management Support Concept

Team Morale Table 20 provides an overview of the Team Morale concept.

		Interview					
Team Effectiveness Concept	**Measure**	I1	I2	I3	I4	I5	I6
Team Morale	User Story age						
	The number of backlog items						
	The number of times the sprint goal is achieved						
	The number of scrum team formation changes						
	The number of changes to the product backlog after a sprint review						
	The number of retrospective items solved after a new sprint						
	The number of releases to production without bugs						
	The number of team events at least two members are present						

Table 20: Measures that concern the Team Morale Concept

Stakeholder Satisfaction Table 21 provides an overview of the Stakeholder Satisfaction concept.

		Interview					
Team Effectiveness Concept	**Measure**	I1	I2	I3	I4	I5	I6
Stakeholder Satisfaction	Average velocity on a certain number of sprints						
	The number of uncommitted features delivered within a release/sprint.						
	The finished user stories compared to the predicted number of user stories that need to be fulfilled in the sprint.						
	The number of times the sprint goal has been achieved.						
	Downtime						
	Business value						
	The lead time of a feature compared to the expected delivery time of a feature.						

Table 21: Measures that concern the Stakeholder Satisfaction Concept

Appendix D

This appendix provides an overview of the four expert opinions on whether a measure is related to team effectiveness. In addition to the opinions, 17 literature could be found to support the opinions of the experts.

Measure	Interview 1	Interview 2
Test time	Efficiency will increase if test time decreases. The quicker the test time, the faster the team will be able to anticipate changes and receive feedback. As a result, a test pipeline that is automated becomes more crucial. Ultimatley, the performance of the team will be positively impacted by a reduction in test time.	This could be a disturbing factor for the team process. The shorter the test time, the faster the team can continue building features. This improves team morale. Moreover, a reduction in test time helps to improve problems for the customers faster.
Measure	**Interview 3**	**Interview 4**
Test time	During the software development process, you have different phases, one of these phases is testing. Which can be measured in test time. Whenever this can be measured, we can revise and improve the test time. Ultimately, if the test time gets improved, the team will receive faster feedback and can ship features faster. This will improve stakeholder satisfaction.	The general rule is here, the faster features and user stories can be tested, the better the performance and productivity of the team will be. This will ultimately improve the effectiveness of the team. The explanation of the built time and review time measures are comparable to this measure.

Table 25: Four expert opinions on the "test time" measure

Measure	Interview 1	Interview 2
Built time	All these times are in a sequential order executed. Also on a code level. Similar to test time, as built time is shorter, the team will be able to anticipate changes and get feedback more quickly.	This could be a disturbing factor for the team process. The shorter the built time, the faster the team can continue building features. This improves team morale. Moreover, a reduction in built time helps to improve problems for the customers faster.
Measure	**Interview 3**	**Interview 4**
Built time	During the software development process, you have different phases, one of these phases is building the feature or user story. Which can be measured in built time. Whenever this can be measured, we can revise and improve the built time. Ultimately, if the built time gets improved, the team will receive faster feedback and can ship features faster. This will improve stakeholder satisfaction.	The general rule is here, the faster features and user stories can be built, the better the performance and productivity of the team will be. This will ultimately improve the effectiveness of the team. The explanations of the test time and review time measures are comparable to this measure.

Table 26: Four expert opinions on the "built time" measure

Measure	Interview 1	Interview 2	Interview 3
Software quality (SonarQube)	When the code quality is high, a scrum team will encounter fewer surprises. This leads to fewer unexpected costs for a stakeholder and fewer bugs, which improves the satisfaction of scrum team members. As a result, code quality has an impact on both the perceived satisfaction of the team members and stakeholder satisfaction.	This measure serves as an indicator of the team's products or deliverables. It may not be a problem created by the team. However, the team must examine the underlying causes of the issue. Furthermore, if the software quality falls short of your quality benchmark, it affects the team morale and therefore team effectiveness.	Software quality influences many different factors such as, Technical Debt and Done Work. There needs to be a balance in software quality since too high software quality could lead to reducing the number of features which would negatively influence stakeholder satisfaction. However, low software quality could both influence stakeholders' and team member satisfaction since the software generates more bugs or more rework.
Measure	**Interview 4**	**Literature**	
Software quality (SonarQube)	This measure provides information on software quality instead of the performance or productivity of the team. In contrast, software quality might have a negative impact on stakeholder satisfaction. The team may experience delays as a result of poor software quality, which lowers stakeholder satisfaction.	Jones et al. (2011) concluded that an improvement in software quality will lead to an improvement in cost savings and monitoring. Jones, C., & Bonsignour, O. (2011). *The economics of software quality* . Addison-Wesley Professional. Graziotin et al. (2018) concluded that software quality and happiness both influence each other happy developers built products with a higher software quality and unhappy developers built products with a lower software quality. Graziotin, D., Fagerholm, F., Wang, X., & Abrahamsson, P. (2018). What happens when software developers are (un) happy? Journal of Systems and Software, 140, 32-47.	

Table 23: Four expert opinions and literature on the "software quality" measure

Measure	Interview 1	Interview 2	Interview 3
Release time	A reduction in release time will improve team effectiveness, especially for the stakeholder's satisfaction. Stakeholders will faster get new features and bugs will be solved faster. This generates value for the stakeholder.	This could be a disturbing factor for the team process. The shorter the release time, the faster the team can continue building features. This improves team morale. Moreover, a reduction in test time helps to improve problems for the customers faster.	During the software development process, you have different phases, one of these phases is releasing. Which can be measured in release time. Whenever this can be measured, we can revise the release time and improve the release time. Ultimately, if the release time gets improved, the team will receive faster feedback and can ship features faster. This will improve stakeholder satisfaction.
Measure	**Interview 4**	**Literature**	
Release time	In general, the team wants the release as soon as possible. The faster a release can be done, the more effective a team is. This could mean that the team's effectiveness will get improved. However, this is not always the case. It could also be possible that a team is effective and that the saves all the features for a release. Therefore, this measure could be interpreted in multiple ways.	The number of releases is related to the release frequency of software projects. Khomh et al. (2012) investigated the difference between a rapid-release model (projects with a high frequency of releases) and a traditional release model. There are several benefits of an increase in release frequency. Bugs are fixed faster under rapid-release models, but proportionally fewer bugs are fixed compared to the traditional release model. Finally, as expected, users of a software system developed following a rapid release model tend to adopt new. versions faster compared to a traditional release model. Khomh, F., Dhaliwal, T., Zou, Y., & Adams, B. (2012, June). Do faster releases improve software quality? an empirical case study of mozilla firefox. In 2012 9th IEEE working conference on mining software repositories (MSR) (pp. 179-188). IEEE.	

