DEPARTEMENT OF INFORMATICS

DER TECHNISCHEN UNIVERSITÄT MÜNCHEN

Master's Thesis in Information Systems

An Evaluation Framework for Artifact-based Software Process Improvement Procedures

Claudia Konopka

DEPARTEMENT OF INFORMATICS

DER TECHNISCHEN UNIVERSITÄT MÜNCHEN

Master's Thesis in Information Systems

An Evaluation Framework for Artifact-based Software Process Improvement Procedures

Entwicklung eines Evaluationsframeworks für Artefakt-basiertes Software Process Improvement

Author:	Claudia Konopka
Supervisor:	Prof. Dr. Dr. h.c. Manfred Broy
Advisor:	Dr. habil. Marco Kuhrmann
	Dr. Daniel Mendéz Fernández
Submission Date:	July 15th, 2014

I confirm that this master's thesis is my own work and I have documented all sources and material used.

Garching, July 15th 2014 Claudia Konopka

Acknowledgements

“Ich bin nicht bis hierher gekommen, weil ich davon träumte oder darüber nachdachte, sondern weil ich meinen Weg gegangen bin."
Estée Lauder

Als erstes danke ich meinem Aufgabensteller Prof. Dr. Dr. h.c. Manfred Broy, dass ich die Möglichkeit bekommen habe, mich als wissenschaftliche Hilfskraft mit dem V-Modell XT näher zu befassen und zudem dieses Thema behandeln durfte.

Ich danke meinen Betreuern Dr. habil. Marco Kuhrmann und Dr. Daniel Mendéz Fernández für die enge Zusammenarbeit und Betreuung während der Bearbeitung meiner Masterarbeit, die Unterstützung, den Input, die Reviews, das Feedback, den Spaß, die Unterhaltung, für das zur Seite stehen und eine tolle Atmosphäre während meiner Tätigkeit als Hiwi, sowie während meines Studiums.

Weiterhin danke ich allen meinen Kolleginnen und Kollegen für die schöne Zeit am Lehrstuhl. Insbesondere danke ich Silke Müller, Sebastian Eder, Henning Femmer, Jonas Eckhardt, Benedikt Hauptmann, Jakob Mund, Diego Marmsoler, Maximilian Junker und Thomas Kofler für die Unterstützung und Hilfestellungen bei Fragen zu meiner Masterarbeit, Hiwitätigkeit und meines Studiums.

Ich danke allen meinen Kommilitoninnen und Kommilitonen, allen wissenschaftlichen Mitarbeitern und Freunden, die mit mir durch die Hölle gegangen sind, mir beigestanden und geholfen haben und dafür gesorgt haben, das Studium an der TU München angenehmer zu gestalten.

Zu allerletzt danke ich meiner Familie, meinem Ehemann und besten Freund, Robert, sowie unserem größten Glück, Alexander, für die große Unterstützung und Hilfe, das zur Seite stehen und Trösten während meines Studiums. Ohne Euch hätte ich das nie geschafft.

Ich danke Euch allen von ganzem Herzen!

Eure Claudia

Zusammenfassung

Software Process Improvement (SPI) befasst sich mit der Verbesserung der Prozesse im Software Engineering und ist daher ein wissenschaftliches Gebiet, welches oft untersucht wird. Jedoch fokusiert sich die Mehrheit der SPI Forschungsarbeiten auf Erfolgsfaktoren oder menschliche Aspekte. Konkrete Vorgehensweisen, wie SPI Projekte zu initiieren und durchzuführen sind, werden jedoch kaum erforscht. Zudem ist meist unklar, wie effektiv bestimmte SPI Vorgehensweisen, im Hinblick auf deren Umsetzung, sind.

In meiner Arbeit wurde eine systematische Literaturanalyse (SLR) durchgeführt um Kriterien, die die Güte von Prozessmodellen beschreiben, zu bestimmen. Basierend auf diesen Kriterien wurde anschließend ein Bewertungsrahmen für Artefact-based Software Prozessverbesserungsmethoden entwickelt, der die Güte von Prozessmodellen bewerten soll.

Abstract

Software Process Improvement (SPI) deals with the improvement of processes in software engineering and is, for this, a research area what is often investigated. However, the majority of software process improvement (SPI) research aktivities are focused on success factors or human aspects. Concrete approaches, how to initiate and perform SPI projects are rarely researched. In addition, it is often unclear how effective certain SPI procedures are with regard to realization.

In this work, a Systematic Literature Review (SLR) was undertaken to identify criteria for evaluating integrity of process models. Based on this criteria, an evaluation framework artifact-based software process improvement procedures was developed to evaluate the integrity of process models.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

There are numerous companies in the automotive industry, electric, electronic and systems business. All of them generate more than 50 percent of their sales with software products, software systems or equipment with embedded software. Software is therefore an important factor in production and contributes fundamentally the product success [Wal07a].

Furthermore, thousands of people are employed in IT and software development. They work in projects which are enormously large and internationally distributed. This means that software is a major key and success factor for companies dealing with time-to-market, shorter lead times to gain a first-move advantage, high flexibility because of high requirements volatility, cost reduction and enabling innovations. For managing all human and organizational aspects a well defined business and development process is essential and needed [Wal07a, PW10].

“A focus on processes provides the stability and the management infrastructure required to deal with the ever-changing and competitive market.” [MMACM12].

As Münch et al. [MAKS12] quoted the “Chaos Report", many projects fail because of not achieving the given target or are challenged. Challenged projects have been indeed completed, but needed extensive additional effort and/or budget than planned. Lots of problems, occurred during the project execution, can be related to processes. For instance in the area of project management, configuration and change management, and validation and verification - all lack of indispensable processes. These problems result in risks, such as financial or liability risks, risks of losing important customers or lack of reputation that, at least, strains the organization with potential damage. In addition there is another more fundamental challenge. The lack of understanding of the relationship between processes and their effects in concrete developments, e.g. the relationship between a specific test process and the resulting reliability in the automotive domain.

According to Münch [MAKS12] et al. and Unterkalmsteiner et al. [UMG$^+$12] there is a widely accepted assumption about an existing relationship between process and product quality.

The product is visible to the customer. It is the *what*, that matters in the end. *How* this goal can be achieved is represented by the process [CGM98]. As Fuggetta [FFM88] mentioned, "[s]oftware development is a collective, complex, and creative effort. As such, the quality of a software product heavily depends on the people, organization, and procedures used to create and deliver it."

In figure 1.1 is shown that ad-hoc realization (quadrant 1) with low process quality often leads to bad product quality, prone to defects. In contrast, investing in SPI promises high process quality and mature software development processes (quadrant 3) thus lead to high product quality. Quadrant 2 shows that high product quality ,while process quality is low, can oftentimes be reached with excellent developers combined with overtime work and enormous effort. But this not beneficial for organizations. With such conditions success of developing high quality products is difficult to repeat and institutionalize. Furthermore, these organizations have to deal with high personnel fluctuations, where experts leave the company and organizational knowledge flows away with them. As Münch et al. [MAKS12] quotes, mature software development processes hence "can be seen as a prerequisite for high- quality software products." Consequently can success be repeated more easily with a good process quality and an understanding of the effects of processes in a specific development environment.

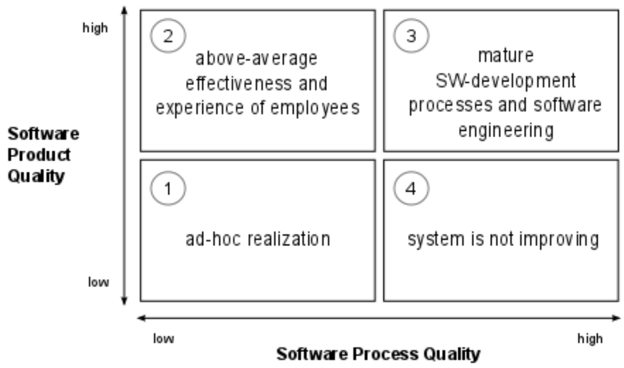


Figure 1.1: Relationship between process quality and software product quality [MAKS12].

Definition 1.1 (Software Process):
According to Humphrey [Hum89], a **software process** is the "set of activities, methods, and practices that are used in the production and evolution of software."

With improved development and efficiency activities, organizations "can develop, maintain and deliver high quality products, meeting business objectives (usually focused on quality improvement and cost and time reduction) and obtaining a higher customer [and employee] satisfaction." [MMACM12, BF] Customers, nowadays, ask not only for qualitative products, they require also quality in the services they receive. That means that companies deal not only with deploying their software development processes, but also with satisfying the ongoing demand for better IT services [MMACM12].

Jones and Bonsignour [JB] listed 121 attributes for software quality and ranked them on a scale from +10 (extremely valuable) to -10 (extremely harmful)."The rankings come from observations in about 600 companies and 13,000 projects." [Fra]. According to Fuggetta [FFM88] research payed more attention to understand and improve the quality of software in software development. This can be achieved through several techniques and approaches. Because of the underlying assumption , that there is a relationship between software quality and software process, one of the main directions in research is the study and improvement of software processes.

Definition 1.2 (Process Quality):
According to Dikici et al. "Process quality determines the appropriateness of the processes to the needs and expectations stated in the process definition." [DTD12] Quality is affected by certain quality attributes [Kan02].

In the literature we call the approach of improving the quality of processes *Software Process Improvement (SPI)*. According to Ravichandran and Rai [RR03] it "has emerged as an important paradigm for managing software development."

Definition 1.3 (Software Process Improvement):
Software Process Improvement in general is the set and repetitive activities of "changes implemented to a software process that bring about improvements." [OHK+89, SM10].

With the improvement of processes we need not only developments of structured programming languages and design methods and principles. But also definitions of software lifecycles, because they are related directly to processes. In the lifetime of software products, software lifecycles define different stages, such as requirements specification, software design, software development, testing and integration, rollout and maintenance. Furthermore, software lifecycles also define guidelines and principles according to which these stages have to be executed [FFM88, Bro12].

In the last years software life cycles, or software development models, in the field of software development changed from rigid, linear process models, for instance the waterfall model, which where split into inflexible phases to more agile and incremental ones. In agile development frameworks component based development is used. This leads to shorter development durations and cycle times between single releases.

Definition 1.4 (Software Process Model):
A "[o]ne specific embodiment of a software process architecture." [Hum89], or in short "[...] a model [(description)] of a software process." [MAKS12].

For the implementation of flexible, incremental process models we need clear, elaborated concepts and a straight forward realization. For this, experts and managers are confronted with the challenge of assessing processes, detecting weak points and finally deducing and implementing improvement measures to make sure that these processes and practices work efficiently and qualitative [Wal07a, PlGFF].

But, "[...] adopting a specific lifecycle is not enough to practically guide and control a software project." [FFM88] Processes need to be continuously assessed and improved, because they undergo refinements and changes to increase their ability of dealing with expectations and requirements of stakeholders and the consumer market.

Process assessment is an important part of process improvement. Monitoring processes can lead them to higher maturity levels. This enables a global understanding, a better vision of the workload and, finally, a more consistent and repeatable work. Higher maturity also improves technical skills of the staff and maximizes the use of technology [FFM88, MMACM12].

Definition 1.5 (Software Process Maturity):
Software Process Maturity is "The extent to which a specific process is explicitly defined, managed, measured, controlled, and effective. Maturity implies a potential for growth in capability and indicates both the richness of an organization's software process and the consistency with which it is applied in projects throughout the organization." [WGW+93]

The aim of SPI is to reduce production costs and time-to-market, increasing the product quality and reliability, employee and customer satisfaction as well as increasing productivity of software development [UMG+12, PGP07, IN06].

"Software process improvement is a complicated, systematic, and highly professional activity in software engineering that requires theory and models, skilled technical and managerial staff, as well as motivated top management commitment." [WK].

For this, improvement methods and quality models have been developed.

Definition 1.6 (Quality Model):
According to Fuggetta [FFM88], "A quality model [...] defines the requirements of an ideal company [...]." It is a reference model, that is used to assess the state of a company and the degree to which improvements have to be achieved or are already achieved. One example of a quality model is Capability Maturity Model (CMM) [CMU95].

Definition 1.7 (Improvement Method):
"An improvement method suggests the steps to be accomplished in order to improve the quality of a software process." [FFM88] it carries out how the process of improving a process. An example of an improvement method is ISO/IEC 15504 (SPICE) [FFM88, Wal07b].

Eduardo et al. [ELPC12] have already quoted the literature, that an implementation of SPI mainly occurs growth and organizational improvement within a software company. It enhances company and employee productivity of software development with lesser development time, better quality. It leads to client contentment and finally to organizational success [SM10, PGP07].

1.1 Problem Statement

After introducing software process improvement is now the question *How can the success of SPI projects be measured?*

According to Abrahamsson [Abr00] "Quality managers, change agents and researchers are often troubled in defining and demonstrating the level of success achieved in software process improvement (SPI) initiatives."

The study of Dyba [Dyb05] investigated six key factors of success in SPI, that were confirmed by Sulayman and Mendes [SM10]. These are:

1. Business orientation,
2. Involved leadership,
3. Employee participation,
4. Measurement,
5. Exploitation and
6. Exploration [Dyb05].

Business orientation is the extend to which SPI goals and activities are aligned with concrete business strategies and goals. This was identified as one of the factors with the strongest influence on the success of SPI.
Involved leadership is the extent of management commitment in SPI and has an strong impact on its success.
Also employee participation was investigated as one of the factors with the strongest influence on SPI success. It is the extent to which employees use their experiences and knowledge to decide and act and also take responsibility for SPI.
As fourth, measurement has a strong and highly significant correlation to overall success of SPI. It is the extent to which quality data is collected and utilized to assess and guide the effects of SPI initiatives.
Finally, exploitation of existing knowledge and exploration of new knowledge are two fundamentally different learning strategies. The results show, that success of SPI is positively associated with these two learning strategies. *How things are* can force to move organizations toward *how things should be* [Dyb05].

However, the issue of success is deemed to be problematic. Abrahamsson [Abr00] mentioned, there are two opinions in the research area of SPI. Some authors call for learning from failures, while others call from learning from success. According to Unterkalmsteiner [UMG+12], software measurement is a necessary part of every SPI program. For this, it is essential in the improvement of software processes and products. The problem is, that the SPI effort could be addressing the wrong issue, since if the process (or the result) is not measured. Since that, undertaking of process measures could prove to be a difficult task. Initiating even the simplest metrics programs which provides relevant and valid information to support decision-making and information for managers about the state of their software development practices, serious problems occur that faces companies. Abrahamsson states, that any direct measure of success addressing one the five process success dimension remains inadequate if other dimensions are not considered. The importance of these dimensions also vary depending on the stakeholder who evaluates it [Abr00, UMG+12, IN06].

Unterkalmsteiner et al. [UMG+12] inspected 148 research papers published between 1991 and 2008. The result is, that quality was most measured attribute with 62 percent of the contributions. Cost could be find with 41 percent, followed by schedule with 18 percent.

Also Verhoef [Ver] quantified the effects of a large SPI program. They assess three basic IT-metrics: function point size that counts the actually delivered functionality, time and costs. Their results show, when it comes to interrater reliability, the function point countering practice is state-of-the art. As a second, they investigated actual costs for assessing the success of a SPI-effort. According to Verhoef, improving costs over time is a sign of the successful implementation of a SPI-program. Furthermore, software production productivity increases when time passes. Verhoef assessed time series that compare estimates with actual values over time for measuring duration [Ver]. Another metric that can be measured is defect density or the return on investment (ROI) in SPI [HLM11, GA06, ZP].

Of course, cost reduction and a high ROI on implementing SPI projects are an interesting issues for research. But, the focus of this work is concentrated to process engineering and not economic aspects. For this we are interested in criteria for evaluating the quality of process models.

Not only technical issues of management and accuracy, such as costs, time and size have been evaluated. SPI measurement programs also involve management practices, organizational culture and policy, stakeholder influence and resource allocation [IN06].

According to Unterkalmsteiner et al. [UMG+12], there is little agreement on what should be measured. In addition, there is a lack of systematic and reliable measurement procedures. This is regarded as a reason why SPI initiatives fail so often and are in 70% unsuccessful [DTD12, IN06]. According to Abrahamsson, "there does not exist a universal framework with which the success of a project could be measured and assessed." [Abr00] Furthermore, analysis shows that these frameworks do not provide a comprehensive view from all relevant stakeholders that are involved in SPI.

For this, Iversen and Ngwenyama [IN06] suggest baed on the evidence from their longitudinal study, that the development, implementation and success of a measurement program can benefit from a comprehensive proper framework. It could help to guide managers in their work. A framework as blueprint for SPI projects can save much cost and time [IN06].

Since the beginning of the 90s the Software Engineering community has a special interest in Software Process Improvement. This fact show the growing number of articles dealing with SPI and the great number of standards for SPI, "such as CMM, CMMI, ISO/IEC 15504:2004, SPICE (ISO/IEC15504:1998), ISO/IEC 12207:2004 and ISO 9001:2000." [PGP07] and others. According to Mesquida et al. [MMACM12] Capability Maturity Model Integration (CMMI) and SPICE are similar the "two most internationally used SPI models (...)".

CMMI, here as representative example of maturity models, is represented in five stages, which are levels of process maturity for organizations. A collection of process capabilities that largely focus on supporting process areas is defined for the first time in level 2. In Level 2, the project's process is under the effective control of a project management system. Organizations with CMMI maturity level 4 and 5 are able to successfully use formal process control metrics. With these metrics, they can provide measures of the effect of their SPI projects. For this, there exist a *conceptual software process framework* used in CMMI. An organization establishes and maintains a set of software process assets, that are available for use by the *spi* projects in developing, implementing and maintaining their defined software process [WGW+93].

With every stage occur additional requirements for other increasingly sophisticated process areas to the stages before. By undergoing an appraisal and improvement service, companies can reach a maturity level by assessing their capability to operate. A, B and C are the three classes of CMMI appraisal, while class A is most resource-intensive, most costly and time-consuming appraisal. Findings about the organization's capability will be reported, but only class A can result in a publicity-reportable rating [SNJ+07, HLM11].

While CMMI is a good ability for companies to assess their processes, not all organization adopt CMMI. Staples and et al. [SNJ+07] investigated reason why organization do not adopt CMMI. The most frequent reason was the size of the company. It was too small, so it may be that CMMI seems not being profitable for them. Other reasons are that CMMI is too costly, the organization was using another SPI approach such as ISO 9001 or or the organization has no time for the time-consuming assessment. An interesting reason for us was the reason of not adopting CMMI, because of no clear benefit. [SNJ+07]

What we know from maturity models, they only assess existing processes. With the CMMI services, organization can reach higher maturity levels of their processes, but nowhere is mentioned how they can improve their processes for the next CMMI level. There is a lack of code of practice for SPI approaches. In addition, there is also no measurement how effective the SPI procedure is. Organization pay lots of money for assessing their processes and to reach the next maturity level. They adopt CMMI without considering the effectiveness of CMMI.

1.2 Contribution

The overall goal of this work is to develop a framework for evaluating software process improvement procedures.

For this, evaluation criteria and their relationships have to be investigated by conducting a SLR.

With these identified criteria and additional further expert material, we will create an evaluation framework that supports organizations in evaluating the success of their SLR approach.

The evaluation framework will exemplarily be verified by means of two variants of the V-Model XT[1]

As main task of this work, we will conduct a case study with two variants of the V-Modell XT.

1.3 Outline

Chapter 2 gives an overview of related work and fundamentals for conducting this study.

In chapter 3, we describe our research design for undertaking a systematic literature review. We used further expert material, that will be introduced here.

In chapter 4, we present the results of the systematic literature review and discuss the results found in the contributions. Then, we describe our second literature review and present the investigated criteria and the contents of the framework, evaluating these criteria.