Table 27: Four expert opinions and literature on the "release time" measure

Measure	Interview 1	Interview 2	Interview 3
The number of bottlenecks in a scrum visualized by a value stream map.	The number of bottlenecks will severely impact the scrum team's productivity. A team's flow will be disrupted by bottlenecks, which has an impact on how well the team performs.	The general rule is, the fewer bottlenecks the better. Each bottleneck lowers your team's productivity. As a result, the team faces greater challenges in completing tasks and user stories. Besides the team performance, the bottlenecks has also a negative influence on team morale.	Depends on the bottleneck in the process. But, taking into account that the bottleneck is in a crucial position, it will influence effectiveness on the performance side of the team. The consquence of a bottleneck have been mentioned in a n example in which a user story is for a long time on the user storyboard.
Measure	**Interview 4**	**Literature**	
The number of bottlenecks in a scrum visualized by a value stream map.	Similar to the number of solved retrospectives items, this does not necessarily provide information on team effectiveness. It could be the case that some bottlenecks are identified. However, whenever teams have some easy workarounds to pass these bottlenecks, your team's effectiveness and performance will not be negatively influenced. Furthermore, it does not quantify team effectiveness.	The ultimate goal of VSM is to identify all types of waste in the value stream and to take steps to try and eliminate these. Furthermore, a value stream map aids communication since it provides a common language about the processes. Abdulmalek, F. A., & Rajgopal, J. (2007). Analyzing the benefits of lean manufacturing and value stream mapping via simulation: A process sector case study. *International Journal of production economics*, *107*(1), 223-236. Besides, that value stream maps are applied in optimizing logistical processes, it is also applicable for software development. In software projects, the use of a vsm led to realistic improvements with a high likelihood of implementation. Ali, N. B., Petersen, K., & De França, B. B. N. (2015). Evaluation of simulation-assisted value stream mapping for software product development: Two industrial cases. Information and software technology, 68, 45-61.	

Table 22: Four expert opinions and literature on the "number of bottlenecks in scrum a value stream map" measure

Measure	Interview 1	Interview 2
The number of changes to the product backlog after a sprint review	This measure provides information on the involvement of stakeholders in the scrum team. Likely, the team did not create the right features when there are a lot of changes. It can also imply that the PO and the stakeholders have not been communicating well. Last but not least, when a team creates something that will not be used by the stakeholders, it demotivates the team. The level of team member satisfaction will be reduced as a result.	The number of product backlog changes influences team effectiveness. This measure indicates that stakeholders have different opinions on what is important as the output. When this happens more frequently, it may be a sign that stakeholders are not being very actively involved. The team needs to adjust the product backlog more frequently, which may frustrate stakeholders, and hurts their level of satisfaction.
Measure	**Interview 3**	**Interview 4**
The number of changes to the product backlog after a sprint review	Whenever there are changes to your product backlog, this means that your stakeholders are involved in your development process. This can be seen as very positive for the team. Ultimately, the high involvement of stakeholders will have a positive influence on team morale.	After a sprint review, if the product backlog is changed, it may indicate that the priorities of stakeholders have been changed or the wrong stakeholders are in present in the meeting. If this happens repeatedly, stakeholder management does not working properly. The level of stakeholder satisfaction may be affected as a result.

Table 28: Four expert opinions on "the number of changes to the product backlog after a sprint review" measure

Measure	Interview 1	Interview 2
Business value	In Team Effectiveness, the business value is related to the stakeholder aspect. Business value helps to prioritize user stories, which provides clarity for stakeholders. The stakeholders will be pleased with this. Additionally, provides also clarity for the team. The team understands where to focus their efforts due to prioritizing.	Business value provides information on the impact and the outcome of a certain feature or user story. Business value helps to visualize the features with the most value for the customer. In team effectiveness, this aspect influences stakeholder satisfaction. Furthermore, it is a communication tool between the scrum team and the stakeholders.
Measure	**Interview 3**	**Interview 4**
Business value	A business value would positively influence both team member satisfaction and stakeholder satisfaction. Business value helps to prioritize user stories. Prioritizing user stories helps to provide clarity to both stakeholders and team members.	The business value of a user story indicates a certain value. Whenever the team delivers user stories with a high business value, stakeholder satisfaction will be positively influenced. Since stakeholders will have features that they admire, and therefore stakeholder satisfaction will be improved.

Table 29: Four expert opinions on the "business value" measure

Measure	Interview 1	Interview 2
The number of stakeholders attending a sprint review meeting	This measure provides information on how a scrum team's stakeholders are involved and on the quality of the sprint reviews. When there are no stakeholders attending sprint review meetings, it is incredibly demotivating for a scrum team and hurts team member satisfaction.	The number of stakeholders attending meetings affects team morale. Each time there are no stakeholders present at a meeting, the team feels as though they are not being heard. Furthermore, the session has been organized to inform stakeholders. The entire meeting is not necessary if there are no stakeholders present.
Measure	**Interview 3**	**Interview 4**
The number of stakeholders attending a sprint review meeting	Many stakeholders attending meetings helps to improve team morale. The team knows where they are building the for. Furthermore, you will receive feedback from stakeholders.	This measure includes information on the team's stakeholders' involvement. A high meeting attendance by stakeholders has a positive impact on team member satisfaction and, consequently, team effectiveness.

Table 30: Four expert opinions on "the number of stakeholder attending a sprint review meeting" measure

Measure	Interview 1	Interview 2
The number of acceptance tests 'first time right'	This measure shows how effective an acceptance test is, which can be seen as very positive for team morale. However, is "first time right" quantifiable? The measure leaves a lot of space for debate over the definition of "first time right".	This is a process indicator, whenever many bugs appear, it can tell that the number of acceptance tests needs to be improved. However, it is hard to link to team effectiveness.
Measure	**Interview 3**	**Interview 4**
The number of acceptance tests 'first time right'	This measure provides information on Team Effectiveness, especially for productivity, performance, and stakeholder satisfaction whenever no bugs are found after a release. Most teams have an acceptable flow to define the first time right? For instance, no bugs within a certain period, for instance, a week.	In my opinion, "first time right" indicates that an acceptance test goes without problems. Problems refer to disruptions in the software or dissatisfaction from stakeholders. As a result, this measure does influence team effectiveness, since whenever successful production without bugs occurs, this team means that the team is effective in their performance and would lead to the happy stakeholder and team members.

Table 31: Four expert opinions on "the number of acceptance tests 'first time right' " measure

Measure	Interview 1	Interview 2	Interview 3
The number of times the same feedback is addressed by stakeholders	This measure indicates that the team does not respond to requests for feedback. As a result, stakeholder satisfaction will decrease. Stakeholders could believe that their voices are not being heard, which might have a negative effect on the response time of requests.	This may positively or negatively impact the team's capacity to satisfy stakeholders. Stakeholder satisfaction is positively impacted whenever there is positive feedback has been repeated. When the feedback is negative, it will negatively impact the team morale and thereby team effectiveness.	For stakeholder satisfaction, this could be very negative. If the feedback item concerns something within the power of the team, this only negatively influences stakeholder satisfaction. Whenever the feedback item concerns something outside the team, this negatively both stakeholder satisfaction and team member satisfaction.
Measure	**Interview 4**	**Literature**	
The number of times the same feedback is addressed by stakeholders	This measure does not provide much information on productivity but regards more stakeholder management and thus stakeholder satisfaction. Whenever stakeholders repeat feedback that is in the scope of the team, then stakeholder satisfaction will be reduced. However, if the same feedback is provided that is outside the scope, then there is no influence on team effectiveness.	This measure can be seen as part of stakeholder management. Within stakeholder management, communication is key in a factor to success. Furthermore, conflicts with stakeholders could negatively impact stakeholder satisfaction. Olander, S., & Landin, A. (2005). Evaluation of stakeholder influence in the implementation of construction projects. International journal of project management, 23(4), 321-328.	

Table 32: Four expert opinions and literature on "the number of times the same feedback is addressed by stakeholders" measure

Measure	Interview 1	Interview 2
The difference between the items/stories that are created in this sprint compared to the previous sprints.	A change in the definition of done or the outcomes of a sprint retrospective may be expressed by a decrease or increase in the number of items/user stories. The refinement of user stories may be crucial in this process. As a general rule, a team can work more effectively the smaller the user stories are.	Does not influence team effectiveness.
Measure	**Interview 3**	**Interview 4**
The difference between the items/ stories that are created in this sprint compared to the previous sprints.	It could be possible that there are changing market conditions. It should not directly regard team effectiveness. It could be that the result is that there are too many items on the backlog which could reduce focus, but no direct impact.	This does not say much about team effectiveness. It could be that more stories are refined, which could lead that there being more stories in a sprint. In general, I see no link between team effectiveness and this measure

Table 33: Four expert opinions on "the difference between the items/stories that are created in this sprint compared to the previous sprints" measure

Measure	Interview 1	Interview 2
The time it takes to execute an integration	A reduction in integration time will improve team effectiveness, especially for the stakeholder's satisfaction. Stakeholders will faster get new features and bugs will be solved faster. This generates value for the stakeholder.	This could be a disturbing factor for the team process. The shorter the time of integration, the faster the team can continue building features. This improves team morale. Moreover, a reduction in time of an integration time helps to improve problems for the customers faster.
Measure	**Interview 3**	**Interview 4**
The time it takes to execute an integration	During the software development process, you have different phases, one of these phases is the integration of a product. Which can be measured as integration time. Whenever this can be measured, we can revise the integration time and improve integration time. Ultimately, if the integration time gets improved, the team will receive faster feedback and can ship features faster. This will improve stakeholder satisfaction.	The integration time with another system can be seen similar as to the teams that work on the same product. Whenever the team needs to integrate with new systems, this could bring dependencies and negatively influences performance and thus team effectiveness. Furthermore, a long integration has a negative influence on your lead time.

Table 34: Four expert opinions on "the time it takes to execute an integration" measure

Measure	Interview 1	Interview 2	Interview 3
Number of Releases	An increase in the number of releases will improve team effectiveness, especially for stakeholder satisfaction. A feedback loop is created for the team. Moreover, stakeholders will faster get new features and bugs will be solved faster. As a result, an increase in the number of releases generates value for the stakeholder.	The team's flexibility increases with the number of releases. The more releases, the more comfortable the teams get with releases. Releases indicate that teams have control over the situation. They also succeed in their goals. A release is a risk-involved change. As a result, it is excluded from stakeholder satisfaction. It more closely relates to improving the team performance, since teams will grow in release experience.	An increase in releases generates fast value for the customer. The faster the release, the faster it generates value for the customer. This improves stakeholder satisfaction. However, more code could result in more bugs. This could be a downside.
Measure	**Interview 4**	**Literature**	
Number of Releases	In general, a team's effectiveness increases with the number of releases it makes. The main reason for this is that the team gets more comfortable with releasing. As a consequence, the team will operate more efficiently and experience fewer surprises. Stakeholder and team member satisfaction both increase as a result.	The number of releases is related to the release frequency of software projects. Khomh et al. (2012) investigated the difference between a rapid-release model (projects with a high frequency of releases) and a traditional release model. There are several benefits of an increase in release frequency. Bugs are fixed faster under rapid-release models, but proportionally fewer bugs are fixed compared to the traditional release model. Finally, as expected, users of a software system developed following a rapid release model tend to adopt new. versions faster compared to a traditional release model. Khomh, F., Dhaliwal, T., Zou, Y., & Adams, B. (2012, June). Do faster releases improve software quality? an empirical case study of mozilla firefox. In 2012 9th IEEE working conference on mining software repositories (MSR) (pp. 179-188). IEEE.	

Table 35: Four expert opinions and literature on the "the number of releases" measure

Measure	Interview 1	Interview 2
Review time	All these times are in a sequential order executed. Also on a code level. Similar to test time, as review time is shorter, the team will be able to anticipate changes and get feedback more quickly.	This could be a disturbing factor for the team process. The shorter the review time, the faster the team can continue building features. This improves team morale. Moreover, a reduction in review to improve problems for the customers faster.
Measure	**Interview 3**	**Interview 4**
Review time	During the software development process, you have different phases, one of these phases is reviewing. Which can be measured in review time. Whenever this can be measured, we can revise the review time and improve the review time. Ultimately, if the review time gets improved, the team will receive faster feedback and can ship features faster. This will improve stakeholder satisfaction.	The general rule is here, the faster features and user stories can be reviewed, the better the performance and productivity of the team will be. This will ultimately improve the effectiveness of the team. The explanations of the test time and built time measures are comparable to this measure.