1 because of discretion commitment we will present only the result from quantitative analysis and no model contents.

Finally in chapter 5, we set the context to the artifact model for software process improvement and management and analyze the results from the evaluation framework, that will be discussed in the end.

Chapter 6 will complete this work with a summary and statement about future work.

Chapter

2 Fundamentals

In this section, we discuss the fundamentals and work related to the thesis.

At first, I will give a brief overview of the current state of the art in research activities in the area of process engineering related to SPI criteria and procedures.

After that, I will introduce and describe a selection of research methods applied in the publications found from the SLR and which were also used to create my own study. These methods are case study (2.2.1), experiment (2.2.1), document review/analysis (2.2.2), mixed methods (2.2.3), the Goal Question Metric Paradigm (2.2.4) and survey or interview (2.2.5).

Finally, in section 2.3 I will mention in which way these methods are used for developing and conducting my own study.

Übersicht

2.1 State of the Art in SPI

In this section we will have a look on related work in SPI.

The main tasks in software process improvement is to assess how well a project did and what should be done as improvement in the future. Because of the wide range of data and because of complicated relationships between them , this is not always easy. For this, von Mayrhauser et al. [MCWO00] suggest to use production models (or Data Envelope Analysis) to analyze object variable and their impact on efficiency. The study contains an identification of relevant software production factors that have a major influence on characteristics of the produced software. Then they identified relevant variables that relate to software quality measures and efficiency of the production of software processes. Finally, transformation of resources into production outputs were modeled. They illustrate how production models can be combined with other approaches for assessing and understanding software project data. The developed approach is verified on a data set from the NASA-SEL database consisting of 46 software projects [MCWO00].

Prakash deals with the lack of formal specifications of methods. Without a formal specification, application engineers are free to interpret method concepts as they like. Accordingly, there can be a dispute about their true meaning. The researcher attempts to represent methods independently of any underlying paradigm or way-of-work. He/she couples process and product aspects of methods tightly to remove the process/product dichtonomy and merpits extensibility of methods [Pra97].

Rainer and Hall [RH01] conducted a secondary study were they analyzed 39 publications that report issues relating to SPI. The main issues are organizational stability and process expertise, but several other specific issues are identified and discussed. In their study, the researchers investigate how the specific and main issues relate to each other and how they relate to software process improvement.

One of the important methodologies is software process assessment and improvement is software engineering process benchmarking. Wang et al. conducted a survey to derive a national benchmark of software engineering practices in the software industry in Sweden. The *Swedish national benchmark* is then contrasted with the *IBM European benchmark*. In the end, a new benchmark-based *spi* approach is proposed [WK].

The publication of Berry and Jeffery "reports on the development and validation of an instrument for the collection of empirical data on the establishment and conduct of software measurement programs." [Ber00] To trial the instrument, a pilot study of 14 measurement programs was carried out [Ber00].

2.2 Methods

In general, all empirical approaches can be divided into four categories:

- *Hypothesis method.* A theory will be developed to explain a phenomenon. Then a hypothesis will be proposed and alternative variations of the hypothesis will be tested, too. Data will be collected to verify or refuse the hypothesis.
- *Treatment method.* A solution to a hypothesis will be developed and tested. Based on the results the solution will be improved until no further improvement can be done.
- *Case Study and Experiment (confirmatory methods).* As a means to validate a given hypothesis a statistical method is proposed. There may not be a theory or formal model describing the hypothesis. Just like the scientific model scientists collect data for verifying the hypothesis.
- *Analytical method.* Results derived from a developed formal theory can be compared with empirical observations [ZW98].

In the following we make a distinction between confirmatory and exploratory methods.

2.2.1 Confirmatory Methods

As confirmatory methods we distinguish between case study and experiment. By applying a case study or an experiment one or more research questions will be answered completely or partially [UMG$^+$12] where it “adds to existing knowledge by being based on previously established theory [...] or building theory." [RH08]. To evaluate attributes is a crucial part on conducting an experiment and case study. In software engineering experimentation and conducting a case stuyd can help to “[...] determine whether methods used in accordance with some theory during product development will result in software being as effective as necessary." [ZW98]. The collection and empirical analysis of data can be done on either the development process, e.g. in the case of SPI, or the product itself [ZW98].

The Process. Zelkowitz and Wallace [ZW97] developed a taxonomy for software engineering experimentation that describes 12 different experimental approaches. In the following we will see the five steps according to Wohlin et al. [WRH$^+$00]:

- “Scoping
- Planning
- Operation
- Analysis and interpretation
- Presentation and package" [WRH$^+$00].

These process steps are used to conduct the methods case study, experiment and ..., that will be described in the following.

Case study. According to Runeson and Höst [RH08] the three definitions of Benbasat et al. [BGM87], Robson [Rob02] and Yin [Yin03] agree on that a case study is of empirical nature and aims at investigating contemporary phenomena in their context. The purpose is either exploratory and used as initial investigation of some phenomena deriving new hypotheses and building theories [ESSD08], that means “finding out what is happening, seeking new insights and generating ideas and hypotheses for new research." [RH08], or it is confirmatory, used to test existing theories [ESSD08]. They offer an "in-depth understanding of how and why certain phenomena occur [...]" [ESSD08].

A Case Study (CS) can be applied as comparative research methods or as a form of manipulation. Used as comparative method, results will be compared by using one method. In the second case, results were manipulated by using another approach.

Characteristics. Mostly at the expense of the level of control, case studies are conducted in real world settings, therefore they have a high degree of realism. They are characterized by a flexible design process, also called quantitative design study. That means that key parameters of the study may be changed during the course of the study. In a case study, a project, or more precisely a certain attribute, will be monitored by researchers and qualitative data, involving words, descriptions, pictures and diagrams etc., will be collected from multiple sources over time that may characterize that attribute, derived from a specific and clearly defined goal for the project. It is a precondition for conducting a case study to formulate a clear research question used to derive a study proposition.

Often similar data from a class of projects will be collected to build a baseline. The baseline is then used to represent an organization's standard process for software development. Sometimes, a single case is sufficient for the investigation, but is then a critical (extreme or unique) case for testing a well-formulated theory [ZW97, ZW98, RH08, UMG^{+}12, ESSD08]. Study objects can be either "1) private corporations or units of public agencies developing software rather than public agencies or private corporations using software systems; [or] 2) project oriented rather than line or function oriented; and 3) the studied work is advanced engineering work conducted by highly educated people rather than routine work." [RH08].

Runeson and Höst [RH08] see multiple case studies as field studies. They usually offer a greater external validity [ESSD08]. They also notice that a case study will never provide conclusions with significance. But, in contrast, a strong and relevant conclusion will be reached by linking together "many different kinds of evidence, figures, statements [and] documents." [RH08].

Experiment. "A controlled experiment is an investigation of a testable hypothesis where one or more independent variables[, called factor, [ZW97]] are manipulated to measure their effect on one or more dependent variables." [ESSD08]. With controlled experiments, researchers are able determining "in precise terms how the variables are related and, specifically, whether a cause-effect relationship exists between them. Each combination of values of the independent variables is a treatment." [ESSD08]. An experimental unit or subject is each agent that is been studied and for what data is collected on.

"The goal of an experiment is to collect enough data from a sufficient number of subjects, all adhering to the same treatment, in order to obtain a statistically significant result on the attribute of concern, compared to some other treatment." [ZW98].

2.2.2 Exploratory Methods

Document Review/Analysis. Document review or analysis is a procedure of the grounded theory methodology. According to Glaser and Strauss [GS99] grounded theory is "the discovery of theory from data - systematically obtained and analyzed [...]."

It is a general method for text analysis and provides researchers with relevant interpretations, explanations and applications. The approach of grounded theory is used to structure raw data or for the formulistic use of analysis steps, e.g. keywords in context.

Grounded theory was developed as a qualitative method that combines two data analysis processes. It is used to convert qualitative data into a rough quantifiable form for testing a hypothesis or generating theoretical ideas with new categories, hypotheses and interrelated hypotheses.

Within the framework of data analysis, strategies like coding are used in the first process. Further strategies are open sampling, targeted sampling or discerning sampling. After the data have been coded the codes will be analyzed systematically to verify or prove a given proposition. In the second process the data will be inspected for properties of categories. For this, the researcher uses memos to track the analysis and develops theoretical ideas [Sea, BH09, WM06, CS].

Canons and Procedures. In the following, we will briefly introduce the canons and procedures of grounded theory approach.

These are:

1. The processes of data collection and analysis are interrelated. Data will be collected unless the questions or areas of observation prove to be irrelevant during analysis.
2. The basic units of analysis are concepts. Incidents, events and happenings from the data collection were labeled conceptually with codes as indicators of phenomena.
3. Concepts are grouped to categories by comparison for similarities and differences. Categories must be related to form a theory.
4. Sampling proceeds on theoretical grounds in terms of concepts, their characteristic, variations and dimensions.
5. Analysis will be done by constant comparison.
6. Looking for patterns of regularity and variations assist with integration and gives order to the data.
7. The process must be built into the theory, because of several meanings in grounded theory.
8. Writing theoretical memos is an integrated part of grounded theory. Code notes of first impressions, in form of memos, were produced during the process of coding and then related to the formulation of theory and its revision.
9. "Hypotheses about relationships among categories are developed and verified as much as possible during the research process." [CS].
10. The researcher need not work alone. Discussing with collegues helps getting new insights and increases theoretical sensitivity, as well as guarding against bias.
11. Broader structural conditions, including economic conditions, another cultural values, trends and movements must be brought into the analysis [CS].

2.2.3 Mixed Methods

At the design of mixed methods (MM) research, also called multi-method research [OR12], an "investigator collects and analyzes data, integrates the findings, and draws inferences using both [monomethod designs [TT06], which are] qualitative [or] quantitative approaches or methods[,] in a single study or program of inquiry." [CTBP07].

According to the Methods-Strands Matrix of Teddlie and Tashakkori [TT06], a typology of research designs featuring MM, there exist two types of MM designs. The one are Mixed Methods Monostrand designs. These designs involve only one strand of a research study and are, hence, the simplest ones of the MM designs. They include both, qualitative and quantitative components. With this design one type of qualitative or quantitative inference is made and only one type of data is analyzed. Because of this, Teddlie and Tashakkori labeled these designs as quasi-mixed methods.

The second type, Mixed Methods Multistrand designs, are the most complex ones of the designs. All of them contain "mixed methods and at least two research strands." [TT06]. In sum there are four of these designs, considered to be the most valuable and which are families of designs. These are Concurrent Mixed Design, Sequential Mixed Designs, Conversion Mixed Designs and Fully Integrated Mixed Designs. Based on other design criteria, e.g. priority of methodological approach, several permutations of members of these families may be. For more information on the families of designs I refer to the literature. In the following we will see an example approach for MM.

MM Research Approach. The design for mixed-methods research can be conducted according to the evolutionary research method by O'Leary and Richardson [OR12]. This approach was designed as advised by Franz et al. [Fra86]. It conducts two or more research methods in one project and uses triangulation between methods and the data to allow more plausible interpretations to emerge.

The five stages are:

1. Analysis of the current domain knowledge by conducting a literature review and expert opinion workshops.
2. Conducting a case study
3. Conducting an academic comparative analysis to develop a process framework, collecting data on and comparing it to another process.
4. Evaluating the process model by a systematic evaluation.
5. Data source triangulation of the research results with the use of multiple information sources.

In the subsections before we have seen several methods for collecting and analyzing data. Before searching for data we have to ask ourself what data we would like to analyze. In our case we are looking for metrics and procedures for SPI. Now the question is *"How to define metrics?"*

Data can already be available or either be defined for the purpose of the study. In the case where the question is what data, most suitable for the research question under investigation, have to be collected can be done with a goal-oriented measurement technique [RH08].

In the following I will introduce the Goal Question Metric (GQM) method which is a good technique to answer the question how metrics can be defined.

Definition 2.1 (Metric):
A metric, or measure, is a "variable to which a value is assigned as the result of measurement." [Wag13]

2.2.4 Goal Question Metric

The GQM Paradigm is used as top-down approach to precisely identify what metrics to measure or as bottom-up approach for analyzing, interpreting and justifying collected data. Thirdly, GQM enables also an assessment of the validity of a drawn conclusion by an explicit rationale. With the top-down approach goals were formulated at first. Based on these goals follows the refinement of questions. Finally, metrics were derived based on the questions. The bottom-up approach analyzes carefully applied software practices. Improvement goals were selected derived from the analyzes of collected data and measurements support the management of improvement activities.

The final objective of GQM is the improvement of products and processes [DHL, vSB99, Bas, RH08].

Principles for GQM-based Measurement Programs. Planned and performed measurement programs based on GQM should follow five principles.

These principles are:

- By using a detailed measurement plan analysis tasks that should be performed must be specified explicitly and precisely.
- Based on goals and questions, metrics must be derived in a top-down approach.
- Metrics must be at the basis of an explicitly documented rational. "The rational is embodied in the series of questions [...][and][...] is used for justifying data collection and for guiding data analysis and interpretation."[DHL].
- To support the interpreting of the data subject to the assumptions and limitations behind the rational, collected data must be interpreted in a bottom-up way.
- Stakeholder perspectives must be involved in the process of definition and interpretation of measurement goals [DHL].

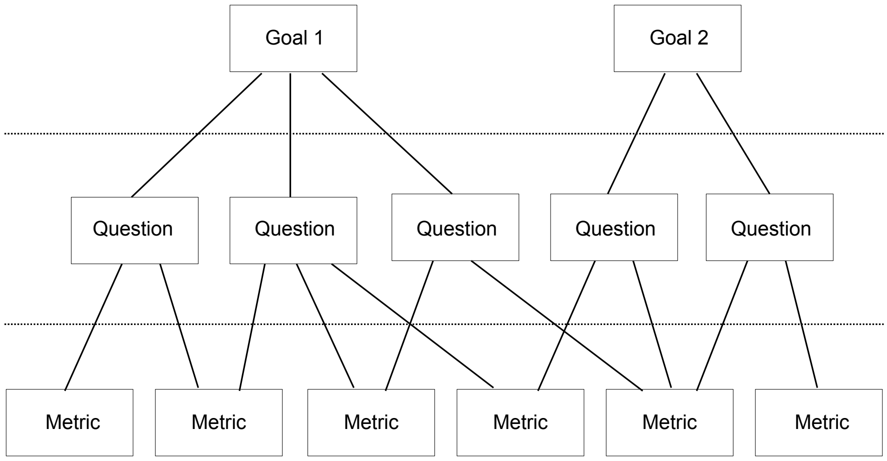


Figure 2.1: The GQM model hierarchical structure [Bas]

There are three ways of performing the GQM Paradigm. For development a GQM plan the techniques goal templates, GQM abstraction sheet and product- and process-oriented GQM plans can be used [DHL].

We used the approach described by Basili et al. [Bas] to perform GQM. The structure of this approach is shown in figure 2.1. The structure consists of three levels. The top level is the conceptual level where the goals are defined. The middle level is the operational level. Here you can find the set of questions used to characterize the way the assessment of a specific goal will be performed. Finally the lowest level contains the set of data associated with every question [Bas].

2.2.5 Survey

The characteristic of a survey is that it is an exploratory and comparative method. It is a comprehensive research method to gather standardized information about phenomena “to describe, compare, or explain their knowledge, attitudes, and behaviour." [Fin03]. Information can be collected directly by asking people in form of questionnaires, by reviewing literature or indirectly by examining documentations (written, oral or visual) of thoughts and actions of people. Data can be quantitative and/or qualitative and subjective (concerning opinions, preferences and attitudes) and/or objective (demographic information e.g. age or education), collected by conducting questionnaires or semi-structured interviews (via telephone or face-to-face). Further kinds of surveys are open-ended and closed questions, participant observation and data mining exercises.

Surveys can be supervised assigning one researcher to each respondent or administered to a group for elaborating instructions or clarifying questions. The survey can either be semi-supervised, where the researcher just explains the objective and format of the survey and then leaves the respondent alone for providing information on their own. Or the survey is unsupervised, that means that the questionnaire is mailed to selective respondents or published online with notifications in forums or social networks.

People in the case of SPI are customers of the organization or targeted employees affected by the process improvement initiative.

The purpose of a quantitative survey is to produce statistics of some aspects of the studied population. Though, generally information will by collected only from a fraction of the population, rather than asking every member of the population [KP08, Fin03, UMG+12, PK01, Rob02, Sea].

"Gathering information from customers, on the other hand, can provide insight on how the improvement affects the quality of products or services as perceived by their respective users. This can be valuable to assess external quality characteristics, such as integrity, reliability, usability, correctness, efficiency, and interoperability [39], which otherwise would be difficult to evaluate." [UMG+12].

Seven stages for Undertaking a Survey. In the following we will see the main activities needed to be considered.

These are:

- Setting specific and measurable survey objectives
- Selecting the most appropriate survey design
- Developing the survey instrument (i.e. a questionnaire or checklist)
- Preparing and evaluating the survey instrument for reliability and validity
- Obtaining valid data by administering and scoring the instrument
- Analyzing the collected data
- Reporting the results [KP08, Fin03, PK01].

2.3 Summary

In this chapter, we presented a brief overview of the current state of the art in research activities in the area of process engineering related to SPI criteria and procedures.

Afterwards, the methods of case study (2.2.1), experimentation (2.2.1), document review/-analysis (2.2.2), mixed methods (2.2.3), the Goal Question Metric Paradigm (2.2.4) and survey or interview (2.2.5), used in this study, were described.

The method of document analysis/review is used to investigate the publications from the SLR. With this approach we were able to get an overview about characteristics and procedures for evaluating the integrity of process models in the area of SPI.

As a whole work we used a mix of methods to conduct a study for SPI.

The GQM approach was made to gather metrics for an additional literature review.

In the end an evaluation framework in form of a generic questionnaire was developed to investigate the two cases of the V-Modell XT. The way in which the questionnaire was built will be described in more detail in chapter 4 in the section 4.3.

Chapter 3

Research Design

In this section, we will give basic information how to gather and structure information.

The overall research design with the procedure of data collection and analysis will be described.

For this, we will explain the steps of undertaking a systematic literature review.

Afterwards, we will explain our data analysis procedure with the harmonization process and the 2-staged voting process. Furthermore, the approach of paper coding used to analyze in-depth the collected publications will be introduced.

Finally the further expert material, used as second input for the development of an evaluation framework, will be described.

Übersicht

3.1 Overall Description

Our approach of collecting, analyzing and interpreting data is a mix of methods (2.2.3), mentioned in chapter 2.

The design of the study is structured in the same way as described by Runeson and Höst [RH08]. In the following we define our research questions and describe how we selected the case. Afterwards we describe the data collection procedure and the analyzing approach.

For the data collection a SLR is conducted according to the publications of Kitchenham [Kit] and Kitchenham et al. [KPB+09].

After the collection of all research literature, the data sets will be cleaned up and harmonized. From the refined final data set a randomly selected number of 50 papers were chosen for the analysis process. This subset then will be reviewed by the approach of document review/analysis (2.2.2). We analyzed the papers for research methods used for their author's study and attributes, investigated and evaluated by conducting these methods.

The attributes and methods found from the publications are listed and combined to categories in a separate spreadsheet. With this intermediate result we were able to built classes of areas and could map the categorized attributes to a certain area in the field of SPI, in which these attributes are settled and measured.

Than the final outcome of the SLR could be interpreted and, according to this result, a final statement can be given in 4.1.

As second input for developing the evaluation framework, further expert material was provided. This expert material, in the form of questionnaires, comes from research projects and maintenance projects, as well as student projects in lectures from the chair of *Software & Systems Engineering*.

3.2 Systematic Literature Review

For providing a SLR the research design is designed by a combination of methods described in [KPB+09].