Table 36: Four expert opinions on the "Review view" measure

Measure	Interview 1	Interview 2	Interview 3
Lead time release/story	All these times are in a sequential order executed. Also on a code level. Similar to cycle time, as lead time is shorter, the team will be able to anticipate changes and get feedback more quickly.	This could be a disturbing factor for the team process. The shorter the lead time of a user story, the faster the team can continue building features. This improves team morale. Moreover, a reduction in lead time helps to improve problems for the customers faster.	During the software development process, you have different phases, the lead time is the cycle time, combined with the time a user story is on the product backlog. In other words, lead time is the time a user story is created until the feature has been integrated. Whenever this can be measured, we can revise the lead time and improve the lead time. Ultimately, if the lead time gets improved, the team will receive faster feedback and can ship features faster. This will improve stakeholder satisfaction.
Measure	**Interview 4**	**Literature**	
Lead time release/story	The influence of lead time depends on the situation. In my opinion, whenever a user story needs to be fulfilled, the story will be put into the product backlog. However, it is unknown how long the story will be in the backlog and thus how long the lead time is. Therefore, it does not say much about team effectiveness. In general, this also is the case for the user story age and the cycle time. Therefore, the same explanation will be given.	The benefits of a shorter lead time for the stakeholder have been discussed. For instance, a shorter lead time leads to an early enrollment of the products. As a result, the product conforms more to the expectations of the market Petersen, K. (2010). An empirical study of lead-times in incremental and agile software development. In New Modeling Concepts for Today's Software Processes: International Conference on Software Process, ICSP 2010, Paderborn, Germany, July 8-9, 2010. Proceedings (pp. 345-356). Springer Berlin Heidelberg.	

Table 37: Four expert opinions and literature on the "Lead time release/story" measure

Measure	Interview 1	Interview 2
User Story age	The longer a user story is open, the more waste it creates. Team members invest time in organizing the backlog and the more items it contains, the more time it costs. Furthermore, this could mean that team members already have looked into the user story, which takes time. Then, a user story has been parked for some reason and team members have to continue with the user story after some time. However, most of the information will certainly be lost. So, team members have to invest more time in the user story. A negative effect on the efficiency and productivity of the team.	The shorter the user story age, the better. The focus of the team has to be on finishing tasks. A user story that stays open costs results in focus and concentration issues. This would effect the productivity and the team performance of the team.
Measure	**Interview 3**	**Interview 4**
User Story age	Age stories that have a high age influence team effectiveness. The team will lose focus. Whenever stories have no priority, the team should delete them. It generates waste in the backlog. This also could mean that something went wrong up front. Ultimately, the loss of focus will decrease the performance and satisfaction of the team.	The influence of lead time depends on the situation. In my opinion, whenever a user story needs to be fulfilled, the story will be put into the product backlog. However, it is unknown how long the story will be in the backlog and thus how long the lead time is. Therefore, it does not say much about team effectiveness. In general, this also is the case for the user lead time and the cycle time. Therefore, the same explanation will be given.

Table 38: Four expert opinions on the "User story age" measure

Measure	Interview 1	Interview 2
Cycle time	All these times are in a sequential order executed. Also on a code level. Similar to lead time, as cycle time is shorter, the team will be able to anticipate changes and get feedback more quickly.	This could be a disturbing factor for the team process. The shorter the cycle time, the faster the team can continue building features. This improves team morale. Moreover, a reduction in cycle time helps to improve problems for the customers faster.
Measure	**Interview 3**	**Interview 4**
Cycle time	During the software development process, you have different phases, the total time of creating the user story untill shipping the feature is called cycle time. Whenever this can be measured, we can revise the cycle time and improve the cycle time. Ultimately, if the cycle time gets improved, the team will receive faster feedback and can ship features faster. This will improve stakeholder satisfaction.	The influence of lead time depends on the situation. In my opinion, whenever a user story needs to be fulfilled, the story will be put into the product backlog. However, it is unknown how long the story will be in the backlog and thus how long the lead time is. Therefore, it does not say much about team effectiveness. In general, this also is the case for the lead time and the user story age. Therefore, the same explanation will be given.

Table 39: Four expert opinions on the "Cycle time" measure

Measure	Interview 1	Interview 2
The ratio between the working hours and meeting hours	The general rule here is, the more time team members attend meetings, the less time they can code. In an ideal situation, you want to minimize the number of meetings to the scrum principles. This measure can be seen as a reflection indicator since the number of meetings can be seen as a consequence of why developers are not finishing their work. The satisfaction of stakeholders may decline if this is the case.	This measure is an indicator of the team's performance of the team. The Agile way of working has been designed to be built as fast as possible. Meeting cultures do not produce results. However, it could be an indirect indicator of team performance. However, if the output is fine, this measure does not say anything. Therefore, this measure is situation-dependant.
Measure	**Interview 3**	**Interview 4**
The ratio between the working hours and meeting hours	Depends on what can be seen as a meeting. In a way, a helpful meeting could improve the effectiveness, since they can generate new knowledge. However, many meetings outside the usual scrum rituals could reduce effectiveness. These meetings could lead to an overhead or are not beneficial for the team.	The difference between the meeting hours and work hours does influence team effectiveness. The team will have less time to complete user stories and build features the more meetings they attend. As a result, this measure provides information on team effectiveness, particularly on the productivity of the team.

Table 40: Four expert opinions on the "The ratio between the working hours and meeting hours" measure

Measure	Interview 1	Interview 2
The number of managers attending a sprint review meeting	Can be seen as similar to the number of stakeholders attending a sprint review meeting meaure. This measure provides information on how a scrum team's stakeholders are involved. When there are no stakeholders attending sprint review meetings, it is incredibly demotivating for a scrum team and has a negative impact on team member satisfaction.	In a normal situation, management is not involved in review meetings. A manager only attends a meeting to check the team's performance. Therefore, the attendance of managers could lead to a reduction in team morale, since the performance of a team is not as expected.
Measure	**Interview 3**	**Interview 4**
The number of managers attending a sprint review meeting	Managers can be seen as a stakeholder. So it could relate to the evaluation of the number of stakeholders attending meetings. However, this can be seen differently if the manager has a justice system. If this is the case, it does not say much about Team Effectiveness.	The management has something magical. Whenever management participates in meetings and shows interest, the team feels that they are taken seriously. This could improve team member satisfaction and thus team effectiveness. However, if management is not interested and does appear in these meetings, team member satisfaction could be reduced.

Table 41: Four expert opinions on the "The number of managers attending a sprint review meeting" measure

Measure	Interview 1	Interview 2
Management's response time to requests.	Can be seen as similar to response time to stakeholders measure. This measure can be seen as a bottleneck. Without the input of stakeholders, features could in some cases not be finished unless the stakeholders provide answers or explanations. A high response time has a negative influence on team performance and team member satisfaction.	The number of management requests should be minimal. The team should be enabled. Therefore, the number of requests and a high response time to requests has a negative influence on team member satisfaction.
Measure	**Interview 3**	**Interview 4**
Management's response time to requests.	When response time is short it indicates the importance of a team. It clarifies the situation for the team and boosts morale.	This measure can be seen as similar to the response time to stakeholder measures. Therefore, the same explanation of the response to stakeholders will be provided.