3.2.1 Research Questions

Before starting to formulate research questions, the overall goal of the study and the problem statement in he selected research area, with the call for necessity of conducting a SLR, must be clear. In our case we aim to investigate literature, published over more than 30 years, for identifying criteria and methods used to evaluate SPI. The outcomes from the SLR aim to lay to foundation for defining a set of attributes to measure SPI in various dimensions. Of special interest are the criteria used to evaluate SPI (attributes), and the methods applied to determine those criteria. For this, the SLR aims to answer the following research questions:

To answer these research questions a SLR was conducted. The purpose of this research method is to to gather all evaluation criteria, SPI approaches and methods for evaluating these approaches that contributes SPI.

RQ 1 *Which criteria where identified to evaluate artefact-based SPI?* We investigate criteria (or KPIs) used to determine the effects of SPI.

RQ 2 *Which methods are used to determine the identified criteria?*
We analyze the methods appield to determine the criteria (values), an relate them to classes of criteria in order to work out pattern, which can be used to define proper measurement approaches for SPI projects.

3.2.2 Case Selection

In the following we describe the procedure in which way we selected the cases. Upon determining the overall topic software process improvement approaches and experiences, we held a workshop were we collected related keywords by the creativity technique *Brainstorming* [Rei07]. In the second step the collected keywords were classified to the research criteria *lifecycle*, *techniques*, *optimization*, *reference model*, *experience*, *software process model* and *software process improvement*. With this classification general search strings were created and logically combined. All general search strings are listed in table 3.1. Afterwards we combine these general search strings to our final search strings (table 3.2) which we use for our search in several literature databases.

Before using the final search strings from table 3.2 we validated them in test runs on *Google Scholar*[1]. Our expectation on the search was a collection of numerous contributions related to the search context and a considerable overhead as outcome. Afterwards, all collected data sets have to be filtered by filter queries, based on our inclusion criteria, that narrows down the problem.

3.2.3 Data Collection Procedures

As mentioned in 3.2.2, with the technique of brainstorming we collected keywords of most common terminology in a list and grouped them into categories like *lifecycle*, *techniques*, *optimization*, *reference model*, *experience*, *software process model* and *software process improvement*. Within these categories we designed general search strings containing all keywords, logically combined by conjunction and disjunction. To get more precise results form the SLR we considered different spelling of words, e.g. *life-cycle*, *lifecycle* and *life cycle*.

To get relevant publications in the context of software processes and software process improvement we combined all search strings from *S1* to *S8* with *S9* and *S10*.

In table 3.2 you can find all combinations of all general search strings for building final search strings, that were used for a full-text search in several literature databases.

The automated full text-search was undertaken in the databases *ACM Digital Library*, *SpringerLink*, *IEEE Computer Society Digital Library*, *Wiley*, *Elsevier (Science Direct)* and *IET Software*. These are established literature databases which are most appropriate in the process engineering domain, according to research experiences of the chair in the past. Each result set from the database was exported as *BibTeX* file, converted into the form of *CSV*, and was then transferred in a spreadsheet where all data sets could be filtered automatically. In the end of the data collection process, we had one spreadsheet for each database and, for this, collected results separately in accordance to the database. We had to do this, because the content information of the search results is different in each database.

1 http://scholar.google.de/

The sum of all results built the basis of the filter procedures and the in-depth analysis.

3.2.4 Definition of General and Final Queries

Based on the classification of keywords we could define ten general search strings. Table 3.1 lists all combination of the brainstormed keywords.

	Search String	**Addresses . . .**
S1	(life-cycle **or** lifecycle **or** life cycle) **and** (management **or** administration **or** development **or** description **or** authoring **or** deployment)	software process management: general life cycle
S2	(life-cycle **or** lifecycle **or** life cycle) **and** (design **or** modeling **or** modelling **or** analysis **or** training)	phases of the software process's life cycle
S3	modeling **or** modelling **or** model-based **or** approach **or** variant	process modeling
S4	optimization **or** optimisation **or** customization **or** customisation **or** tailoring	process customization and tailoring
S5	measurement **or** evaluation **or** approach **or** variant **or** improvement	general measurement ans improvement
S6	reference model **or** quality management **or** evaluation **or** assessment **or** audit **or** CMMI **or** Capability Maturity Model Integration	reference models and quality management
S7	SCAMPI **or** Standard CMMI Appraisal Method for Process Improvement **or** SPICE **or** ISO/IEC 15504 **or** PSP **or** Personal Software Process **or** TSP **or** Team Software Process	reference models and assessment approaches
S8	(feasibility **or** experience) **and** (study **or** report)	reported knowledge and empirical research
S9	software process **and** (software development model **or** process model)	*context definition:* software process
S10	SPI **or** software process improvement	*context definition:* software process improvement

Table 3.1: Search strings used as basis for our search queries.

After defining general search strings follows the combination of all search string from *S1* to *S8* with the search strings *S9* and *S10*. The result of this step are the final search strings in table 3.2.

3.2.5 Definition of Filter Queries

When conducting a full-text search we expected a considerable number of contributions within the resulting papers, that include some overhead. Because of this, there was a need of filter queries to reduce the sets by duplicates and publications that are out of scope for the research area. Hence, we designed a search tree that contain our inclusion criteria. With the left path in the tree we aim to find all contributions in the field of SPI that deal with approaches, practices or addresses the management. The right path of the tree presents another inclusion criteria for papers. Here we aimed at finding reports of experience or reports that analyze feasibility in the context of SPI.

According to our inclusion criteria, we defined filter queries shown in table 3.3. This filter queries are used to filter the abstracts of the initial result set. Unfortunately, not all databases

Combination of search strings	Final Search strings in words
S1 **and** (S9 **or** S10)	[(life-cycle **or** lifecycle **or** life cycle) **and** (management **or** administration **or** development **or** description **or** authoring **or** deployment)] **and** [(software process **and** (software development model **or** process model)) **or** (SPI **or** software process improvement)]
S2 **and** (S9 **or** S10)	[(life-cycle **or** lifecycle **or** life cycle) **and** (design **or** modeling **or** modelling **or** analysis **or** training)] **and** [(software process **and** (software development model **or** process model)) **or** (SPI **or** software process improvement)]
S3 **and** (S9 **or** S10)	(modeling **or** modelling **or** model-based **or** approach **or** variant) **and** [(software process **and** (software development model **or** process model)) **or** (SPI **or** software process improvement)]
S4 **and** (S9 **or** S10)	(optimization **or** optimisation **or** customization **or** customisation **or** tailoring) **and** [(software process **and** (software development model **or** process model)) **or** (SPI **or** software process improvement)]
S5 **and** (S9 **or** S10)	(measurement **or** evaluation **or** approach **or** variant **or** improvement) **and** [(software process **and** (software development model **or** process model)) **or** (SPI **or** software process improvement)]
S6 **and** (S9 **or** S10)	(reference model **or** quality management **or** evaluation **or** assessment **or** audit **or** CMMI **or** Capability Maturity Model Integration) **and** [(software process **and** (software development model **or** process model)) **or** (SPI **or** software process improvement)]
S7 **and** (S9 **or** S10)	(SCAMPI **or** Standard CMMI Appraisal Method for Process Improvement **or** SPICE **or** ISO/IEC 15504 **or** PSP **or** Personal Software Process **or** TSP **or** Team Software Process) **and** [(software process **and** (software development model **or** process model)) **or** (SPI **or** software process improvement)]
S8 **and** S10	[(feasibility **or** experience) **and** (study **or** report)] **and** (SPI **or** software process improvement)

Table 3.2: Final search strings for database search.

export the abstracts for their papers. For this, we were not able to filter the data sets of the databases *ACM Digital Library* and *SpringerLink* in this way. Thus, the data sets were filtered manually during the following voting process.

	Filter Query
F1	(SPI **or** software process improvement) **and** (approach **or** practice **or** management)
F2	(SPI **or** software process improvement) **and** report **and** (feasibility **or** experience)

Table 3.3: Filter queries for the harmonization process.

3.2.6 Analysis Procedures

The first step of the analysis procedure is the preparation of the initial data sets from the database search by an harmonization process. The second step is a two-staged voting process where the collected data will be prepared for the final in-depth analysis.

Harmonization. After the data collection we expected contributions that occur multiple times in the data set. Firstly, the data set has to be cleaned up by eliminating duplicates. We did this step at first within the spreadsheet of a single database and then comprehensively across all databases. As a second, all data sets that contain abstracts were filtered by the filter queries *F1* and *F2* from table 3.3. For this, the data sets were filtered again. By applying the filters, we could expect that our final result set contains papers, that address the inclusion criteria. During this process of harmonization we also removed contributions that are not from software engineering and conference papers followed by journal articles. In the end, we merged the spreadsheets of all databases and cleaned the final result set for duplicates again.

To verify the harmonization process tag clouds were created and analyzed. One tag cloud bases on the author keywords that can be find in every paper and that are exported by the data base as content information in the four data bases *IEEE Computer Society Digital Library*, *Wiley*, *Elsevier (Science Direct)* and *IET Software*. Figure 3.1 shows an example of such a tag cloud. We can see that it contains keywords related to the studied field. In conclusion, we can say that our final result set contains publication that contribute SPI.

Figure 3.1: Tag cloud of author keywords

Voting. To classify the papers as relevant or irrelevant we performed a two-staged voting process. At first, the spreadsheet was prepared for the voting. The spreadsheet contains *meta data*, an paper ID an the citation-key, and *content information* e.g. authors, paper title, abstract and year. The spreadsheet was modified by an additional column containing the attribute *relevance*. The first voting stage is an individual voting. Each of the two researchers got one column for an independent decision. The task is to identify relevant papers and vote them with binary numbers. For each paper in one row the cell in the column is filled either with one (1), if the contribution is evaluated as relevant, or with zero (0), if it is evaluated as not relevant. When both researchers agree on the contribution, it is finally marked with two (2). That means the contribution is automatically in the set of contributions for further investigation. If it is finally rated with zero (0) the contribution is excluded and eliminated from the result set. If a contribution is marked with 1, that means the researchers have to discuss the evaluation in a second voting. The criteria for the voting were on one hand, the title of the contribution and on the other hand, the abstract regarding the filter queries.

The second voting aims at considering contributions that were not finally decided in the first stage. The goal of the second stage is the agreement on contributions that are relevant for the in-depth analysis. For this, the researchers read through the abstracts and discuss the certain contribution in several workshops. After striking an agreement it will be decided whether to include the contribution in the result set.

The analysis procedure will continue with the in-depth analysis.

3.2.7 In-depth analysis

As in-depth analysis procedure we used the document review/analysis from the grounded theory, described in 2.2.2.

This procedure bases on the process of paper coding. Paper coding is a strategy for extracting values for quantitative variables from qualitative data. The process of coding transforms qualitative data into quantitative data. This is done in order to perform some type of quantitative or statistical analysis [Sea]. In our case we will investigate the papers for evaluation procedures and criteria used in SPI initiatatives. For this, we selected 50 contributions randomly from the final result set, that are evaluated as relevant.

All 50 papers were reviewed and coded. Because we were looking for methods and criteria, we used the codes *M* for methods, e.g. case study, document review, systematic literature review and so on. Criteria like reliablity, functionality, defect density etc. were coded with *C* for criteria.

In the end all methods and criteria, related to certain papers, were allocated in an additional spreadsheet. We assigned an item with 1, if it exists in the paper. After that, we mapped all criteria to the categories *performance*, *organization*, *quality*, *cost*, *context*, *model* and *product*. Also the methods were coded. Codes related to methods are shown in table 4.2.

In the end of this procedure a matrix of criteria categorized to targets and classes were matched to the methods used in the papers.

3.3 Further Expert Material

After investigating and analyzing research literature the first part for the development of an evaluation framework is done. The second part will be further expert material, in form of questionnaires, used as additional input for the development process.

All further expert material can be found in appendix B. The material comes from research projects and maintenance projects, as well as student projects in lectures from the chair of *Software & Systems Engineering*.

Questionnaire V-Modell XT Evaluation (Public Administration) was developed to investigate the application of the V-Model XT in the authorities' practice with an analysis, evaluation and improvement. The goal of this questionnaire is to gather data that give information about the application of the V-Model XT in the authorities' environment. For the development of the questionnaire the *GQM Paradigm* was used. The study considered two core issue. For the first issue, the researchers would like to know what context factors are necessary for a successful application of the V-Model XT. Secondly, what effects does appliance of the V-Model XT (VMXT) have. Based on this two main goals, questions were derived from.

The second questionnaire, V-Modell XT Evaluation (BORKOR), addresses evaluation of the application of a customized V-Modell XT. The goal of this questionnaire is to identify problems with the application of this customized V-Modell XT and determining improvement potentials. The questionnaire consists of three categories:

- Meta data according to the person and the process
- Area process model that addresses the exported documentation of the V-Modell XT
- Area product templates that addresses exported product templates, e.g for reporting

Questionnaire Artifact-based RE Improvement and KPIs concerns the evaluation of the overall artefact-based RE improvement approach (*ABREImp*), i.e. the workshops performed to define the new RE reference model of an industry partner of the chair.

The fourth, questionnaire V-Modell XT Evaluation (BNetzA), addresses also the evaluation of the application of the customized V-Modell XT BNetzA. The goal of this questionnaire is to identify problems with the application of the V-Modell XT BNetzA and determining improvement potentials. The questionnaire consists of same three categories like the second one (meta data, area process model, area product templates).

The questionnaire V-Modell XT Evaluation (Public Administration - Code Book) evaluates experiences in the handling of the V-Modell XT in the daily routine of projects.

Questionnaire number six, Artifact Model Evaluation (GloBuS), addresses also the evaluation of the application of the customized V-Modell XT BNetzA. The goal of this questionnaire is to identify problems with the application of the V-Modell XT BNetzA and determining improvement potentials. The questionnaire consists of same three categories like the second and fourth one (meta data, area process model, area product templates).

The last two questionnaires are exercise material from the lecture *Vorgehensmodelle im Software Engineering* from the winter term 2011/2012. Software Process Modeling Course (1) was the questionnaire in accordance to exercise 12 of the exercise sheet 7.1 and questionnaire Software Process Modeling Course (2) from the exercise sheet 7.2, according to the exercises 13 and 14.

3.4 Summary

In this chapter we presented the overall description of the research design. We introduced the literature review method SLR, that was used to conduct this study. We presented our research questions and described the selection of the case. Then our data collection procedures was described in detail and the databases, that we used for the literature search, were mentioned. We defined our general queries and the combination of them to final queries, used for the database search. After the data collection, it was described in which way we cleaned and filtered the result set and our filter queries were defined. Then we explained our data analysis procedure with the harmonization and voting process and how we used paper coding for our in-depth analysis of the contributions. Finally, we described what further expert material was used as additional input for the research design.

4 Chapter

Results

In this chapter we will summarize all results from the Systematic Literature Review and the further expert material in the form of questionnaires. Afterwards the main work of this thesis, an evaluation framework for artefact-based SPI procedures, will be created.

We will show and discuss that the results from the SLR produce no satisfactory results. With the SLR we hoped to be able determining quality metrics and approaches in the area of process models. To hold the output from the SLR in a manageable manner we chose a random number of papers for the further steps of the document analysis. In this selection our expectations were disappointed, because of missing statements for process models. For this, we had to rework the literature research with the snowballing approach to find acceptable results in other publications. This process and its approach will be described in detail.

Fortunately, we could find criteria for measuring process model quality. We will classify these criteria according to the type of evaluation. The criteria, that can be verified manually, will then be used to create a questionnaire.

For this, we will describe the structure of the questionnaire in detail and will constitute which question from the questionnaire addresses what criteria.

Übersicht

4.1 Systematic Literature Review

In chapter 3 the approach of a SLR used in this study was described. After the collection of data by a full-text search in several databases duplicates were eliminated and the data sets were purged by inclusion and exclusion criteria to filter out contributions that are out of scope for the investigated domain. Based on these preliminary results from a 2-staged voting process was performed to classify the papers as relevant or irrelevant for further investigation. From the results of relevant papers a randomly number of publications were selected as representative subset. Afterwards the papers were investigated through an in-depth analysis using the paper coding approach. All criteria and methods described in the subset were identified through this process and collected in a spreadsheet. All attributes found in the literature were classified to the category classes shown in table 4.1. These categories, in turn, were also clustered to classes, called targets. The overall criteria candidates and classification can be found in appendix A from A.1 to A.4.

Code	Target	Categorization Class
P	Performance	Time Effectiveness/Efficiency Productivity
Q	Quality	Number of Problems Number of TestCases Non-Functional Quality Precision Customer Stability
M	Model and/or Product	Model/Product Properties
C	Cost	Cost
CTX	Context	Personnel Knowledge Compliance Environment Awareness

Table 4.1: Classes and Categories

We also counted the number of criteria that we could found in the specific target, related to a certain method. In table 4.2 we see a matrix of targets and classes matched to the methods used in the papers from where we could find the particular criteria. This gives us a broad overview about methods used in certain areas of SPI and criteria, to be precise, classes of criteria address by these methods.

		Applied Methods											
Target	**Class**	**ADC**	**CMM**	**CS**	**DR**	**EXP**	**FM**	**FS**	**MM**	**SIM**	**SLR**	**STM**	**SUR**
P	Time	1	2	10	1	2	0	0	0	0	4	0	3
	Effectiveness/Efficiency	0	0	6	0	2	0	0	0	0	0	1	5
	Productivity	0	0	4	0	2	0	1	0	0	1	0	0
Q	Number of Problems	1	2	2	0	2	0	0	0	0	1	0	2
	Number of TestCases	0	0	0	0	1	0	0	0	0	1	0	0
	Non-Functional	0	0	2	0	1	0	1	0	0	3	0	0
	Quality	0	0	3	0	1	0	1	0	0	0	0	3
	Precision	0	0	0	0	1	0	0	0	0	0	0	1
	Customer	0	0	1	0	0	0	0	0	0	0	0	2
	Stability	0	0	1	0	0	0	0	0	0	0	0	0
M	**Model/Product**	**0**	**0**	**14**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
C	Cost	0	0	2	0	0	0	0	0	0	0	0	0
CTX	Personnel	0	0	1	0	0	0	0	0	0	3	0	1
	Knowledge	0	0	0	0	0	0	0	0	0	1	0	0
	Compliance	0	0	1	0	0	0	0	1	0	2	0	1
	Environment	0	0	6	0	0	0	0	0	0	0	0	1
	Awareness	0	0	1	0	0	0	0	0	0	0	0	0

Code	**Class**
P	Performance
Q	Quality
M	Model and/or Product
C	Cost
CTX	Context

Code	**Applied Method**
ADC	Automated Data Collection
CMM	CMM(I) Assessments
CS	(comparative) Case Study
DR	Document Review/Analysis
EXP	Experiment
FM	Financial Methods
FS	Field Study
MM	Mixed Method
SIM	Simulation

Code	**Applied Method**
SLR	Systematic Literature Review
STM	Statistical Metthods
SUR	Survey

Table 4.2: SLR Results

The most interesting column for this study is M - Model and/or Product. As we can see in table 4.2 there exist only a number of 14 criteria attributes for the applied method case study. These criteria come from the publication of Canfora et al. [CGP+06]. In their work they present results from the Framework for the Modeling and Measurement of Software Processes (FMESP) collected in a software company for development and maintenance of software for information systems. "The aim of FMESP is to provide companies with a conceptual and technological framework for the management of their process models and measurement models in an integrated way" [CGP+06]. After describing the FMESP a case study at the company Cronos Iberica is introduced. All software processes at Cronos were represented by using the Software & Systems Process Engineering Metamodel Specification (SPEM) [Obj08]. The software measurement models were defined and two work products were chosen for measurement as the most relevant, because Crono's main professional activity is the development of database applications. For this, Canfora et al. chose the conceptual and the logic data model and present them by using entity/relationship diagrams and relational database schemas. The definition of metrics was performed by an analysis based on the definition of direct metrics, defined on the domain metamodel which represents the software entity, furthermore indirect metrics, calculated by applying a measurement function on other direct or indirect metrics, and finally indicators, obtained by applying an analysis model also on other direct and/or indirect metrics. The results of this analysis is referenced in appendix A in table A.1 as well as in the tables A.2 and A.3.

As we can see from the results of the SLR our sample of 50 publications contains only one paper with 14 criteria for the target Model and/or Product. Unfortunately, the outcome of the publication of Canfora et al. [CGP+06] results in direct and indirect metrics, defined on the E/R meta model and functions calculating numbers of relationships in an E/R diagram, and indicators obtained by applying an analysis model. We cannot figure out any criteria for the integrity of process models, what was the purpose of this study.

As an outcome of the SLR we can make the statement that our representative subset generates no result of apt metrics for process models. In conclusion we make the assumption that there exist no criteria for the integrity of models. That means we found a research gap on what we can focus on now in more detail.

4.2 Further Expert Material

In this section, we will describe in which way further expert material contributes the development of the evaluation framework.

The goal of the first questionnaire,V-Modell XT Evaluation (Public Administration), is the evaluation and data acquisition that gives information about the application of the V-Model XT in the institutional practice. Because the researchers used the *GQM Paradigm* for the development of the questionnaire, we used this questionnaire as input for performing our own GQM application. It is a good example and orientation, in which way goals and questions can be formulated.

The questionnaire V-Modell XT Evaluation (BORKOR) was developed to evaluate the application of a customized V-Model XT. As mentioned in section 3.3, the questionnaire consists of three categories: *Meta data*, the *Area process model* and the *Area product templates*. The structure of his questionnaire gave us the biggest input for our evaluation framework. Our questions are formulated in accordance to this. We also use the Likert scale as evaluation type.

The *Evaluation of Artefact-based RE Improvement and Quality Indicators*, we took the evaluation of the *artefact-based RE reference model (ÒABREÓ)* as input for developing questions evaluating a process model. The questions refers to the area *Model Evaluation (absolute)*.

The structure of the questionnaire evaluating the V-Modell XT BNetzA gave an strong input for our questionnaire. The areas *Bereich Prozessmodell und Prozessdokumentation* and *Bereich Produktvorlagen* are a good basis for the development. Also the layout was recopied.

Questionnaire V-Modell XT Evaluation (Public Administration - Code Book) gave us input for the area *meta data* and for scaling types.

The questionnaire Artifact Model Evaluation (GloBuS) looks the same as the questionnaire V-Modell XT BNetzA, and was, for this, also integrated in our questionnaire.

The both exercise materials from the lecture *Vorgehensmodelle im Software Engineering*, Software Process Modeling Course (1) and Software Process Modeling Course (2) provides input for the area *tool support*.

All of the further expert material can be find as input for the questionnaire in appendix C.

4.3 Initial Synthesis and Construction

According to our results from 4.1 our sample of 50 papers from the SLR we cannot make a statement for methods and metrics in the area of process models.

In the following we will analyze the results of Canfora et al. [CGP+06] and explain why they not suffice for the development of evaluation framework for process models.

Canfora et al. present in their publication results and lessons learned in the application of the Framework for the Modelling and Measurement of Software Processes (FMESP) in a software company. The aim of this work was providing companies with a conceptual and technological framework for the management of their process models and measurement models in an integrated way.

Canfora et al. [CGP+06] evaluated the conceptual and the logical data model as a result of two small projects with regard to the application of FMESP. Then, the investigated metrics from above were calculated automatically with the tool GenMETRIC [GRCP03].

For the measurement models of E/R diagrams direct metrics (A.1), indirect metrics (A.2) and indicators (A.3) were used.

Direct metrics: number of entities, number of simple attributes, number of composite attributes, number of derived attributes, number of M:N relationships, number of 1:N relationships (including 1:1 relationships), number of binary relationships, number of N-Ary (no binary) relationships, number of IS A relationships (generalization/especialization), number of reflexive relationships

Indirect metrics: total number of attributes in an E/R diagram (considering the entities and relationships attributes), total number of relationships in an E/R diagram (without considering IS A relationships)

Indicators: entity maintenance indicator (ratio of attributes and entities of the diagram), diagram connectivity indicator (ratio of relationship and entities of the diagram), diagram inheritance indicator (ratio of inheritance relationships and total number of relationships, including inheritance), N:M relationship indicator (ratio of N:M relationships), Relationship complexity (this indicator represents the ratio of N-Ary relationships)

From the research perspective's point of view the lessons learned as a result of the utilization of FMESP allowed the researchers to confirm the importance of the following aspects.

- "*Rigorous definition of the software processes* as a prerequisite for their evaluation and improvement.
- *Effective and integrated representation of the measurement process related information* by means of a suitable metamodel based on a consistent terminology, given the diversity of process-related entities that are candidates for measurement.
- *Application of the measurement process* not in an isolated way, but *integrated in the context of evaluation and improvement of software processes* by identifying the relevant entities which are necessary to measure and registering the results obtained in a homogeneous and consistent way.
- *Suitable software tools* to automate and manage the software processes in order to promote their improvement. The FMESP application has allowed the researchers to refine the developed software prototypes" [CGP+06].

Conceptual and the logical data models are necessary to illustrate software processes and analysis of relationships between entities. With the FMESP application, the complexity of software processes can be evaluated based on an E/R diagram.

In our case, the aim of this work is to evaluate an artifact-based process model, such as the V-Model XT. The V-model XT, for instance, bases on a *XML schema*, is supported by tools and exports an process documentation. For this, direct and indirect metrics, as well as indicators does not help to develop and artifact-based evaluation framework for software process improvement procedures that should measure the structure of process models.

Because of this fact, it was necessary to conduct a literature review again in the form of the Snowballing Process 4.3. [Woh14], [Gro07]

Because of this we had to change our focus to quality metrics for process models in general.

Before creating search queries wu used the approach of GQM to define goals for further development of questions and selection of metrics.

The main goals and resultant questions for identifying metrics can be found in table 4.3.

Goal 1	*Measuring the integrity of process models*	Metric
Q1	What characteristics describe the integrity of process models?	Literature Review
Q2	In which way can integrity of process models be measured?	analysis of the methods found from the SLR
Goal 2	*Identifying improvement goals*	Metric
Q1	In which way can quality be measured?	analysis of the results found from the SLR
Q2	How can structure be measured?	

Table 4.3: The GQM Model.

Based on the goals and questions in 4.3 we derived search keywords from and performed an additional database literature search on Google[1]. Our search strings for the additional search were formulated as written in the following table 4.4.

Strings	Additional Search strings in words
A1	Güte AND Evaluierung AND Kriterien
A2	software AND metric AND model
A3	software AND metric AND quality model

Table 4.4: Additional Search Strings for database search.

The Snowballing Process

As mentioned in 4.3 there was the need of performing an additional database literature search on Google[2]. The procedure in the second literature review is similar as the SLR in the first two steps. It begins also with the phase of the case selection and data collection. The difference in this process is, at first, the brainstorming of search keywords according to the technique of GQM. Keywords were also combined to search queries for data base search.

1 www.google.de
2 www.google.de

The second difference is that the outcome of the data base search in no *BibTex*- or *CSV-file*. Because the search was done via Google[3], the publications can be reviewed separately and directly. The first review of the literature occurred online to identify publications as relevant or not. When I decided that a publication is relevant followed an more detailed review. This more detailed review of the publication contains a backward citation searching for gathering more publications, and for citing from the original source.

Then follows the analysis approach in three stages. After the data collections, an in-depth analysis of the relevant literature for criteria of integrity of process models followed in stage one. As we know from the SLR and as our result from the sample showed us, the number and content of publications with criteria for integrity of process models are rare. Because of this fact, I used criteria for software quality, that occurred from the second literature review, and adopted them in the case of process models. To get a valid result in adopting criteria of software product quality, there was the need of two consolidation steps.

For the first consolidation, all criteria were collected in a spreadsheet. In the first consolidation of stage one, all criteria were refined. The first result of this stage are 11 publications that contain 37 criteria. After the refinement of the first result, 30 criteria remained for the second stage.

Based on the literature, the 30 criteria from the first stage were clustered to aspects (areas) of evaluation in the second stage. These areas of evaluation are *model professionalism*, *model conformity*, *model validity*, *model design* and *model documentation*. Additionally, all criteria were enriched by definitions and measurements that can be found in the glossary of appendix A. Finally in the second stage follows the second consolidation. In this second consolidation all clustered criteria were classified into top-level criteria and sub-criteria. From this perspective we were able to determine by what sub-criteria a top-level criteria is expressed.

In the last stage of the second literature review, all criteria were determined as automatically and/or manually verifiable. Then, all automatically and/or manually verifiable criteria were labeled with acronyms. All criteria and acronyms can be found in table 4.5.

The acronym *CA* is for criteria that can be verified automatically, *CM* for criteria that can be verified manually and *CAM* lables criteria that can be verified in both ways, automatically and/or manually.

After the classification in top-level criteria and sub-criteria and the labeling of them, we created a network to visualize the relationships between them. Because we made a distinction between manually and automatically verifiable criteria, the network containing all criteria can be divided into two sub-networks.

Table 4.5 summarizes all characteristics found from the literature. The results address goal 1 *Measuring the integrity of process models* with question Q1 *What characteristics describe the integrity of a process model?* from the GQM model in table 4.3.

Afterwards the glossary A from appendix A extends table 4.5 with methods (see item *Input*) with which the investigated characteristics can be measured. This glossary addresses question Q2 from table table 4.3 *In which way can criteria of integrity of process models be measured?*.

The graph 4.1 contains the overall properties investigated through the secondary literature review.

After showing all criteria in one graph, follows the division in the two sub-networks. Graph 4.2 contains all criteria that can be evaluated automatically and graph 4.3 the criteria that have to be evaluated manually.

3 www.google.de

Acronym	Criteria
CA1	Ability to Achieve Consensus
CA2	Consistency
CA3	Granularity
CA4	Redundancy
CA5	Reuseability
CM1	Ambiguity
CM2	Appropriatenes
CM3	Completeness of Realization
CM4	Consistency between Model and Documentation
CM5	Expressiveness
CM6	Implementability
CM7	Learnability
CM8	Level of Abstraction
CM9	Minimality
CM10	Readability
CM11	Simplicity
CM12	Structuredness
CM13	Tailoring Adaptability
CM14	Tool Support
CM15	Understandability
CAM1	Analyzability
CAM2	Compliance
CAM3	Connectivity
CAM4	Evolution Adaptability
CAM5	Expandability
CAM6	Generalizability
CAM7	Interoperability
CAM8	Modularity
CAM9	Replaceability
CAM10	Seclusiveness
CAM11	Stability
CAM12	Validity

Table 4.5: Criteria for Evaluating the Integrity of a Process Model.

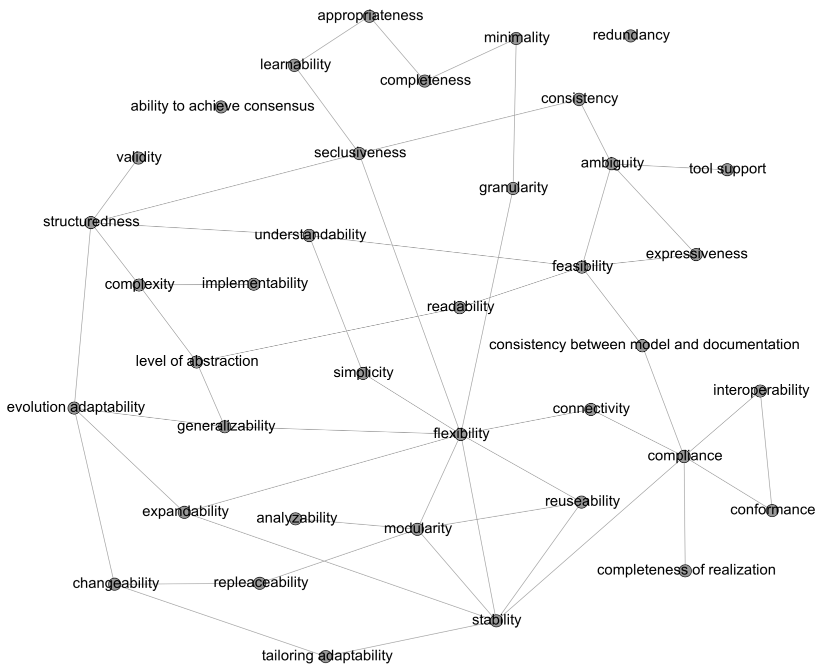


Figure 4.1: relationship between all properties

The Evaluation Framework. The evaluation framework consists of two parts. The first part will evaluate the automatically verifiable criteria. Automatically verifiable criteria need to be evaluated in different ways.

Validity can be measured by a structural validity check (e.g. XML schema validation). Consistency can be evaluated by coupling. Coupling measures how well two model components are data related, i.e. how independent they are. It is the degree of independence between modules [YC79]. Compliance can be evaluated by analytical determination, according to a conformity program, by mapping of model contents directly between models and processes at an established joint and method neutral consolidation basis for consistency reliability [KTF09].

Determination to what extent de facto standards are considered and what supported the model development measures the ability to achieve consensus. That means if the process of building consensus took place, and what functional inputs served the model development (e.g. data model, schema integration of application object models with expert consultation, identification of use cases with Joint Application Design (JAD)) [KS].

The criterion granularity can be measured by "comparing models made from 1-second versus 1-minute data. Compression of temporal data by arithmetic mean, therefore, can be an effective method for decreasing knowledge discovery processing time without compromising learning." [Ver]. Reusability can be measures by correlation of complexity/size [Fle].

Logic complexity and control structure are measured to evaluate strucuredness [Fle, Bas]. And finally, redundancy can be measured by "[...] clone detection approaches, e.g. clone count, for process-representation-based similarity on syntactic similarity." [Wag13]

To measure all criteria automatically verifiable would go beyond the scope of this master thesis. Because of this, we will only demonstrate an example of automated evaluation by a consistency check.

The second part of criteria must be evaluated manually. For this, we created a questionnaire

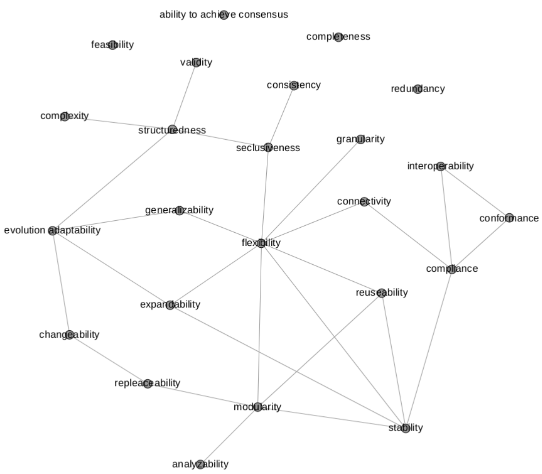


Figure 4.2: relationship between all properties convenient for automated evaluation

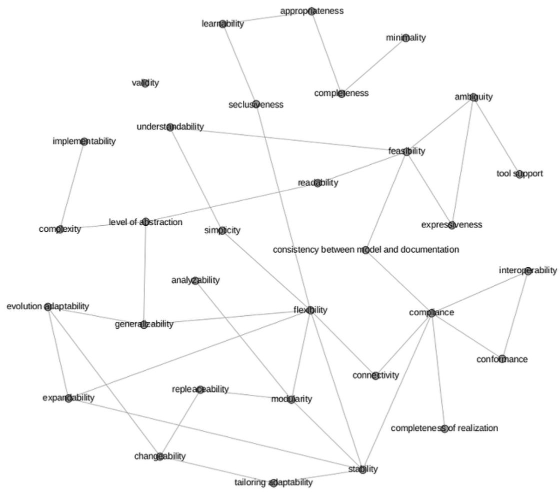


Figure 4.3: relationship between all properties convenient for manual evaluation

in which manually verifiable criteria will be measured by the evaluation of two respondents.

In the following paragraph we will see what questions address what criteria and how the questionnaire is built-on.

The Questionnaire

The questionnaire exists of two parts. The first part is a self assessment for one single process model. The second part is used as comparative assessment where a model can be compared with another one. The self assessment is subdivided into five areas. The first one is the area *Meta Data*. Here, personal information of the respondent, the process model used for the evaluation and his/her experience with the model are requested. Table 4.6 contains all questions of this area.

No.	Question	Further Information	Type
1.1	Participant (Name)		Open
1.2	What processes and version did you use?		Open
1.3	What export of the process documentation did you used?		Closed (SC)
1.4	Did you use the product templates generated from the process?		Closed (SC)
1.5	Did you get a process specific training?		Closed (SC)

Table 4.6: Area meta data.

The second area examines the absolute model evaluation. The questions from this area (table 4.7) evaluate the model in general and addresses the criteria understandability, appropriateness, readability, structuredness, level of abstraction, consistency, expressiveness and comprehensibility, connectivity in form of ability of components to communicate with others, modularity for easily reconfiguring model components and interoperability - the ability of a model to interact with other models.

The third area evaluates the process model and process documentation. The questions from table 4.8 evaluate criteria like learnability that mirrors the learning effort of the user, struc-

No.	Question	Further Information	Type
2.1	The model is complete.		Likert
2.2	The model is understandable.	CM15	Likert
2.3	Is the model concise.	CM2	Likert
2.4	The model gives a good overview about all essential elements and their relations.	CM10	Likert
2.5	The model is well structured.	CM12	Likert
2.6	The content structure is self-explaining.	CAM10	Likert
2.7	The level (degree) of abstraction of the model is appropriate.	CM8	Likert
2.8	The model content is free of contradictions.	CA2	Likert
2.9	The model allows for flexibility.		Likert
2.10	The model contains ambiguity.	CM1	Likert
2.11	The model is correct with regards to contents.	CM2	Likert
2.12	The naming of artefacts ist understandable.	CM5	Likert
2.13	Relationships between artefacts are correct with regards to contents.	CAM3	Likert
2.14	All required extensions to a given model could be completely realized.	CAM5	Likert
2.15	Modularity was considered sufficiently.	CAM8	Closed (SC)
2.16	It was clear to me how to use the interfaces, without loss of consistency of the integrated model.	CAM7	Closed (SC)
2.17	Wich parts of the model would you like to extend?	CM2	Open
2.18	Wich parts of the model would you like to reduce?	CM2	Open
2.19	Wich additional contents would you like to add to the model?	CM2	Open
2.20	Do you have further comment? (free text)		Open
2.21	What school grade would you give for the model (overall assessment)?		Closed (SC)

Table 4.7: Area Model Evaluation (absolute).

turedness of the model, minimality of structures, consistency between model and documentation, and replaceability of model components.

Area four deals with tool support. Here will be evaluated how easy the user could use the tool, with what the model is realized. The questions (table 4.9) address the criteria simplicity, tool support, tailoring adaptability, completeness of realization, implementability and validity.

The fifth area evaluates product templates. It will be evaluated how easy the user could understand and use these templates.

As mentioned before, the questionnaire contains two evaluation parts. Area one to five belong to the self assessment in part one. The second part is a comparative assessment that evaluates a model relatively to the approaches used so far. The whole area addresses the criterion evolution adaptability and question 1.10 in table 4.11 the criterion compliance of the model content to other models.