Table 42: Four expert opinions on the "The number of managers attending a sprint review meeting" measure

Measure	Interview 1	Interview 2
Availability and recognizability of management. (in hours available)	This provides insight into how much management cares about the team. The more the management is available, the more they get a feeling that they are being heard. This will lead to an improvement in team member satisfaction.	Management should be transparent and enabled. This measure concerns team effectiveness on this matter. So, whenever this is not the case, the team morale could be reduced.
Measure	**Interview 3**	**Interview 4**
Availability and recognizability of management. (in hours available)	When management allocates a significant amount of time to a team, it indicates that the team is important. It clarifies the situation for the team and boosts morale.	Similar to the managers attending review meetings. If the management is not visible or does not show interest, team member satisfaction will be reduced. However, it could also work the other way around. Whenever management shows interest, team member satisfaction can be boosted.

Table 43: Four expert opinions on the "The number of managers attending a sprint review meeting" measure

Measure	Interview 1	Interview 2
Availability and recognizability of management. (in hours available)	This provides insight into how much management cares about the team. The more the management is available, the more they get a feeling that they are being heard. This will lead to an improvement in team member satisfaction.	Management should be transparent and enabled. This measure concerns team effectiveness on this matter. So, whenever this is not the case, the team morale could be reduced.
Measure	**Interview 3**	**Interview 4**
Availability and recognizability of management. (in hours available)	When management allocates a significant amount of time to a team, it indicates that the team is important. It clarifies the situation for the team and boosts morale.	Similar to the managers attending a sprint review meeting meaure . If the management is not visible or does not show interest, team member satisfaction will be reduced. However, it could also work the other way around. Whenever management shows interest, team member satisfaction can be boosted.

Table 44: Four expert opinions on the "The Management's response time to requests" measure

Measure	Interview 1	Interview 2	Interview 3
The amount of technical debt in a sprint/release	Technical debt is the result of past actions that could result in future work debt, whether you are aware of it or not. It can be seen as a postponement of work. Finishing a user or feature in a specific way will result in the team more work in the future. More bugs or code factoring will result from this. Surprises would happen more frequently, which would be bad for both team members and stakeholders.	Increased technical debt harms the team's performance and thus team effectiveness. Technical debt is also related to your software quality. A team works to improve the quality. However, earlier decisions resulted in technical debt. Additionally, it indirectly affects shareholder satisfaction since user stories and releases take more time to complete.	An increase in technical debt has a bad influence on team morale. Much technical debt reduces the quality of the product. This could frustrate team members since it might result that a great length of sprint needing to be used to fix the technical debt.
Measure	**Interview 4**	**Literature**	
The amount of technical debt in a sprint/release	The amount of technical debt is related to the quality of your software. However, for me, it is unsure whether it is also related to team effectiveness. Technical debt is often a choice that has been made for the team. For instance, there is a strict deadline for the stakeholder. One choice could be to deploy the platform now and address bugs later. This does not imply that the team is ineffective. Technical debt has no effect on team effectiveness when the stakeholder receives value. The debt will be resolved in a later stage of the project.	Technical debt both negative influence stakeholders. For stakeholders technical debt creates significant long-term problems, such as increased maintenance costs Brown, N., Cai, Y., Guo, Y., Kazman, R., Kim, M., Kruchten, P., ... & Zazworka, N. (2010, November). Managing technical debt in software-reliant systems. In Proceedings of the FSE/SDP workshop on Future of software engineering research (pp. 47-52). And for team member satisfaction, Addressing technical debt is a mundane task for many developers, and it seems likely that developers would find the effects of technical debt frustrating in the long term. Tom, E., Aurum, A., & Vidgen, R. (2013). An exploration of technical debt. Journal of Systems and Software, 86(6), 1498-1516.	

Table 45: Four expert opinions and literature on "the amount of technical debt in a sprint/release" measure

Measure	Interview 1	Interview 2
Resources (euros)	It is not always beneficial to add more developers or to provide more money for the team. You have to make a broad analysis and set budgets. Therefore, in my opinion, there is no connection between resources (euros) and effectiveness. This measure can be interpreted in a wide context.	Does not say anything. Fit for purpose is the most important.
Measure	**Interview 3**	**Interview 4**
Resources (euros)	This measure is linked to stakeholder engagement. Often, this measure is a consequence of something. Whenever management provides many resources, the management sees value in the team. Furthermore, the team feels that they are being heard, which could boost team morale.	Not much can be concluded about team effectiveness from this measure. It is much more useful to get insight into how you spend these resources, instead of just an increase or decrease in resources.

Table 46: Four expert opinions on the "Resources" measure

Measure	Interview 1	Interview 2	Interview 3
The number of reviews/acceptance tests executed by external parties	The more external reviews/acceptance tests the team has, the more frustrated the team will be. Especially if the external reviews/acceptance is a quality gate. The team must frequently wait for others to approve the review or acceptance test. It does, however, bring benefits. The team is driven to deliver excellent quality since the more external reviews and acceptance tests there are, the longer it will take for the feature or user story to be accepted. The measure can therefore be interpreted in a very wide context.	It does not influence external reviews. Acceptance tests do. The acceptance test is a waste of time whenever a company or stakeholder is unavailable. Additionally, the team is more reliant on stakeholders. Additionally, this might increase the wait time and decreases the team's performance.	The number of external review/acceptance tests could both negatively or positively team morale. Negatively, since you have to wait for other people to revise the work. This increases waiting time, which negatively impacts the performance of the team. Furthermore, it could also positively impact team morale, since the team gets a feeling that stakeholders are involved.
Measure	**Interview 4**	**Literature**	
The number of reviews/acceptance tests executed by external parties	The number of external reviews/acceptance tests will have an impact on team effectiveness. However it depends on how much work the group must put in to execute the external reviews and acceptance testing. Whenever many things have to be organized for the external review/acceptance test, this will negatively influence the team's performance. Since the team members must spend a lot of time to these tasks, they are unable to continue developing features.	Overall, the results showed that team autonomous helping benefited team role-based functions and ultimately team effectiveness, whereas team dependent helping hindered them Lee, S. H., Liu, Y., Koopmann, J., Seo, J. Y., Zhou, L., & Yu, Y. (2023). Not Always Helpful: Linking Intrateam Helping Types to Team Effectiveness From a Role Theory Perspective. Journal of Management, 01492063221149676.	