4.4 Summary

In this chapter we presented the results of the systematic literature review described in section 3.2 of chapter 3. The results from the further expert material were also presented. In section 4.3 we explained why the result of the SLR was not useful for our study and that there was a need to perform a second literature review. The preparation for this second literature review was done by the Goal Question Metric technique. We described the process of data collection, analysis and consolidation and presented the result in form of 30 criteria for integrity of process models. Then we visualized all criteria in a network of relationships between them. Finally construction of the questionnaire of the evaluation framework was described and explained what questions of the questionnaire addresses what criterion.

No.	Question	Further Information	Type
3.1	The export is clear and all essential information is easily accessible.		Likert
3.2	The export is easily accessible for persons not familiar with the process model ("outsiders").	CM7	Likert
3.3	The overall process is clearly represented and gives a good overview of all essential elements.	CM12	Likert
3.4	The export allows for straightforward inspection of releationships between process elements and the identification of consistency violations.		Likert
3.5	The scope of the process model is satisfying.	CM9	Likert
3.6	The number of the roles is appropriate.	CM2	Likert
3.7	The scope of the process artifacts is satisfying.		Likert
3.8	The level (degree) of abstraction of the process artifacts is appropriate.	CM8	Likert
3.9	Relationships between process artifacts and roles are clearly represented.		Likert
3.10	The scope of the activities (tasks) is satisfying.		Likert
3.11	The level (degree) of abstraction of the activities (tasks) is appropriate.		Likert
3.12	The process documentation is consistent with the process model.	CM4	Likert
3.13	Replaceable process parts could be properly designed and realized.	CAM9	Likert
3.14	The process model is a meaningful representation of the target domain and fulfills the functional requirement.		Likert
3.15	Wich parts of the process documentation would you like to extend?	CM2	Closed (SC)
3.16	Wich parts of the process documentation would you like to reduce?	CM2	Likert
3.17	Wich additional contents would you like to add to the process model?	CM2	Likert
3.18	What actions would recommend to meet these demands?		Likert
3.19	What school grade would you give for the process model (overall assessment)?		Closed (SC)

Table 4.8: Area Process Model and Process Documentation.

No.	Question	Further Information	Type
4.1	The used tool is intuitively applicable and the approach of process modeling is straightforward.	CM11	Likert
4.2	I was immediately familiar with the tool and could instantly use all necessary functionality for the realization of the designed process elements.	CM14	Likert
4.3	The tool was useful to realize the designed model.	CM14	Likert
4.4	If not, why?		Open
4.5	The used tool was useful to realize the model with little overhead.	CM14	Likert
4.6	I was aware about the consequences of the taken realization decision at any time.		Likert
4.7	All required tailoring profiles could be completely realized.	CM13	Likert
4.8	The process documentation could be created for the respective context.		Likert
4.9	I could completely realize all designed elements of the modeled overall process.	CM3	Likert
4.10	I could realize all roles.	CM3	Likert
4.11	I could realize all artefacts.	CM3	Likert
4.12	I could realize all realationsships between artefacts and roles.	CM3	Likert
4.13	I could realize all activities (tasks).	CM3	Likert
4.14	According to the designs in the realization concept, I could completely realize the overall process.	CM6	Likert
4.15	I could export the model at any time.		Likert
4.16	The used tool was straightforward and all requirements could be realized.	CM14	Likert
4.17	I could identify/easily check consitency problems at any time.	CAM1	Likert
4.18	If I had the choice I would coose the same tool.		Closed (SC)
4.19	Why?		Open
4.20	I could realize all process requirements.		Likert
4.21	What process requirements were not completely realized?		Open
4.22	What would be necessary to realize missing/incomplete requirements?		Open

Table 4.9: Area Tool Support.

No.	Question	Further Information	Type
5.1	All product templates are completely realized.		Likert
5.2	The product templates are clear and all essential information is easily accessible.		Likert
5.3	The product templates are easily accessible for persons not familiar with the process model ("outsiders").		Likert
5.4	The product templates allow for direct inspection of releationships to further products and the identification of consistency violations.		Likert
5.5	Which parts of the product templates would you like to extend?	CM2	Open
5.6	Which parts of the product templates would you like to reduce?	CM2	Open
5.7	Which additional contents from the process would you like to add to product templates?	CM2	Open
5.8	What would be necessary to realize missing/incomplete requirements?		Open
5.9	What school grade would you give for the product templates (overall assessment)?		Closed (SC)

Table 4.10: Area Product Templates.

No.	Question	Further Information	Type
1.1	The model simplyfies the communication between stakeholders (project partners and contracting party).		Likert
1.2	The model simplyfies the collaboration between stakeholders (project partners and contracting party).		Likert
1.3	The model simplyfies the data exchange between stakeholders (project partners and contracting party).		Likert
1.4	The model is more felxible than approaches used so far.		Likert
1.5	The model is better structured than approaches used so far.		Likert
1.6	The model is better suitable for tool support than approaches used so far.		Likert
1.7	The model is more complete than approaches used so far.		Likert
1.8	The level of abstraction in this model is higher than in approaches used so far.		Likert
1.9	Taken as a whole, the level of abstraction in this model has advantages compared to approaches used so far.		Likert
1.10	The used terminology is more consistent than in approaches used so far.	CAM2	Likert
1.11	The level of abstraction of relationships is higher than in approaches used so far.		Likert
1.12	Taken as a whole, the level of abstraction of relationships has advantages compared to approaches used so far.		Likert
1.13	What school grade would you give for the model in comparison with approaches used so far (overall assessment)?		Closed (SC)

Table 4.11: Area Model Evaluation (relatively to the approaches used so far - comparative evaluation only).

Chapter

5 Application and Validation

In this section we will set the context to ArSPI, an artifact model for software process improvement and management. For this, we map criteria from chapter 4 to certain process artifacts of the model.

Based on this results, we will refine the evaluation framework. We will discuss what abilities a process engineer has to evaluate processes in the context of SPI.

Then, we explain the validation and analysis procedure for the artifact-based evaluation framework and present our results. Finally, we will discuss the results and will give the final statement.

Übersicht

5.1 Introduction

Software processes are very complex today, because projects are done in different application domains. They are distributed globally and beyond company boarders and involve people with different cultures. With an increasing complexity, there is an increasing demand for structuring and implementing software processes in a systematic manner. To satisfy this demand, software process models provide a blueprint of relevant artifacts, roles and activities. They also support entities which are necessary to implement a particular software process [KDK13].

According to Kuhrmann et al. [KDK13] there are two major strategies to design a software process. The first one is a *acitivity-based* strategy, that is concentrated on activities. Activities and methods will be analyzed and the software process will be defined on the basis of the behavior of project teams. The second one is an *artifact-oriented* strategy that is concentrated on artifacts.

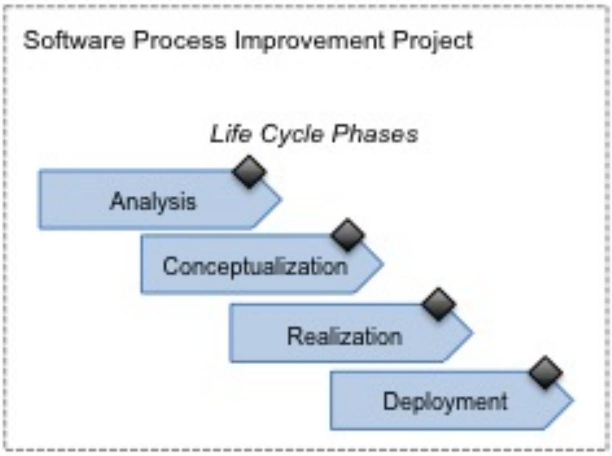
In this work, we focus on artifact-based process models, like the V-Model XT. Because such process models are oriented on artifacts and results, with an artifact-based process models all stakeholders of a project know at any time, *who* has *what* do and *when* [Bro12, Xt]. For this, the ArSPI was developed as equivalence for software process improvement projects.

5.2 Setting the Context

A software process improvement initiatative is a project in which a process or a specific process asset is improved or developed. Like in a software project, a SPI project realizes certain process requirements.

A SPI project is also, like a software project, divided into phases. In figure we see an overview of the respective life cycle of a SPI initiative [Kuh13].

Figure 5.1: SPI life cycle phases [Kuh13].



Analysis. In the analysis phase the actual process with its process documentation, that shall be improved, is analyzed to determine the current state of practice. Process goals and process requirements have to be determined at the end of this phase.

Conceptualisation. In this phase, a software process, to be improved, is designed by descriptive process modeling techniques. A first prototype can be created, depending on the selected process development environment. Drafts of the process will be created without concrete technical implementations.

Realization. In the realization phase a creation approach will be defined. It contains a selection of concrete modeling environments and tools. It can also contain a realization plan, "which is tailored in order to implement, deploy and test the process using [a] selected technique." [Kuh13] Moreover, the requirements and designs are refined for implementation. In addition, a realization and quality assurance plan is created.

Deployment In this phase, a certain process release or a specific change of an existing process variant will be rolled out. Furthermore, an process evaluation of the process by using e.g. process assessment is recommended.

All phases recur in several iterations, including feedback loops [Kuh13].

"From the perspective of a *process engineer (PE)*, the actual process is the one that reflects the current style of working in a project [...]." [Kuh13]. The PE organizes and manages a particular SPI project.

During the conception phase the process engineering team have to get familiar with the potentially new process engineering framework and the corresponding tools. This contains a first considerations regarding the process infrastructure. As the whole process and its environment will be designed. This usually requires a fully instantiated artifact model.

If a variant will be created in a software process line (SPL), the process engineer has to figure out which operations are available and ÒallowedÓ during the realization phase. Usually, SPLs have certain conformance constraints. For this a *SPL-Delta report* or several *assessments* have to be created.

In the deployment phase, the *PE* decides which export of the documentation has to be provided, selects the required tools and their version etc. for the *process release*. In addition, he/she crosschecks the respective requirements.

The PE checks whether the artifacts are complete after the artifact modelÕs instantiation. He/she should also track the project flow where quality issues can occur [Kuh13].

In which way created artifacts can be evaluated by a PE will be described in the following section.

5.3 Refining the Evaluation Framework

As we have seen in 5.2, a *process engineer* is responsible for the organization and management of a particular SPI project. He/she evaluates artifacts by *assessments*, such as *process requirements*, developed during the process life cycle. Furthermore, the PE creates the *SPL-Delta report* in the case of variant development in a *software process line*. In the end, he/she check whether the process release contains all documentations, tools and training material [Kuh13].

An evaluation framework would help the PE performing his/her tasks. With an evaluation framework, the PE would have guideline for the evaluation of developed process artifacts, that leads him/her through the whole process life cycle.

The ArSPI model is also concentrated on artifacts, that should be produced in every SPI project. The five key artifacts of ArSPI are:

- Process Requirements
- Conceptual Process Design
- Technical Process Design
- Process Life Cycle Support
- Process Release [Kuh13].

Process requirements contain all requirements of the process to be improved or developed. They influence the "processÕs conception, the technical design, and also the planning-related tasks, e.g. delivery, measurement, or training." [Kuh13] In the *conceptual process design*, a conceptual non-technical design will be created based on the all non-technical requirements. All technical requirements will be translated into the *technical process design*. The *process life cycle support* contains all definitions and descriptions of processes to support, e.g. the process development, deployment or training etc. Finally, the outcome of a SPI project is the *process release*.

According to the ArSPI model there are also complementing artifacts, e.g. artifacts for (general) project management and quality management. They contribute to a SPI project and to an overall software process management. Quality assurance tasks assign all artifacts and process assets to a *test specification* and *test protocol*. Test specifications and protocols that will be created are defined in the *quality assurance manual (QA Manual)*. The emphQA Manual contain quality assurance methods and defines the criteria for a good process quality, including definitions, metrics, templates etc. In addition, scheduled measures for the respective artifacts are defined in *quality assurance plan*.

Quality management processes are based on QA-related artifacts. These are:

- Process Assessment: addresses conformance and compliance and refers to the measurement and evaluation strategy in the process life cycle support documentation
- Conformity Assessment Report: refers to the measurement and evaluation strategy in the process life cycle support documentation
- SPL-Delta Report is declared in the *technical process design*
- Metric Catalog (incl. Metric): includes the definition of metrics and the measurement in the projects

In figure 5.2, we see an overview of the SPI life cycle with the five key artifacts, as well as the evaluation artifacts of the *quality assurance*. Based on our *glossary* A we can map the artifacts of the ArSPI model to the criteria.

Process requirements have to be valid, consistent, unambiguous and free of redundancy etc. Consistency can be verified automatically and validity automatically and manually. Redundancy must be verified automatically, while ambiguity must be verified manually.

In addition to the *process requirements*, a quality assurance task is to develop a *metric catalog*. The *metric catalog* contains initial definitions, which metrics and KPIs are of interest to be measured with regard to project- and process performance. For this, performance will be verified automatically.

Process life cycle support contains all definitions and descriptions of processes. For this, it has to fulfil the same criteria like the *process requirements*.

As part of *QA* the *process assessment* are evaluation activities regarding, e.g. process performance, and in special cases, compliance or conformance are proven to customer requirements or certain standards. Compliance can be verified manually and/or automatically. Conformance, as top-level criteria, is expressed by interoperability and compliance. Interoperability can also be verified manually and/or automatically.

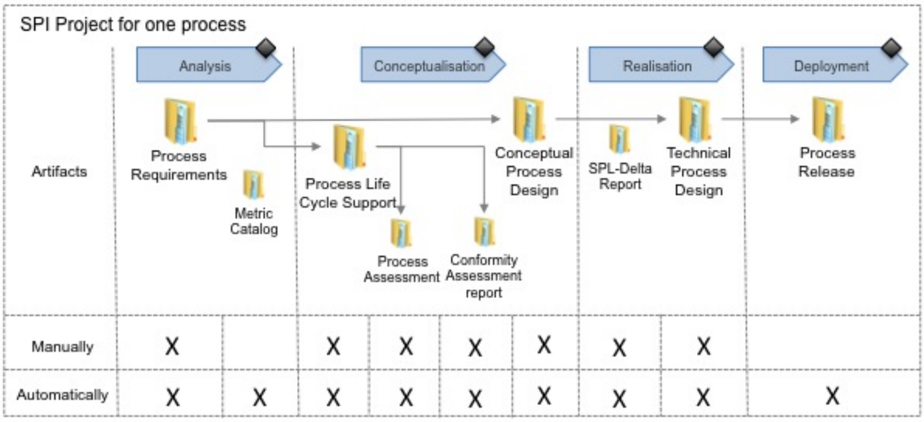
Furthermore, referring to the *measurement and evaluation strategy*, the *process assessment*, that addresses conformance or compliance, is conducted and documented in a *conformity assessment report*. For this, the conduction of the assessment can be done manually or automatically.

Criteria that addresses the the *conceptual process design* and *technical process design* are flexibility, granularity, stability, minimality, expandability, simplicity and so forth. Granularity has to be verified automatically, minimality manually, expandability can be verified automatically and/or manually and simplicity has to verified manually. Flexibility is the top-level criteria and is expressed though the other sub-level criteria. For this, it will be verified by them.

Conformity of a process variant will be checked in a SPL-based setting and explicitly document changes will be reported in a *SPL-Delta Report*. A conformity check, can be done manually and/or automatically.

The *Process Release* contains all relevant process assets and tools, as well as the training material. Everything has to be collected and appropriately configured [Kuh13]. For this, the process release has to be complete, containing all necessary sub-artifacts. Completeness is expressed by appropriateness and minimality, that both have to be verified manually.

Figure 5.2: Evaluation of artifacts based on the ArSPI model [Kuh13].



5.4 Validation

In this section we will describe the procedure with which we analyzed the results of the consistency check and the results of the questionnaire. We will see the results from the questionnaire as net chart and bar chart. In the end, we will discuss the results and give a final statement for the evaluated quality of each process model.

5.4.1 Case Description

We chose two process models for the appliance of the artifact-based evaluation framework, that was developed in this work. One is the *V-Modell XT BNetzA 1.0*. This VMXT is an adapted version, especially designed for the needs of the *Bundesnetzagentur*. It is an evolution of the V-Model XT BNetzA 0.9. The *VMXT BNetzA* is a merge of the V-Model XT 1.3 from the year 2009 and the V-Model XT Bund 1.0 form the year 2010, that are used as basis for the development of the adaption.

The second one is the *V-Modell XT ZIVIT version 1.0*. This VMXT is also a is a merge of the V-Model XT 1.3 and the V-Model XT Bund 1.0, developed to satisfy the demand of the *Zentrum für Informationsverarbeitung und Informationstechnik (ZIVIT)*.

These two adaptions of the VMXT are used to perform the evaluation framework.

At first we demonstrate a consistency check of the two models as reference for an automated evaluation. The V-Model XT offers the function *"Konsistenz prüfen"* in the *V-Model XT Editor*. For this, the consistency check could easily be done.

Then, we apply the questionnaire for the manually evaluation. The questionnaire was responded by two experts. For each VMXT, one questionnaire was responded. So in the end, we had a sum of 4 answered questionnaires, two as representation of the *V-Modell XT BNetzA* and two as representation of the *V-Modell XT ZIVIT*.

In the following we will introduce our analysis procedure.

5.4.2 Analysis Procedure

As we can see in 4.3, we use open and closed questions, including selective choice as special variety of closed questions. Open questions are questions that are used to get more information about the respondent him-/herself, his/her practice or experiences. The respondent replies in a free text form in certain text field. These answers cannot be analyzed systematically, they have to be reviewed manually. For this, closed questions are easier to analyze and were shown up in the questionnaire as majority. The closed questions can be replied by a Likert scale or as selective choice, e.g. average grade or decision for *"yes"*, *"no"* or *"not sure"*. The Likert scale ranged from *"strongly agree"* to *"strongly disagree"*. In sum, there are eight shades of opinion, that proceed from agreement to disagreement.

For the systematically analysis of the results we wanted to know the answer of the respondents in an overall result for each closed question. All closed questions have to be answered with one cross ("X") in the allocated cell of opinion or choice. It was possible to leave the row out, in the case the respondent cannot answer a question.

In the analysis approach, we compared the answers for one and the same model. We counted all questions that were answered by both experts, irrespective if some were left out. For each cell were a cross was made, we marked the answer in an result row with one

(“1"), and zero (“0") if there was no cross for this opinion shade. Then we calculated the sum of both respondents for every opinion and choice. For this we were able to analyze the overall result of this question and could built net charts and bar charts.

A net charts and bar charts can be created to analyze the results of closed questions in form of Likert scale. For creating these charts it is necessary to map numbers from zero to seven to the eight shades. The mapping can be seen in table 5.1.

Strongly agree	**Agree more**	**Agree**	**Somewhat Agree**	**Somewhat Disagree**	**Disagree**	**Disagree more**	**Strongly disagree**
7	6	5	4	3	2	1	0

Table 5.1: Mapping to seven shades.

Finally, we built also an overall result of the manually verifiable criteria, that were addressed by certain questions. If an criterion occurred multiple times in the questionnaire we calculated the median and took this as average for an overall answer.

5.4.3 Results

We analyzed every area in the questionnaire, except the area *Model Evaluation (relatively to the approaches used so far)*, because this is an comparative assessment. All results can be viewed in the following.

Automatically Verifiable Criteria

After processing the consistency check for the V-Model XT BNetzA we found 245 conflicts in sum. In comparison, we found 15 conflicts in sum in the V-Model XT ZIVIT. All conflicts are listed in table 5.2. The attribute *Criterion* consists of the rule and the conflict.
As we can result from table 5.2 none of the both models are consistent. The *VMXT BNetzA* lack two times of missing cardinalities and 124 times of missing values for a certain attribute. In 118 cases, there is a reference to an id, but the id does not exist. In one case is a link broken. The *VMXT BNetzA* suffers also from missing values in all of 15 consistency violations.

In summary, the automatic checks for consistency revealed several consistency violations. However, of these, no critical violation in terms of being able to export and deploy a process variant could be identified. Nevertheless, the issues found in this analysis point to improvement potential. For this, the automatic consistency check provides process engineers with information about the (technical) model quality, or the quality of the realization of the model respectively.
Further findings from the analysis are: The majority of the consistency violations are caused by the reference model. The automatic analysis thus allows for also identifying the source within the model tree causing problems. LIkewise, improvement measures can be scoped accordingly. As a second observation we found the used analysis alogorithm making no difference between critical and non-critical violations. Hence, process engineers have to walk through the results, and have to judge every item individually. From this, we can also derive an improvement proposal: The analysis algorithm should support a classification of found issues.

Model	No.	Criterion
V-Model XT BNetzA	1	CardinalityCheck, Kardinalität des Links version nicht im Intervall [1,1].
	2	CardinalityCheck, Kardinalitä des Links id nicht im Intervall [1,1]
	3 - 80	EmptyValueCheck, Kein Wert für Attribut version.
	81	EmptyValueCheck, Kein Wert für Attribut Version_intern.
	82 - 126	EmptyValueCheck, Kein Wert für Attribut version.
	127 - 234	InternalRefIdCheck, Id # existiert nicht - es wird aber darauf verwiesen.
	235	InternalLinkCheck, Link “SW-Einheiten | oder" endet vor Wortgrenze.
	236 - 245	InternalRefIdCheck, Id # existiert nicht - es wird aber darauf verwiesen.
V-Model XT ZIVIT	1	EmptyValueCheck, Kein Wert für Attribut Begründung.
	2	EmptyValueCheck, Kein Wert für Attribut VModellExportVerzeichnis.
	3 - 6	EmptyValueCheck, Kein Wert für Attribut Begründung.
	7	EmptyValueCheck, Kein Wert für Attribut VorlagenExportVerzeichnis.
	8 - 10	EmptyValueCheck, Kein Wert für Attribut Begründung.
	11	EmptyValueCheck, Kein Wert für Attribut PlanExportVerzeichnis.
	12 - 13	EmptyValueCheck, Kein Wert für Attribut Begründung.
	14	EmptyValueCheck, Kein Wert für Attribut PTVBegründung.
	15	EmptyValueCheck, Kein Wert für Attribut VBBegründung.

Table 5.2: Consistency check.

Manually Verifiable Criteria

In this paragraph we see the results from the questionnaire in form of net charts and bar charts. At first we look at the results from the V-Modell XT BNetzA, that we can see from figure 5.3 to figure 5.12. Afterwards follow the results of the V-Modell XT ZIVIT from figure 5.13 to figure 5.22

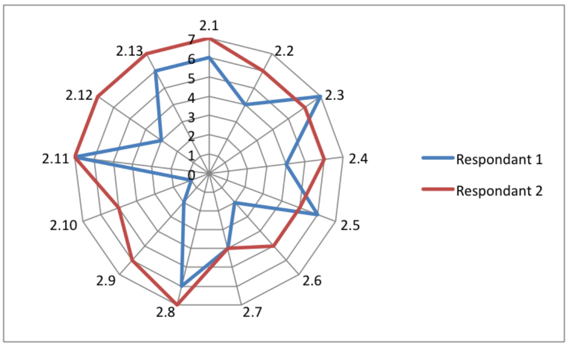


Figure 5.3: BNetzA - net chart of the area model evaluation (absolute)

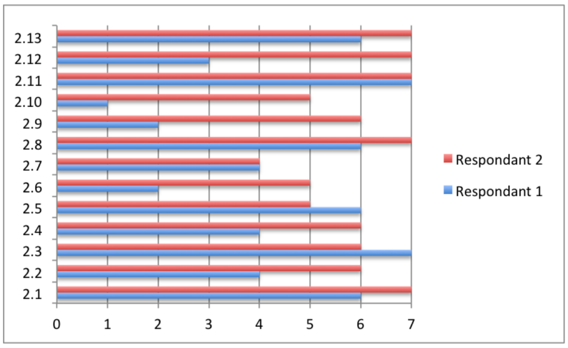


Figure 5.4: BNetzA - bar chart of the area model evaluation (absolute)

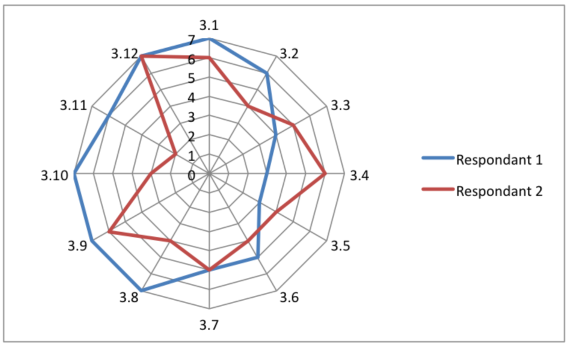


Figure 5.5: BNetzA - net chart of the area process model and documentation

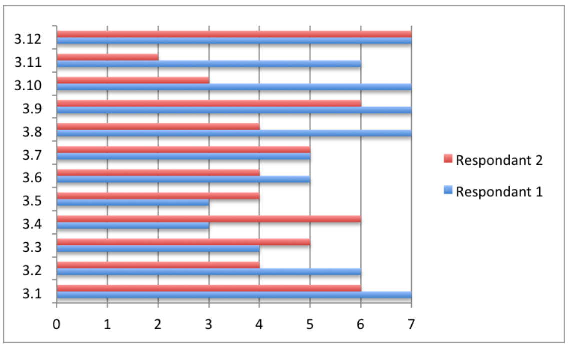


Figure 5.6: BNetzA - bar chart of the area process model and documentation

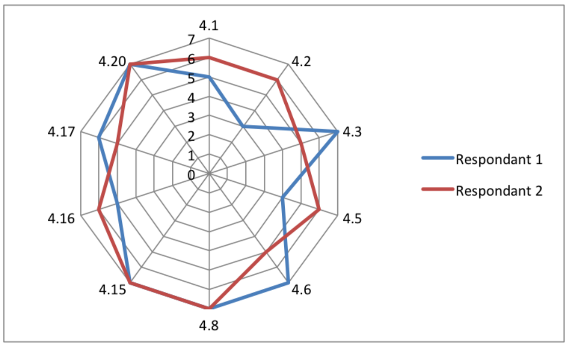


Figure 5.7: BNetzA - net chart of the area tool support

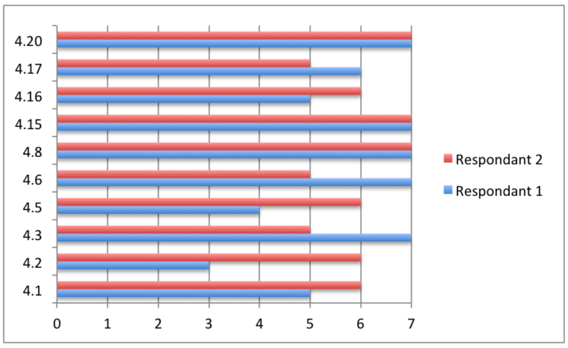


Figure 5.8: BNetzA - bar chart of the area tool support

5.4.4 Discussion

BNetzA - Area Meta Data. Both respondents used the *V-Modell XT BNetzA* version 1.0 with the export of the process documentation as *PDF-file*. Both got no specific training for

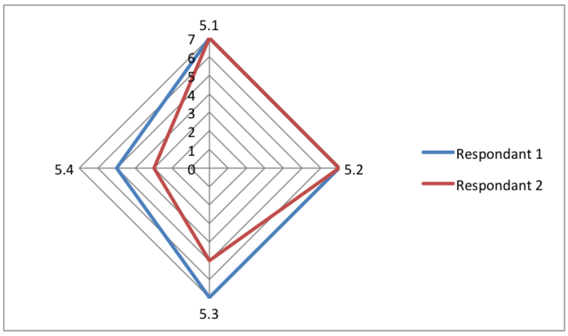


Figure 5.9: BNetzA - net chart of the area product templates

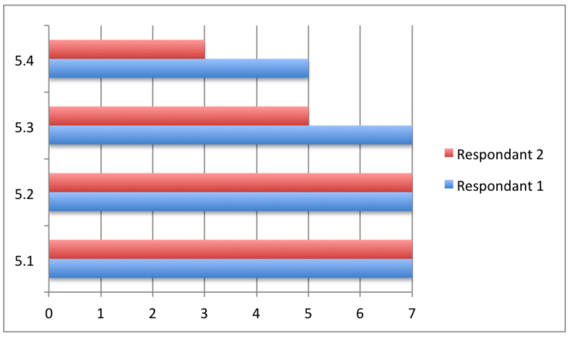


Figure 5.10: BNetzA - bar chart of the area product templates

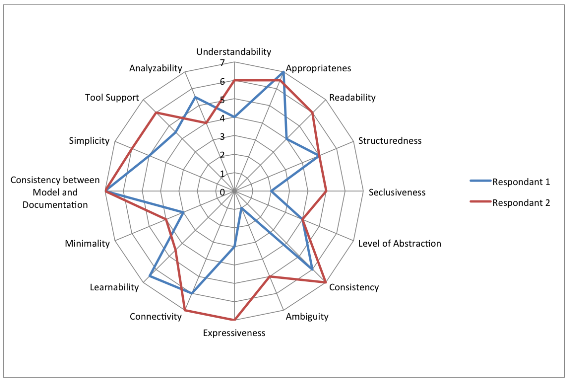


Figure 5.11: BNetzA - net chart of all manually criteria

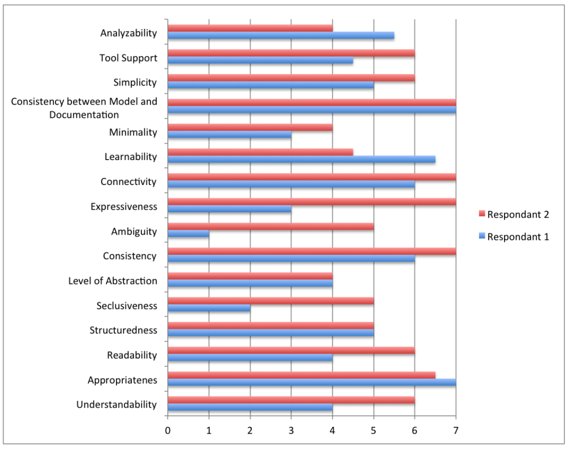


Figure 5.12: BNetzA - bar chart of all manually criteria

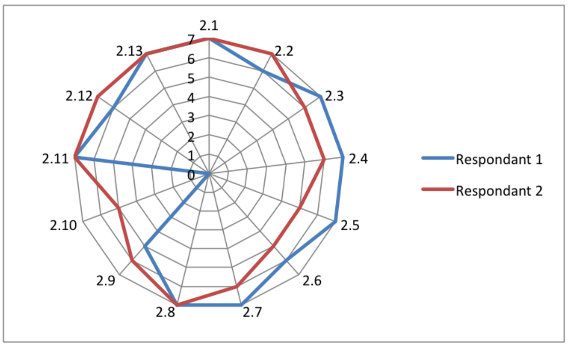


Figure 5.13: ZIVIT - net chart of the area model evaluation (absolute)

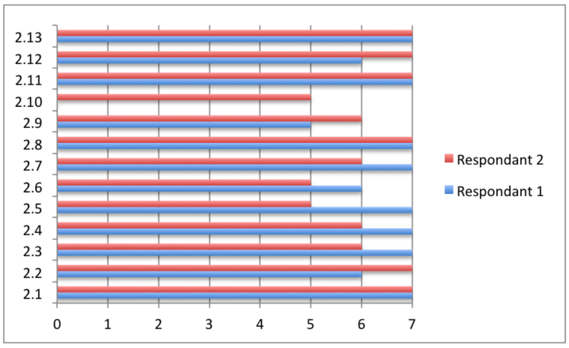


Figure 5.14: ZIVIT - bar chart of the area model evaluation (absolute)

the process model and used the product templates generated from the process.

BNetzA - Area Model Evaluation (absolute). In the area of model evaluation (absolute) (4.7) the respondents evaluated the *V-Modell XT BNetzA* different.

As we can see in the bar chart 5.4, there are wide difference for the questions *2.2*, *2.4*, *2.6*, *2.9*, *2.10* and *2.12*. While respondent 2 agreed strong, the respondent 1 agreed not completely. The widest difference is question *2.10*. Respondent 1 answered that the model contains no

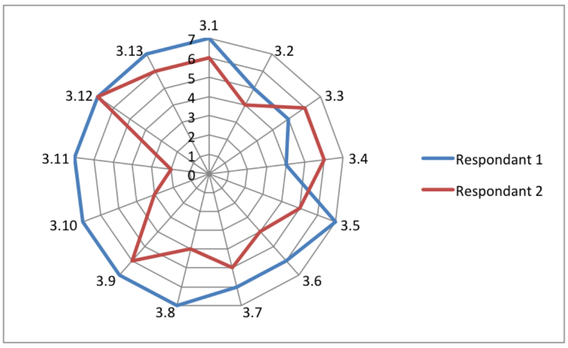


Figure 5.15: ZIVIT - net chart of the area process model and documentation

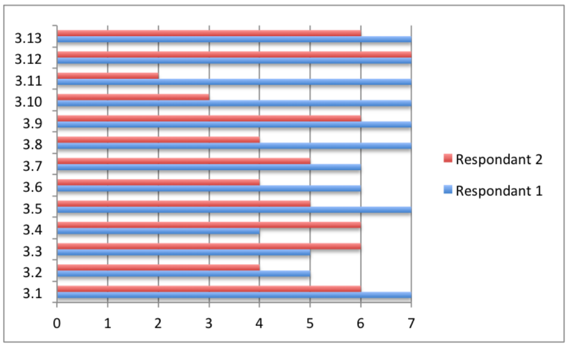


Figure 5.16: ZIVIT - bar chart of the area process model and documentation

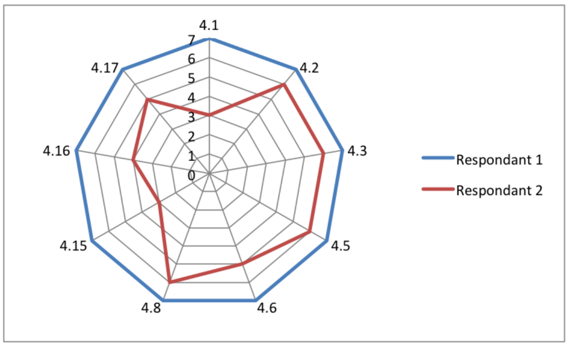


Figure 5.17: ZIVIT - net chart of the area tool support

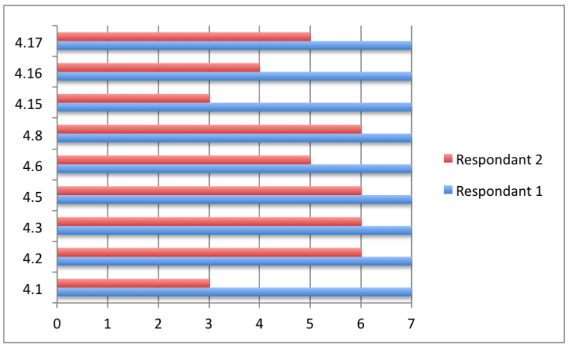


Figure 5.18: ZIVIT - bar chart of the area tool support

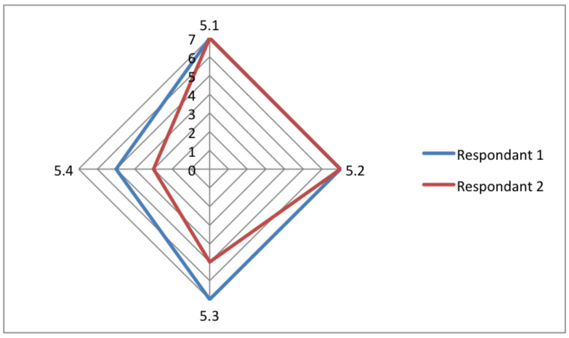


Figure 5.19: ZIVIT - net chart of the area product templates

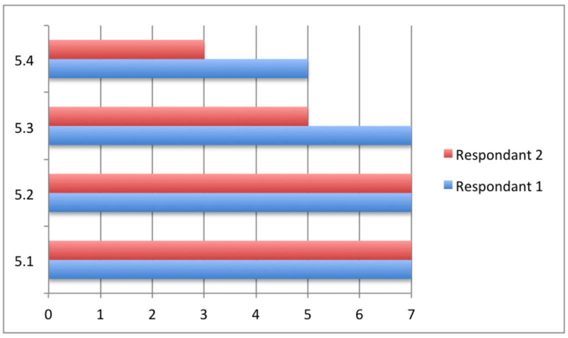


Figure 5.20: ZIVIT - bar chart of the area product templates

ambiguity, while respondent 2 agreed that the “model contains ambiguity."

Respondent 2 would like to extend nothing, while respondent 1 would like to extend the parts “Methods and method descriptions" (question 2.17). Respondent 2 would like also nothing to reduce, while respondent 2 would like to reduce (question 2.18) “Descriptive text from the reference process that does not contribute to the company-specific variant" . Both respondents would like to add no additional contents to the model (question 2.19) and have no further comments (question 2.20).

According to the wide difference of the responses we can see different satisfactory levels with the model as a whole. The answers of respondent 1 are visualized like a star, while the answers of respondent 2 are balanced (5.3). While Respondent 1 agrees somewhat (median

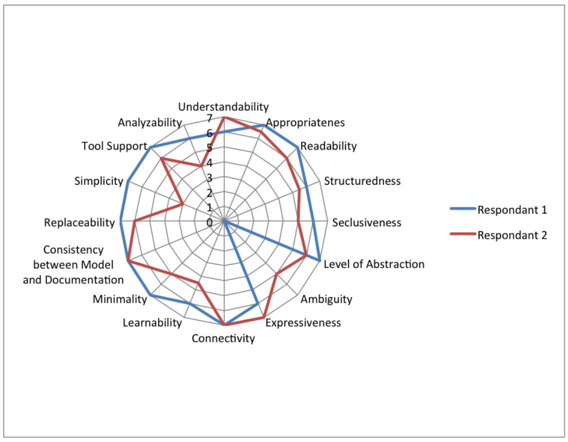


Figure 5.21: ZIVIT - net chart of all manually criteria

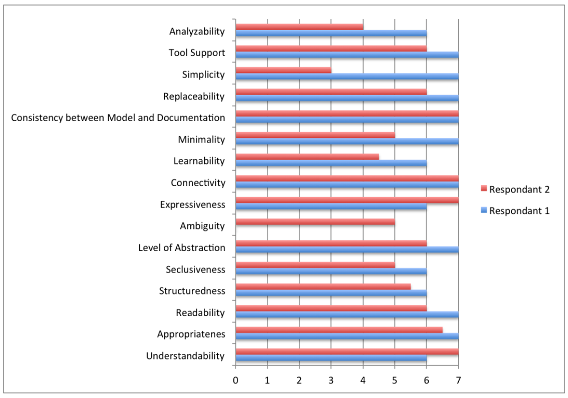


Figure 5.22: ZIVIT - bar chart of all manually criteria

= 4), respondent 2 agrees more (median = 6) with the model. Although, there are differences between the respondents, both would give grade *2* to the model. For this the average school grade is *2* for the overall assessment.

BNetzA - Area Process Model and Process Documentation. In the area of process model and process documentation (4.8) the respondents evaluated the *V-Modell XT BNetzA* also different. But here, the evaluation is converse.

As we can see in the bar chart 5.6, there are difference for the questions *3.2*, *3.4*, *3.8*, *3.10* and *3.11*. While respondent one agreed more strong by the majority, the respondent 2 agreed not completely. The widest differences are in question *3.10* and *3.11*. Respondent 1 strongly agreed that the “scope of the activities (tasks) is satisfying" and agreed more that “level (degree) of abstraction of the activities (tasks) is appropriate", while respondent 2 disagreed somewhat with the scope of activities (question 3.10) and disagreed with level of abstraction of the activities (question 3.11).