Table 47: Four expert opinions and literature on "the number of reviews/acceptance tests executed by external parties" measure

Measure	Interview 1	Interview 2	Interview 3
The number of user stories/items executed by a minimum of 2 scrum team members	Whenever two team members work on a user story, knowledge will be shared. This will have a positive effect on efficiency and productivity, since a team member is sick or leaves the team, the knowledge will stay in the team. In general, it takes a little more time to finish a user story. Furthermore, the quality is oftentimes also higher. However, besides the benefits, the drawbacks are the dependency on your partner will increase, and communication issues could arise. This could negatively influence efficiency.	This action helps to raise the quality of the software. The measure indicates the idea of the four eyes principle. The quality rises as more people work on a specific user story/item. The information also remains within the team if the structure of the team changes. It affects the team's effectiveness and quality. Despite the benefits, dependency issues could create problems in de future.	Depends on the person, since some people like to work on their own and other people like to work in (smaller) groups. Working on user stories together can positively influence you since you share knowledge about a part of the product. It can also negatively influence you since you have more communication lines and tension could arise.
Measure	**Interview 4**	**Literature**	
The number of user stories/items executed by a minimum of 2 scrum team members	Too many developers working on the same user story has an impact on the team's performance. Other user stories might not be able to be completed as a result, which would be negative to the team's performance. The majority of the time, though, the developers are aware how many people are needed to built a certain feature or user story.	Overall, the results showed that team autonomous helping benefited team role-based functions and ultimately team effectiveness, whereas team dependent helping hindered them Lee, S. H., Liu, Y., Koopmann, J., Seo, J. Y., Zhou, L., & Yu, Y. (2023). Not Always Helpful: Linking Intrateam Helping Types to Team Effectiveness From a Role Theory Perspective. Journal of Management, 01492063221149676.	

Table 48: Four expert opinions and literature on "the number of user stories/items executed by a minimum of 2 scrum team members" measure

Measure	Interview 1	Interview 2
The number of backlog items	Here, the distinguishing can be made between a product backlog and a sprint backlog. An extensive sprint backlog can reduce the focus of a team, which negatively affects performance and team morale. For a product backlog, this indicates that a project is becoming bigger and bigger and that the stakeholder demands more from the team.	The quantity of backlog items affects the effectiveness of the team. The number of backlog items has both a lower and an upper limit. Once the group reaches the upper limit of backlog items, the team's concentration will decrease. Whenever the lower limit is reached. The team is unable to look to the future. The team is unable to brainstorm solutions with stakeholders. It affects team effectiveness on both the stakeholder and team member sides. Furthermore, the team's concentration may become less intense whenever there is a product backlog with a large number of user stories.
Measure	**Interview 3**	**Interview 4**
The number of backlog items	Oftentimes, there is a lower and upper limit for backlog times. Whenever the lower limit has been reached, refinement is needed to increase the number of backlog items. Otherwise, a situation could occur that the team can not continue building items, since the backlog. Whenever the upper limit has been reached, it takes longer to finish user stories, clutter could arise and a team loses focus. It mostly influences performance and team morale.	This measure does provide not much information on team effectiveness. In my opinion, each item that is important will be put on the product and sprint backlog. Hence, the effectiveness of the team is unaffected by the size of your backlog. The only drawback is that too few backlog items will make the team less effective, since it is possible that the developers won't be able to continue working.

Table 49: Four expert opinions on "the number of backlog items" measure

Measure	Interview 1	Interview 2	Interview 3
The number of times the sprint goal is achieved	Achieving your sprint goal reveals something about a scrum team's predictability. It becomes simpler to forecast the number of story points for a sprint whenever a team meets its sprint target. This is would satisfy stakeholders. Moreover, a sprint goal's achievement also increases team member satisfaction.	The team morale will improve, the more times the team reaches its sprint goal. For the stakeholders in particular. This results in the team being more dependable and better predictable. Additionally, it can support increasing the team's morale. The completion of stories always stimulates morale. As a result, it affects the effectiveness of the team.	The first question is, How well is the the sprint goal defined. Whenever this is the case, achieving the sprint goal improves both stakeholder and team member satisfaction. It creates joy for team members since they achieve a certain goal. Also for stakeholders, they know that the team is planning.
Measure	**Interview 4**	**Literature**	
The number of times the sprint goal is achieved	If the scrum team is unable to accomplish its sprint goal, this may be a sign that something is wrong within the team. Something must alter in order to improve team performance. In the long run, this will also have a negative impact on team member satisfaction and consequently team effectiveness. The main drawback to this measure is the possibility that the team will set an overly ambitious sprint goal. Hence, the effectiveness of the team inside the team may be good, but their formulation may be excessively ambitious.	The completion of valued tasks, especially in a group, as at work (see below), though this is perhaps more a cause of satisfaction. Some experiences of joy have a dimension of depth, intensity, "absorption" or "flow", for example when tackling a demanding task the completion of valued tasks, especially in a group, as at work (see below), though this is perhaps more a cause of satisfaction. Some experiences of joy have a dimension of depth, intensity, "absorption" or "flow", for example when tackling a demanding task Argyle, M., & Martin, M. (1991). The psychological causes of happiness. Subjective well-being: An interdisciplinary perspective, 77-100.	

Table 50: Four expert opinions and literature on "the number of times the sprint goal is achieved" measure

Measure	Interview 1	Interview 2	Interview 3
The number of scrum team formation changes	This has a negative influence on team effectiveness since it influences both team morale and the safety feeling of team members. Furthermore, onboarding takes also time, in which one team member has to guide the new team member. This distracts him from building new features.	The effectiveness of the team suffers the more team changes there are. Whenever there is a change in the group process, the process of group forming enters a new phase. There will be a decline in team performance. Stakeholders may also experience indirect dissatisfaction after a team change.	After a team change, efficiency and productivity will be reduced. The main reason for this is that the new team member needs to get comfortable with the team. However, the new team member could bring insights, which in the end, could resolve in higher productivity and efficiency.
Measure	**Interview 4**	**Literature**	
The number of scrum team formation changes	A change in your team formation does influence team effectiveness. Every time a team member changes, the other team members are busy with onboarding, or assisting the new team member. Moreover, knowlegde can be lost after a change in the team formation. As a result, performance and perhaps team member satisfaction will decline.	Consequently, the impact of diversity on productivity is identified by relatively marginal changes in the composition of a team that may already have set routines and communication patterns. Hamilton, B. H., Nickerson, J. A., & Owan, H. (2012). Diversity and productivity in production teams. In Advances in the Economic Analysis of participatory and Labor-managed Firms (Vol. 13, pp. 99-138). Emerald Group Publishing Limited. Bonebright, D. A. (2010). 40 years of storming: a historical review of Tuckman's model of small group development. Human Resource Development International, 13(1), 111-120. Tuckman identifies several phases, in which first the team goes from a forming to a storming phase which has negative consequences for the effectiveness of team. However, after a certain period goes from a norming to performing phase in which the team is on a steady performing level.	