Respondent 1 would like to extend nothing, while respondent 2 answered question *3.15* with “Easy access guidelines - here, much improvement is necessary". For question *3.16*, “Wich parts of the process documentation would you like to reduce?", respondent 1 answered “2.1 Produkt- und meilensteinorientierte Grundphilosophie, 2.5 Vorgehensbausteinlandkarte und V-Modell-Kern, 3.1 Systementwicklung (BNetzA), 3.2 Linienaufgaben, 3.3 Systementwicklungsprojekt (AG), 3.4 XÖV-Standardisierungsprojekt" and respondent 2 “Descriptive text from the reference process that does not contribute to the company-specific variant". Both respondents would like to add no additional contents to the model. In the opinion of respondent 1, the process modules map should be taken off. The actions, that respondent 2 would recommend to meet these demands (question 3.18) is the “rework of the export that is used to generate the process documentation from the model".

In sum, we can say that both respondents are nearly satisfied with the process model and documentation as a whole, that can be seen in the net chart 5.5. While respondent 1 agrees more (median = 6), respondent 2 agrees/agrees somewhat (median = 4.5) with the process model and documentation. Although, there are differences between the respondents, both would give grade *2* to the model. For this. the average school grade for the overall assessment of the process model and documentation is *2*.

BNetzA - Area Tool Support. In the area of model tool support (4.9) the respondents evaluated the *V-Modell XT BNetzA* more often the same.

As we can see in the bar chart 5.8, there are only three wide difference for the questions *4.2*, *4.5* and *4.6*. The widest difference is question *4.2*, "I was immediately familiar with the tool and could instantly use all necessary functionality for the realization of the designed process elements." Respondent 1 diagreed somewhat with this staement, while respondent 2 agreed more.

Respondent 2 criticized the the tool was not at all useful to realize the designed model (question 4.3), because "Management of dependencies is poorly implemented in the used tool and requires deep background knowledge. Here, the user interface should be significantly be improved." He/she is not sure to choose the same tool, if he/she had the choice (question 4.18), because (question 4.19) "For the actual context the tool is ok, but barely adoptable to other contexts, and, thus, other tools should be mentioned as well." Respondent 1 would take the same tool, because he/she does not know another one.

Both say, that there are no process requirements, that were not completely realized (question 4.21). Respondent 1 would evaluate "wild-card elements" as necessary to realize missing/incomplete requirements, while respondent 2 answered with "none".

In sum, in chart 5.7 we can see that both respondents are satisfied with tool support. Respondent 1 strongly agrees/agrees more (median = 6.5) and respondent 2 agrees more (median = 6) with tool.

BNetzA - Area Product Templates. In the area product templates (4.10) the respondents evaluated the *V-Modell XT BNetzA* nearly the same.

As we can see in the bar chart 5.10, there are only differences for the questions *5.3* and *5.4*. Respondent 1 agrees strongly that product templates are easily accessible for persons not familiar with the process model ("outsiders") (question 5.3), respondent 2 just agrees. For statement *5.4*, that the "product templates allow for direct inspection of relationships to further products and the identification of consistency violations", respondent 1 agreed, while respondent 2 disagreed somewhat.

Respondent 1 would like to extend part "Systemelemente" of the product templates (question 5.5) and respondent 2 nothing. Both respondents have would nothing to reduce (question 5.6). While respondent 1 need no additional content to add to the product templates (question 5.7), respondent 2 would like to add "Status information and better links to other referred products." For the question "What would be necessary to realize missing/incomplete requirements?" (question 5.8), respondent 1 answered "more information or information in general" and respondent 2 "A change in the export sub-system used to generate the templates."

In chart 5.9 we can see that the respondents harmonize in their assessment. Finally respondent 1 gives grade *1* to the product templates and respondent 2 grade *3*. In average, the product templates gets grade 2 for overall assessment.

BNetzA - Criteria. To get a statement about the integrity of the process model as a whole we analysed the criteria manually verifiable, that were addressed by certain questions.

As we can see in the bar chart 5.12, the respondents evaluated the criteria *understandability*, *readability*, *selclusiveness*, *abiguity*, *expressiveness* and *leanability* different. The widest difference is in the evaluation of *leanability*, *expressiveness*, *abiguity* and *selclusiveness*.

Respondent 1 disagrees with the selclusiveness of the model, while respondent 2 agrees.

Respondent 1 also disagrees somewhat with the expressiveness, but respondent 2 strongly agrees.

Ambiguity was more disagreed by respondent 1. Respondent agrees, that the model is free of ambiguity.

While respondent 1 is in opinion, that the model is easy to learn (median = 6.5), respondent 2 just agrees/agrees somewhat (median = 4.5).

In sum, respondent 1 agrees with the quality of the model (median = 4.75) and respondent 2 agrees more (median = 6), what we can see in the net chart 5.11.

Finally, when we calculate the overall assessment result from respondent 1 and 2 for all criteria evaluated by the questionnaire, we can conclude that the quality of the model is satisfactory (median = 5.375).

ZIVIT - Area Meta Data. Both respondents used the *V-Modell XT ZIVIT* version 1.0 with the export of the process documentation as *PDF-file*. Both got no specific training for the process model and used the product templates generated from the process.

ZIVIT - Area Model Evaluation (absolute). In the area of model evaluation (absolute) (4.7) the respondents evaluated the *V-Modell XT ZIVIT* nearly the same.

As we can see in the bar chart 5.14, there is only are wide difference for the questions *2.5*. While respondent 1 agrees strong, respondent 1 agrees somewhat.

Both respondents say that modularity was considered sufficiently (question 2.15).

For the statement "It was clear to me how to use the interfaces, without loss of consistency of the integrated model." (question 2.16), respondent 1 answered with "yes" and respondent 2 with "not sure".

Both respondents would like nothing to extend (question 2.17).

While respondent 1 would like nothing to reduce (question 2.18), respondent 2 would like to reduce "Descriptive text from the reference process that does not contribute to the company-specific variant".

Both respondents would like to add no additional contents to the model (question 2.19).

While respondent 2 has no further comments (question 2.20), respondent 1 says "The tour through the model is awesome. ;-D".

According to the net chart 5.13, the answers of both respondents are balanced.

In sum, we can say that both respondents are very satisfied with the model as a whole. While respondent 1 strongly agrees (median = 7), respondent 2 agrees more (median = 6) with the model. Respondent 1 evaluated the model with grade *1* and respondent 2 with grade *2*. The average school grade is *1.5*.

ZIVIT - Area Process Model and Process Documentation. In the area of process model and process documentation (4.8) the respondents evaluated the *V-Modell XT ZIVIT* more different. But here, the evaluation is converse.

As we can see in the bar chart 5.16, there are difference for the questions *3.4*, *3.6*, *3.8*, *3.10* and *3.11*. While respondent 1 agreed more strong by the majority, the respondent 2 agreed not completely. The widest differences are in question *3.10* and *3.11*. Respondent 1 strongly agrees that the "scope of the activities (tasks) is satisfying" and strongly agrees more that "level (degree) of abstraction of the activities (tasks) is appropriate", while respondent 2 disagrees somewhat with the scope of activities (question 3.10) and disagrees with level of abstraction of the activities (question 3.11).

Both respondent would like to extend nothing (question 3.15).

Respondent 1 would like to reduce the part (question 3.16), "2.5 Vorgehensbausteinlandkarte und V-Modell-Kern". Respondent 2 answered with"Descriptive text from the reference process that does not contribute to the company-specific variant".

While respondent 2 would like to add no additional contents to the model (question 3.17), respondent 1 would like to rename part 5 chapter "3 Produkte" as "Produkte nach Disziplin" or just "Disziplinen".

The actions, that respondent 2 would recommend to meet these demands (question 3.18) is the "rework of the export that is used to generate the process documentation from the model". Respondent 1 would like"to take off the process modules map".

In chart 5.15, we can see that respondent 1 is nearly completely satisfied with the process model and documentation as a whole. The result of respondent 2 is satisfactory.

While respondent 1 strongly agrees (median = 7), respondent 2 agrees (median = 5) with the process model and documentation. According to this, respondent 1 give grade *1* to the process model, respondent 2 gives grade 2. In average the process model and model documentation is rated with *1.5*.

ZIVIT - Area Tool Support. In the area of model tool support (4.9) the respondents evaluated the *V-Modell XT ZIVIT* also more different.

As we can see in the bar chart 5.18, there are difference for the questions *4.1*, *4.6*, *4.6*, *4.15*, *4.16* and *4.17*.

The widest difference are in question *4.1*, *4.15* and *4.16*.

While respondent 1 agrees strongly on question *4.1*, respondent 2 disagrees somewhat that the"used tool is intuitively applicable and the approach of process modeling is straightforward."

Respondent 1 could export the model at any time (question 4.15), while respondent 2 agrees somewhat.

Also all requirements could be realized and the tool was straightforward for respondent 1, respondent 2 agrees somewhat.

Respondent 2 agrees that the tool was useful to realize the design model (question 4.3) and comments "In this project, a backend component needed to be implemented as well. This was neither tested nor documented, and, thus, tool development and much trail and error work was required, especially to create the deployment packages." But he/she is not sure if he/she would choose the same tool (question 4.19), if he/she had the choice, because "For the actual context the tool is ok, but barely adoptable to other contexts, and, thus, other tools should be mentioned as well. Furthermore, the used backend was not yet tested. It finally worked, but consumes too much effort." Respondent 1 would choose the same tool, because he/she does not know another one (question 4.19). Respondent 2 could realize all process requirements (question 4.20) and has no process requirements, that were not completely realized (question 4.21). Nothing would be necessary to realize missing/incomplete requirements for respondent 2 (question 4.22).

In sum, we can see in chart 5.17 that respondents 1 is completely satisfied with tool support. Respondent 1 strongly agrees (median = 7), while respondent 2 agrees (median = 5) with tool.

ZIVIT - Area Product Templates. In the area product templates (4.10) the result for the *V-Modell XT ZIVIT* is the same like for the *V-Modell XT BNetzA*.

As we can see in the bar chart 5.20, there are only differences for the questions *5.3* and *5.4*. Respondent 1 agrees strongly that product templates are easily accessible for persons not familiar with the process model ("outsiders") (question 5.3), respondent 2 just agrees. For statement *5.4*, that the "product templates allow for direct inspection of relationships to further products and the identification of consistency violations", respondent 1 agrees, while respondent 2 disagrees somewhat.

Respondent 1 would like to extend part "Systemelemente" of the product templates (question 5.5) and respondent 2 nothing.

Both respondents have would nothing to reduce (question 5.6). While respondent 1 need no additional content to add to the product templates (question 5.7), respondent 2 would like to add "Status information and better links to other referred products."

For the question "What would be necessary to realize missing/incomplete requirements?" (question 5.8), respondent 1 answered "more information or information in general" and respondent 2 "A change in the export sub-system used to generate the templates."

In chart 5.19 we can see that the respondents harmonize in their assessment, too.

In sum, respondent 1 strongly agrees with the product templates and respondent 2 agrees more.

Finally respondent 1 gives grade *1* to the product templates and respondent 2 grade *3*. In average, the product templates gets grade 2 for overall assessment.

ZIVIT - Criteria. Finally, we will also get a statement about the integrity of the process model as a whole and analyzed also the criteria manually verifiable for the *V-Modell XT ZIVIT*, that were addressed by certain questions.

As we can see in the bar chart 5.22, both respondents evaluated the criteria nearly the same. We have only two wide difference is in the evaluation of *simplicity* and *ambiguity*.

Respondent 1 agrees strong with the simplicity of the model, while respondent 2 disagrees somewhat.

Ambiguity is strongly disagreed by respondent 1. Respondent agrees, that the model is free of ambiguity.

In sum, respondent 1 strongly agrees with the quality of the model (median = 7) and respondent 2 agrees more (median = 6), what we can see in the net chart 5.21.

Finally, when we calculate the overall assessment result from respondent 1 and 2 for all criteria evaluated by the questionnaire, we can conclude that the quality of the model is very satisfactory (median = 6.5).

5.5 Summary

In this chapter we set the context at first. For this, we described the four process phases of the process life cycle and described the role of the process engineer. Then, we refined the evaluation framework and introduced the key artifacts *Process requirements*, *Technical Process Design*, *Conceptual Process Design*, *Process Life Cycle Support* and *Process Release* and QA-related artifacts. Based on the ArSPI model we determined how the artifacts can be evaluated.

In the subsection 5.4 we described the case and our analysis procedure for the evaluation framework. Finally, we presented the result of the automatically verifiable criteria in form of a consistency check. We showed the results of the questionnaire as net chart and bar chart and described the results of each area of the questionnaire in detail. In addition, we made a overall conclusion about the manually verifiable criteria evaluated in the questionnaire and gave a final statement of the integrity of the analyzed process model.

Chapter 6
Summary and Outlook

In this section we will sum up the content of every part of the work as final conclusion. Finally, we will give a statement for future work.

Übersicht

6.1 Summary of Contributions

In chapter 1 we introduced the term software process improvement and all related definitions, e.g. software process, process quality, software process model, software process maturity, quality model and improvement method.
In subsection 1.1 we firstly described the six key factors of success in SPI. After that, we put our eye on problems of measuring the success of *spi* projects. In subsection 1.2 we identified criteria and approaches for measuring the success of SPI initiatives. Finally, we explain in which way this work contributes the research area and outlines the work.

In chapter 2 we presented a brief overview of the current state of the art in research activities in the area of process engineering related to SPI criteria and procedures.
Afterwards the methods of case study (2.2.1), experimentation (2.2.1), document review/-analysis (2.2.2), mixed methods (2.2.3), the Goal Question Metric Paradigm (2.2.4) and survey or interview (2.2.5), used in this study, were described.

In chapter 3 we presented the overall description of the research design. We introduced the literature review method SLR, that was used to conduct this study.
We presented our research questions and described the selection of the case.
Then our data collection procedures was described in detail and the databases, that we used for the literature search, were mentioned. We defined our general queries and the combination of them to final queries, used for the database search.
After the data collection, it was described in which way we cleaned and filtered the result set and our filter queries were defined.
Then we explained our data analysis procedure with the harmonization and voting process and how we used paper coding for our in-depth analysis of the contributions.
Finally, we described what further expert material was used as additional input for the research design.

In chapter 4 we presented the results of the systematic literature review described in section 3.2 of chapter 3.
The results from the further expert material were also presented.
In section 4.3 we explained why the result of the SLR was not useful for our study and that there was a need to perform a second literature review.
The preparation for this second literature review was done by the Goal Question Metric technique. We described the process of data collection, analysis and consolidation and presented the result in form of 30 criteria for integrity of process models. Then we visualized all criteria in a network of relationships between them.
Finally construction of the questionnaire for the evaluation framework was described and explained what questions of the questionnaire addresses what criterion.

In chapter 5 we set the context at first. For this, we described the four process phases of the process life cycle and described the role of the process engineer.
Then, we refined the evaluation framework and introduced the key artifacts *Process requirements*, *Technical Process Design*, *Conceptual Process Design*, *Process Life Cycle Support* and *Process Release* and QA-related artifacts. Based on the ArSPI model we determined how the artifacts can be evaluated.

In the subsection 5.4 we described the case and our analysis procedure for the evaluation framework. Finally, we presented the result of the automatically verifiable criteria in form of a consistency check. We showed the results of the questionnaire as net chart and bar chart and described the results of each area of the questionnaire in detail. In addition, we made a overall conclusion about the manually verifiable criteria evaluated in the questionnaire and gave a final statement of the integrity of the analyzed process model.

6.2 Future Work

The evaluation framework was developed to evaluate the integrity of process models.

In future, the evaluation framework should be fully integrated in the ArSPI model. We verified the artifact-based evaluation framework for the process models *V-Model XT BNetzA* and *V-Model XT ZIVIT*.

An further approach would be the integration in ArSPI for quality assurance. With this evaluation framework, process engineers are then able to verify their developed process model and procedure for software process improvement projects.

6.3 Summary

In this chapter we presented the summary of the overall work and pointed out future work for the evaluation framework for artifact-based software process improvement procedures.

Appendix A

This appendix contains all criteria found in the SLR (table A.1 to A.4). After that we, from table A.1 to table A.3, we can find the results from Canfora et al. [$CGP^{+}06$]. In the end, we can find the glossary A of the criteria found in the second literature review.

Figure A.1: Classification of Criteria Candidates

No.	Criteria Candidate	Class	Categorization
1	#Problems (per Release)	Q	Number of Problems
2	#frequency of Integration	P	Time
3	duration of 1 cycle	P	Time
4	#of test (cases)	Q	Number of TestCases
5	size of problem backlog		
6	reliability	Q	Non-Functional
7	functionality	Q	Non-Functional
8	safety	Q	Non-Functional
9	(development/project) time	P	Time
10	improvement activities	P	Time
11	(company/process) performance	P	Time
12	budget	C	Cost
13	defects (density)	Q	Number of Problems
14	#of product innovation (/cre Employee)		
15	effectiveness	P	Effectiveness/Efficiency
16	productivity	P	Productivity
17	quality	Q	Quality
18	error rate of size estimate	Q	Precision
19	Ease of use	Q	Non-Functional
20	usefulness		
21	Use		
22	costs of VVT strategy	C	Cost
23	risks of VVT strategy		
24	re-organization		
25	#of sub-oranizations		
26	staff turnover	CTX	Personnel
27	movement of management staff	CTX	Personnel
28	movement of other staff	CTX	Personnel
29	role		
30	understanding		
31	process expertise	CTX	Knowledge
32	improvement strategy		
33	goal: CMM(I) level 2	CTX	Compliance
34	defect removal effectiveness	P	Effectiveness/Efficiency
35	reuse		
36	Estimating	Q	Precision
37	Skill	CTX	Personnel
38	Testing		
39	Satisfaction	Q	Customer
40	Code	Q	Quality
41	Management		
42	Historical		
43	Development/Maintenance environment	CTX	Environment
44	Benchmark		
45	Programming		

Figure A.2: Classification of Criteria Candidates (2)

No.	Criteria Candidate	Class	Categorization
46	CAS		
47	Development/maintenance	P	Effectiveness/Efficiency
48	Project planning & Development	P	Effectiveness/Efficiency
49	Methods/tools		
50	Standards & guidelines	CTX	Compliance
51	User involvement	Q	Customer
52	Design		
53	Requirements design for operation		
54	Project		
55	Organizational orientation		
56	Organizational objectives		
57	Organizational structure		
58	Base of power		
59	Decision making		
60	Leadership style		
61	Compliance		
62	Evaluation of members		
63	Orientation to change		
64	reason of adopting SPI		
65	readiness for SPI	CTX	Awareness
66	process changes		
67	Kano model		
68	respond time of SPI projects	P	Time
69	time-to-market	P	Time
70	tailoring time	P	Time
71	efficiency	P	Time
72	effort	P	Time
73	#of discarded requirements	Q	Quality
74	bottlenecks	P	Effectiveness/Efficiency
75	unnecessary work	P	Effectiveness/Efficiency
76	rework	P	Effectiveness/Efficiency
77	capability		
78	information requirements		
79	stability & certainty of IT group's environemnt		
80	adaption		
81	financial culture within the organization		
82	interest		
83	environemnt		
84	communication of information		
85	emphasis on quality in the IT group		
86	decision-making		
87	organizational culture		
88	discretion work		
89	and relationships		
90	content of changes		

Figure A.3: Classification of Criteria Candidates (3)