Table 51: Four expert opinions and literature on "the number of scrum formation changes" measure

Measure	Interview 1	Interview 2
The number of releases to production without bugs	This measure shows how effective a release test is. However, is first without bugs quantifiable, and in what period? The measure leaves a lot of space for debate over the definition of what is without bugs.	This is a process indicator, whenever many bugs appear, it can tell that the number of acceptance tests needs to be improved. However, it is hard to link to Team Effectiveness.
Measure	**Interview 3**	**Interview 4**
The number of releases to production without bugs	If the number of releases to production without bugs is null, this provides information on productivity, performance, and stakeholder satisfaction. However, it is hard to tell what without bugs means. In some cases, teams define an acceptance flow. For instance, if the acceptance flow is a week, then the first week without bugs can be seen as a successful release to production.	Without bugs can be seen similarly as “first time right”. Therefore, a similar definition of the first-time-right acceptance tests can be applied to this measure. In my opinion, "first time right" indicates that the release to production goes without problems. Problems refer to disruptions in the production platform or dissatisfaction from stakeholders. As a result, this measure does influence team effectiveness, since whenever successful production without bugs occurs, this team means that the team is effective in their performance and would lead to the happy stakeholder and team members.

Table 52: Four expert opinions on ”the number of releases to production without bugs” measure

Measure	Interview 1	Interview 2	Interview 3
The number of team events at least two members are present	This measure is an indicator of team morale. Team members will get a feeling that they have psychological safety. Which can boost team morale. No team activity is a missed opportunity.	This helps the team to come together, other than for instance, to achieve your sprint goal. It is good for team formation and group dynamics. It improves team effectiveness based on team member satisfaction.	This measure provides information on the involvement of the team. Does not contribute to productivity. The measure could be changed so a certain percentage instead of minimum of 2 team members.
Measure	**Interview 4**	**Literature**	
The number of team events at least two members are present	Looking at the satisfaction of the team members, I believe that this measure does affect team effectiveness. Activities for the team build bonding and increase respect among team members. As a result, group dynamics will be strengthened, which will enhance team member satisfaction and therefore boost team effectiveness.	Improve team morale due to social gatherings. Activities such as picnics, family days, organized recreation, philanthropic work, or holiday parties can create an atmosphere of caring and support that will give employees a sense of belonging that will carry over to their work. Hopkins, H. (1995). A challenge to managers: five ways to improve employee morale. Executive Development, 8(7), 26-28.	

Table 53: Four expert opinions and literature on ”the number of team events at least two members are present” measure

Measure	Interview 1	Interview 2	Interview 3
Average velocity previous X sprints	This measurement must show a trend over a defined period of time. The team is improving whenever there is a rising trend in velocity. When this is constant, this indicates that the group has reached its potential. Stakeholders will be happy if it is stable or rising because it says something about your team's predictability	In the area of team effectiveness, it discusses stakeholder satisfaction. The stakeholder will be less satisfied whenever the velocity decreases. Stakeholders will be pleased whenever it increases or remains steady at the benchmark. It serves as an output/outcome indication for the team.	Velocity does not say something explicitly about the productivity and efficiency of the team. However, it could benefit if there is a trend visible with the velocity regarding the problems of why velocity increases or decreases. For stakeholders, a stable velocity helps to improve the predictability of the team. This improves stakeholder satisfaction.
Measure	**Interview 4**	**Literature**	
Average velocity previous X sprints	This measure provides information on the effectiveness of a team, especially team performance. Velocity offers information on whether the team completes more user stories and, as a result, is more productive. It could go either way, as an increase in velocity indicates that the team is more efficient and a decline indicates that the team's performance is declining.	Al-Sabbagh et al. (2018) discussed that the velocity measurement used in this paper reflects the teams' efficiency in accomplishing scrum tasks while planning effectiveness reflects their ability to estimate and deliver, within each sprint, the expected outcome. Al-Sabbagh, K. W., & Gren, L. (2018). The connections between group maturity, software development velocity, and planning effectiveness. Journal of Software: Evolution and Process, 30(1), e1896.	

Table 54: Four expert opinions and literature on the "Average velocity previous sprints" measure

Measure	Interview 1	Interview 2
The number of uncommitted features delivered within a release/sprint.	This measure, in my opinion, tells something about the quality of your planning. You have to commit that you are going to do something extra in your sprint. Stakeholders see something additional as an outcome, which is positive. When additional features are not structural, it is a good thing. If the additional features are structural, your planning is incorrect, and the team may become disorganized.	Whenever this is the case and the team achieves its sprint goal. It means that the team outperformed itself. As a result, it has a direct result on Team Effectiveness. It positively influences the stakeholder and the team members.
Measure	**Interview 3**	**Interview 4**
The number of uncommitted features delivered within a release/sprint.	In general, the negative influences team effectiveness. This could mean that some things are not clear or that communication is lacking. For stakeholders' satisfaction, this is not good, since the team does something different than asked. As a result communication or coordination-wise something has gone wrong.	This has a negative influence on both the team's performance and the stakeholders satisfaction. Considering this measure may show that the team's planning is off or that it is developing the incorrect features. Therefore, this measure can be seen as a red flag for the team. Furthermore, this will also have a negative influence on the stakeholders' satisfaction.

Table 55: Four expert opinions on "the number of uncommitted features delivered within a release/sprint" measure

Measure	Interview 1	Interview 2	Interview 3
The finished user stories compared to the predicted number of user stories that need to be fulfilled in the sprint.	The team will receive a commission whenever the planning is close to the deadlines. This means that the team meets its deadlines. The planning is either off-target or something else is wrong if there is a significant difference between the deadlines and the plans. Whenever this is the case the morale of the team and stakeholder satisfaction are impacted negatively.	Planned and finished story points tell something if the team achieves your sprint goal. Planning helps to indicate the predictability and reliability of the team. Whenever a team is predictable and reliable, this results in happy stakeholders. If the team is not in line with the planning, this results in unhappy stakeholders.	Tells something about predictability. The team predicts what tasks need to be solved and what is delivered. It ultimately affects stakeholder satisfaction. Since targets are not met will lead to negative stakeholder satisfaction and when targets are met, the stakeholder satisfaction will be stable or increase.
Measure	**Interview 4**	**Literature**	
The finished user stories compared to the predicted number of user stories that need to be fulfilled in the sprint.	This measure can be seen as similar to the velocity measurement. If the team is in line with their planning, the velocity of the team will get improved. Velocity offers information on whether the team completes more user stories and, as a result, is more productive. It could go either way, as an increase in velocity indicates that the team is more efficient and a decline indicates that the team's performance is declining. The main difference is that velocity is quantified in story points and this measure provides information on the user stories.	Predictability is an important outcome of this measure. Research shows the influence of predicitablity on Upper-level and project team members. Upper-level management: Upper-level managers were particularly concerned with the organizational-level impacts of the project and view predictability as a key determinant to insuring intended outcomes. Project team members: Concerned with project duration, cost, scope and system functionality, project team members rely on predictability to achieve interim and final project targets. Achieving these team-based goals begets trust amongst the project actors. Lander, M. C., Purvis, R. L., McCray, G. E., & Leigh, W. (2004). Trust-building mechanisms utilized in outsourced IS development projects: a case study. Information & Management, 41(4), 509-528.	

Table 56: Four expert opinions and literature on "the finished user stories compared to the predicted number of user stories that need to be fulfilled in a sprint" measure

Measure	Interview 1	Interview 2	Interview 3
Downtime	Unplanned downtime harms performance since it prevents developers from creating features, which suggests that something went wrong in terms of planning. Stakeholder satisfaction will eventually decline.	A very strong indicator of team effectiveness, from a stakeholder point of view. You have planned and unplanned downtime. Team effectiveness will only be influenced by unplanned downtime. The more unplanned downtime, the unhappier the stakeholders become. However, it depends on your service level agreement with the stakeholders	This is an indicator of stakeholder satisfaction. Whenever the downtime is more than projected. team effectiveness will decrease. To be more precise, stakeholder satisfaction will be reduced.
Measure	**Interview 4**	**Literature**	
Downtime	Downtime affects team effectiveness. The main reason for this is that whenever is down, the team could not merge code or build new features. Furthermore, this will have a negative influence on your satisfaction of stakeholders, since the team could develop new features. However, the only drawback of this measure is that the team is not always responsible for downing the system. When this occurs, it says nothing about the effectiveness of the team.	Literature shows that downtime hampers productivity and negatively influences the effectiveness Al-Aomar, R., Aljeneibi, S., & Almazroui, S. (2016, May). Reducing operational downtime in service processes: a six sigma case study. In 2016 International Conference on Industrial Engineering, Management Science and Application (ICIMSA) (pp. 1-5). IEEE.	

Table 57: Four expert opinions on the "Downtime" measure

Measure	Interview 1	Interview 2	Interview 3
The lead time of a feature compared to the expected delivery time of a feature.	This measure can be seen as similar to the finished user stories compared to the predicted number of user stories that need to be fulfilled in the sprint. measure. Therefore, a similar explanation will be given. The team will receive a commission whenever the planning is close to the deadlines. This means that the team meets its deadlines. The planning is either off-target or something else is wrong if there is a significant difference between the deadlines and the plans. The morale of the team and stakeholder satisfaction are impacted negatively.	It affects the component of stakeholder satisfaction in team effectiveness. It has to do with planning. Stakeholders may be less content or very satisfied when the team complies with the planning, depending on whether they are in line with it. Performance is also a factor because the team examines more than just a sprint. Anytime something takes longer, your stakeholder satisfaction will suffer.	Tells something about predictability. The team predicts what tasks need to be solved and what is delivered. It ultimately affects stakeholder satisfaction. Since targets are not met will lead to negative stakeholder satisfaction and when targets are met, the stakeholder satisfaction will be stable or increase.
Measure	**Interview 4**	**Literature**	
The lead time of a feature compared to the expected delivery time of a feature.	This also has to do with the velocity and therefore similar to the earlier mentioned velocity and the finished user stories compared to the predicted number of user stories that need to be fulfilled in the sprint measure. If the team is in line with their planning, the velocity of the team will get improved. Velocity offers information on whether the team completes more user stories and, as a result, is more productive. It could go either way, as an increase in velocity indicates that the team is more efficient and a decline indicates that the team's performance is declining. Also for this measure, the main difference is that velocity is quantified in story points and this measure provides information on the user stories.	Predictability is an important outcome of this measure. Research shows the influence of predicitablity on Upper-level and project team members. Upper-level management: Upper-level managers were particularly concerned with the organizational-level impacts of the project and view predictability as a key determinant to insuring intended outcomes. Project team members: Concerned with project duration, cost, scope and system functionality, project team members rely on predictability to achieve interim and final project targets. Achieving these team-based goals begets trust amongst the project actors. Lander, M. C., Purvis, R. L., McCray, G. E., & Leigh, W. (2004). Trust-building mechanisms utilized in outsourced IS development projects: a case study. Information & Management, 41(4), 509-528.	

Table 37: Four expert opinions and literature on "the lead time of a feature compared to the expected delivery time of a feature" measure

Measure	Interview 1	Interview 2	Interview 3
Done Work	In every sprint, a productive team delivers a specific amount of completed work items. Team member satisfaction is influenced by the completed amount of work. The group will recognize that the work has been completed. Additionally, completed work fulfills DOD standards, which results in new features. As a result, stakeholder satisfaction will also be improved.	Done work is an indicator of team effectiveness. It indicates how much work has been done (in story points). Looking at trends of done work. Done Work may positively or negatively impact the team's progress. Furthermore, it also influences stakeholder satisfaction since stakeholders are happy when features and user stories are completed.	Done work has a positive influence on team effectiveness. The team makes a direct impact on the product with has a positive effect on the team morale. Furthermore, stakeholders are also more satisfied since they see a direct result.
Measure	**Interview 4**	**Literature**	
Done Work	This measure is an useful measure of team effectiveness. This is primarily due to the fact that this measure indicates how many features or user stories have been delivered. It is an excellent team performance indicator. The completed task will also have a positive impact on team member satisfaction, which will increase the effectiveness of the team.	The completion of valued tasks, especially in a group, as at work, though this is perhaps more a cause of satisfaction. Some experiences of joy have a dimension of depth, intensity, "absorption" or "flow", for example when tackling a demanding task. Argyle, M., & Martin, M. (1991). The psychological causes of happiness. Subjective well-being: An interdisciplinary perspective, 77-100.	

Table 59: Four expert opinions and literature on the "Done Work" measure