No.	Criteria Candidate	Class	Categorization
91	experience	CTX	Personnel
92	influences		
93	result of actions		
94	Patterns		
95	external context	CTX	Environment
96	organisational context	CTX	Environment
97	context of IS function	CTX	Environment
98	context of process structures	CTX	Environment
99	context of structure/action linkage	CTX	Environment
100	product development actions		
101	progress	P	Time
102	stability (repeatability)		
103	equivalence (consistency)		
104	requirements management		
105	project tracking & oversight		
106	quality assurance		
107	configuration management		
108	process focus		
109	process definiton		
110	training program		
111	integrated software management		
112	software product engineering		
113	intergroup coordination		
114	Peer review		
115	quantitative process management		
116	defect prevention		
117	technology change management		
118	process change management		
119	initiatives		
120	process assessment methods		
121	feasibility		
122	#of entities	M	Model/Product Properties
123	#of simple attributes	M	Model/Product Properties
124	#of composite attributes	M	Model/Product Properties
125	#of derived attributes	M	Model/Product Properties
126	#of M:N-relationships	M	Model/Product Properties
127	#of 1:N-relationships (incl. 1:1)	M	Model/Product Properties
128	#of binary relationships	M	Model/Product Properties
129	#of N-Ary relationships	M	Model/Product Properties
130	#of IS_A relationships	M	Model/Product Properties
131	#of reflexive relationships	M	Model/Product Properties
132	total number of attributes	M	Model/Product Properties
133	total number of relationships	M	Model/Product Properties
134	entity maintenance		
135	diagram connectivity		

Figure A.4: Classification of Criteria Candidates (4)

No.	Criteria Candidate	Class	Categorization
136	diagram inheritance		
137	N:M relationship	M	Model/Product Properties
138	relationship complexity	M	Model/Product Properties
139	identity		
140	perception of the environment		
141	organizational straetgy		
142	knowledge		
143	boundaries		
144	interactive process		
145	triggers		
146	Experimentation		
147	communication, processes & internal standards		
148	Information and communication systems		
149	publication)	P	Effectiveness/Efficiency
150	#CPU hours	P	Effectiveness/Efficiency
151	#of changes	Q	Number of Problems
152	#of pages of documentation	Q	Quality
153	#total new SLOC	P	Productivity
154	#total modified SLOC	P	Productivity
155	#total reused SLOC	P	Productivity
156	stability of requirements	Q	Stability
157	degree to which plans where followed		
158	communication overhead	P	Time
159	readability	Q	Non-Functional
160	maintainability	Q	Non-Functional
161	estimation		
162	monitoring		
163	control		
164	customer relations		
165	Goals		
166	characteristics		
167	evolution of process definition		
168	improvement of process definition		
169	benefits		
170	capability level	CTX	Compliance
171	company size	CTX	Environment

Metric (Genero)	**Description**	**Measurement method**	**Definition of the metric on the E/Rmetamodel**
NE	Number of entities	'Count the number of entities of the E/R diagram'	'Count' measurement method on the metamodel MOF class *Entity* (e.g. count the number of instances of this class in the E/R model)
NAS	Number of simple attributes	'Count the number of simple attributes of the E/R diagram'	'Count' measurement method on the metamodel MOF class *Simple Attribute*
NCA	Number of composite attributes	'Count the number of composite attributes of the E/R diagram'	'Count' measurement method on the metamodel MOF class *Composite Attribute*
NDA	Number of derived attributes	'Count the number of derived attributes of the E/R diagram'	'Count' measurement method on the metamodel MOF class *Derived Attribute*
NM:NR	Number of M:N relationships	'Count the number of M:N relationships of the E/R diagram'	'Count' measurement method on the metamodel MOF class *Binary Relationship* with the following constraint: the 'Arity' property of the *Binary Relationship has value* 'M:N'
N1:NR	Number of 1:N relationships (including 1:1 relationships)	'Count the number of 1:N relationships of the E/R diagram'	'Count' measurement method on the metamodel MOF class *Binary Relationship* with the following constraint: the 'Arity' property of the *Binary Relationship has value* '1:N' or '1:1'
NBinaryR	Number of binary relationships	'Count the number of binary relationships of the E/R diagram'	'Count' measurement method on the metamodel MOF class *Binary Relationship*
NN-AryR	Number of N-Ary (no binary) relationships	'Count the number of N-Ary relationships of the E/R diagram'	'Count' measurement method on the metamodel MOF class *N-Ary Relationship*
NIS AR	Number of IS A relationships (generalization/ especialization)	'Count the number of IS A relationships of the E/R diagram'	'Count' measurement method on the metamodel MOF-Association *Parent* Of (e.g. count the number of links of this association in the E/R model)
NrefR	Number of reflexive relationships	'Count the number of reflexive relationships of the E/R diagram'	'Count' measurement method on the metamodel MOF class *Binary Relationship* with the following constraint: 'the *Binary Relationship* has two *Relationship Ends* which *reference* the same *entity*'

Table A.1: Direct metrics and measurement methods defined to measure E/R diagrams [CGP+06].

Metric	Description	Measurement function
NA	Total number of attributes in an E/R diagram, considering the entities and relationships attributes. This total includes simple, composite and derived attributes, each of them takes the value 1	NA = NAS + NCA + NDA
NNR	Total number of relationships in an E/R diagram (without considering IS A relationships)	NNR = NbinaryR + NN-AryR

Table A.2: Indirect metrics and measurement functions defined to measure E/R diagrams [CGP+06].

Metric	Description	Analysis model	Decision criteria
IME	Entity maintenance indicator: ratio of attributes and entities of the diagram. The higher the value, the higher the maintenance difficulty	$IME = \frac{NA}{NE}$	If $IME > 15 \rightarrow$ IME = 'Very High' If $10 < IME \leq 15 \rightarrow$ IME = 'High' If $5 < IME \leq 10 \rightarrow$ IME = 'Medium' If $0 \leq IME \leq 5 \rightarrow$ IME = 'Low'
IC	Diagram connectivity indicator: ratio of relationship and entities of the diagram. The higher the value, the higher the maintenance difficulty	$IC = \frac{NNR}{NE}$	If $IC \geq 2 \rightarrow$ IC = 'Very High' If $1.5 \leq IC < 2 \rightarrow$ IC = 'High' If $1 < IC < 1.5 \rightarrow$ IC = 'Medium' If $0.5 \leq IC \leq 1 \rightarrow$ IC = 'Low' If $0 < IC < 0.5 \rightarrow$ IC = 'Very Low'
IH	Diagram inheritance indicator: ratio of inheritance relationships and total number of relationships (including inheritance). The higher the value, the higher the maintenance difficulty	$IH = \frac{NISAR}{NNR+NISAR}$	If $IH \geq 0.5 < 2 \rightarrow$ IH = 'High' If $0.1 < IH < 0.5 \rightarrow$ IH = 'Medium' If $0 \leq IH \leq 0.1 \rightarrow$ IH = 'Low'
INM	N:M relationship indicator: ratio of N:M relationships. The higher the value, the higher the maintenance difficulty	$INM = \frac{NM{:}NR}{NNR}$	If $INM \geq 0.5 \rightarrow$ INM = 'High' If $0.25 < INM < 0.5 \rightarrow$ INM = 'Medium' If $0 \leq INM \leq 0.25 \rightarrow$ INM = 'Low'
CI	Relationship complexity: this indicator represents the ratio of N-Ary relationships. The higher the value, the higher the maintenance difficulty	$CI = \frac{NN-AryR}{NNR}$	If $CI \geq 0.2 \rightarrow$ CI = 'Very High' If $0.1 < CI < 0.2 \rightarrow$ IH = 'High' If $0.05 < CI < 0.1 \rightarrow$ IH = 'Medium' If $0 \leq CI \leq 0.05 \rightarrow$ IH = 'Low'

Table A.3: Indicators, analysis models and decision criteria to evaluate E/R diagrams [CGP+06].

⇒ Property: Ability to Achieve Consensus

Acronym	CA1
Category	automated
Source	[KS]

Definition: The model is suitable for cross-functional standardization of functional objects.

Input: Determination to what extent de facto standards are considered, what organizations supported the model development, that means if the process of building consensus took place, and what functional inputs served the model development (e.g. data model, schema integration of application object models with expert consultation, identification of use cases with JAD).

⇒ Property: Ambiguity

Acronym	CM1
Evaluation type	manually
Source	[Fle]

Definition: A model has to be free of ambiguity in the case of descriptions and nomenclatures.

Input: number of weak phrases, number of optional phrases

⇒ Property: Analyzability

Acronym	CAM1
Evaluation type	automated and/or manually
Source	[BDE+05]

Definition: Attributes of a model that relate to the effort needed for diagnosis of deficiencies or for identification of parts to be modified.

Input: Likert-Scala

⇒ Property: Appropriatenes

Acronym	CM2
Evaluation type	manually
Source	[Wag13], [BDE+05]

Definition: Appropriatenes expresses that the model shall provide the functionality to the user that fits to their requirements and expectations (includes functional correctness). The process model shall constitute a meaningful representation of the application (process) area and has to be applicable for its context. Model predictions have to be precise and relevant related to functional requirements.

Input: Likert-Scala

⇒ Property: Changeability

Source	[Sea, Ber, BDE+05, Kur, Pre]

Definition: Changeability is expressed by tailoring adaptability, evolution adaptability and replaceability.

⇒ **Property: Completeness**

Source	[Wag13], [BDE+05], [Mac00], [FM07]

Definition: Completeness is expressed by Appropriateness and Minimality.

⇒ **Property: Completeness of Realization**

Acronym	CM3
Evaluation type	manually
Source	[Fle]

Definition: The implementation of the model must completely contain all designed contents.

Input: Task completions, planned task completions

⇒ **Property: Complexity**

Source	[KS, Fle, Bas]

Definition: Complexity is expressed by implementability, the level of abstraction and structuredness.

⇒ **Property: Compliance**

Acronym	CAM2
Evaluation type	automated and/or manually
Source	[KS], [KTF09]

Definition: Different models must be based on the same professional understanding. The content of the models have to be compliant so that these models can be hormonized and integrated. The analytical compliance shall be examined by means of the process documentation and the meta-model with reference to characteristics and contents of the process description, relationsships between contained items and the coverage of contents related to the reference model.

Input: Analytical determination (according to conformity program) by mapping of model contents directly between models and processes at a established joint and method neutral consolidation basis for consitency reliability. To keep comparison of various models practicable, models shall be consolidated, compared and adjusted at a term based level.

⇒ **Property: Conformance**

Source	[KS], [KTF09], [BDE+05], [Ber]

Definition: Conformance is expressed by Interoperability and Compliance.

⇒ **Property: Connectivity**

Acronym	CAM3
Evaluation type	automated and/or manually
Source	[Soc], [Dun95]

Definition: Connectivity is the ability of any model component to communicate with any of the other components inside and outside of the organizational environment.

Input: Likert-Scala

⇒ Property: Consistency

Acronym	CA2
Evaluation type	automated
Source	[KS], [Fle]

Definition: Model predictions have to be clearly and unambiguous. The model must be free of contradictions.

Input: Coupling between common business object (CBO) classes. Coupling is the degree of independence between modules [YC79]. It measures how well two software components are data related, i.e. how independent they are.

⇒ Property: Consistency between Model and Documentation

Acronym	CM4
Evaluation type	manually
Source	[Fle]

Definition: The model documentation has to reflect the fundamental process model in an adequate way.

Input: Verification of the model realization and the exported model documentation in consideration of correct locations of the positioned elements.

⇒ Property: Evolution Adaptability

Acronym	CAM4
Evaluation type	automated and/or manually
Source	[Kur], [Ber]

Definition: Characterizes the ability of the model to change to new specifications or environments.

Input: ad-hoc transformation or implementing an extension of the transformation language, list of environments the model need to be compliant with

⇒ Property: Expandability

Acronym	CAM5
Evaluation type	automated and/or manually
Source	[BDE^{+}05]

Definition: Expandability expresses the ease of adding new functionality to the model.

Input: Likert-Scala

⇒ **Property: Expressiveness**

Acronym	CM5
Evaluation type	manually
Source	[KS]

Definition: In consideration of comprehensibility and expressiveness it has to be clarified by which models and diagrams a process model is documented and what naming and documentation conventions must be strictly adhered to.

Input: Likert-Scala

⇒ **Property: Feasibility**

Source	[KS, Fle, JB]

Definition: Feasibility is expressed by understandability, expressiveness, readability, ambiguity and constencey between model and documentation.

⇒ **Property: Flexibilty**

Source	[Ver, BDE+05, TKM+, Soc, PMMP08, Jes]

Definition: Flexibility is expressed by modularity, seclusiveness, generalizability, connectivity, granularity, expandability and simplicity.

⇒ **Property: Generalizability**

Acronym	CAM6
Evaluation type	automated and/or manually
Source	[PMMP08], [BDE+05]

Definition: Models should not be evaluated on their goodness of fit but on their generalizability —a criterion that is considered the gold standard in statistical model selection. General solutions that by nature are prepared for being utilised in other contexts than the ones for which they were constructed.

Input: Generalizability is measured by trading off goodness of fit for model flexibility (Akaike Information Criterion, the Bayesian Information Criterion, the Bayes factor, and minimum description length, simulation performance).

⇒ **Property: Granularity**

Acronym	CA3
Evaluation type	automated
Source	[Ver], [TKM+]

Definition: In modelling, granularity refers to the degree of detail and precision contained in a model. It is an issue for the flexibility of the data model and artefacts designed from it.

Input: Comparing models made from 1-second versus 1-minute data. Compression of temporal data by arithmetic mean, therefore, can be an effective method for decreasing knowledge discovery processing time without compromising learning.

⇒ Property: Implementability

Acronym	CM6
Evaluation type	manually
Source	[KS]

Definition: With the implementation of the model the utilization concept, respectively the implementability of the model, will be examined concerning to goals.

Input: Existence of a process model with various development paths, guidelines for cutting sub-models or mapping guidelines of model items.

⇒ Property: Interoperability

Acronym	CAM7
Category	automated and/or manually
Source	[Ber], [BDE^{+}05]

Definition: Ability of a model to interact with other specified models or systems.

Input: Relevant check list for each model to assess the interoperability.

⇒ Property: Learnability

Acronym	CM7
Evaluation type	manually
Source	[Ber], [DG81], [BDE^{+}05]

Definition: Learnability mirrors the learning effort for different users, i.e. novice, expert, casual etc. Attributes of the model that relate to the users' effort for learning its process (for example, operation control, input, output).

Input: ease of learning the model, number of reports requested for information, number user errors

⇒ Property: Level of Abstraction

Acronym	CM8
Evaluation type	manually
Source	[KS]

Definition: The level of abstraction defines the depth of the abstraction of the model.

Input: degree of abstraction, Likert-Scala

⇒ Property: Minimality

Acronym	CM9
Evaluation type	manually
Source	[Mac00], [FM07]

Definition: We consider a first-order structures M, where M is the domain, and there may be other symbols for relations, functions or constants in the language. M is o-minimal if every definable subset of M is a finite union of singletons and open intervals. When we say that a structure is model complete we mean that its theory is model complete.

Input: Robinson's test, full exponential function

⇒ **Property: Modularity**

Acronym	CAM8
Evaluation type	automated and/or manually
Source	[Soc], [BD], [BDE+05], [Dun95]

Definition: Modularity is the ability to easily reconfigure (add, modify, switch or remove) model components for in a controlled fashion and is the standardization of business processes for shareability and reusability.

Input: constructs (IF-THEN), norms

⇒ **Property: Readability**

Acronym	CM10
Evaluation type	manually
Source	[KS], [JB]

Definition: Readability of a model can be improved, corresponding to the model partitioning, by the description of the model contents in different levels of abstraction and itemization.

Input: Likert-Scala

⇒ **Property: Redundancy**

Acronym	CA4
Evaluation type	automated
Source	[JJ]

Definition: Redudancy is a duplication of the problem domain knowledge in the representation including all concepts represented in an artifact (e.g structural aspects repeat multiple times, clones based on syntactic characteristics). Redundancy is observed if a shorter representation can be found from which reproduction without loss of information can be done.

Input: top-down approach to discover and avoid redundancy through analysis of the properties of the schema, as well as ontologies as the models of the problem domain, clone detection approaches (e.g. clone count) for process-representation-based similarity on syntactic similarity

⇒ **Property: Replaceability**

Acronym	CAM9
Evaluation type	automated and/or manually
Source	[Sea], [Ber], [BDE+05]

Definition: Replaceability immediately appears to break the physical systems analogy, to enhance or modify the functionality of a system or for flexibility in manufacturing to reduce cost and incorporate a higher quality. Replaceability characterizes the plug and play aspect of model components, that is how easy it is to exchange a given model component within a specified environment. Attributes of the model that relate to the opportunity and effort of using it in the place of specified other models in the environment of that model.

Input: Ease of implementing a new one.

⇒ Property: Reuseability

Acronym	CA5
Evaluation type	automated
Source	[KS], [Fle], [JB], [KTF09]

Definition: Reuseability means that contents of a project progress stage are reusable in several project situations.

Input: Data Coupling ratio. Data coupling measures how closely two components are tied together, with loose data coupling being desirable. Reuseabelitity is also expressed by stability.

⇒ Property: Seclusiveness

Acronym	CAM10
Evaluation type	automated and/or manually
Source	[Jes]

Definition: Seclusiveness means that the content structure follows a well defined standard and is of self-documenting nature.

Input: Individual types of a XML schema represent direct mappings of the formal notations.

⇒ Property: Simplicity

Acronym	CM11
Evaluation type	manually
Source	[BDE+05]

Definition: In general, simplicity fosters understandability. A simple and understandable system is easier to modify than a complex one.

Input: Likert-Scala

⇒ Property: Stability

Acronym	CAM11
Evaluation type	automated and/or manually
Source	[KS], [JP], [DG81], [KTF09], [MFS], [BDE+05], [MFHT02], [Ara94], [CHWL]

Definition: Stability aims at accomplishing model reusability. The way stable models are built should guarantee reusability. Model stability provides a stable core that can serve processes sharing similar core structure. Artifacts develop a hierarchal order of the model objects, from totally stable at the enduring business themes level to unstable at the industrial objects level, through adaptable yet stable at the business objects level. The stable objects of the system are those that do not change over time. A model should be designed is such a way that new use cases to integrate, respectively additional or changed functional requirements, lead only to low, locally bounded model adaptions. It addresses attributes of the model that relate to the risk of unexpected effects of modifications.

Input: Software Stability Model (SSM), direct determination of approved procedures (model architecture for stability, design patterns for flexibility, analysis patterns for reuse, code of modeling for stability) that support a well- model design in the process model. As well as localization of relationships in the model by target-performance comparison in the process of tailoring.

⇒ Property: Structuredness

Acronym	CM12
Evaluation type	automated
Source	[Fle], [Bas]

Definition: The structure of a model is often a good indicator of whether that process is well designed, understandable and easy to modify.

Input: logic complexity, control structure (number of decisions by measuring the number of constructs that represent branches in the flow of control), control of complexity (weighting various types of control structures as simple or complex)

⇒ Property: Tailoring Adaptability

Acronym	CM13
Evaluation type	manually
Source	[Pre]

Definition: A software process model has be able to be adapted to organization's specific project needs.

Input: assessment with checklists

⇒ Property: Tool Support

Acronym	CM14
Evaluation type	manually
Source	[KTF09]

Definition: As general condition it has to be determined which tools are already present, which tools, if necessary, have to be acquired and which have to be lie quit.

Input: Evaluation if tools are adequate and complete.

⇒ Property: Understandability

Acronym	CM15
Evaluation type	manually
Source	[BDE+05]

Definition: The model possesses the characteristic understandability to the extend that it purpose is clear to the process engineer. Attributes of the model that relate to the users' effort for recognizing the logical concept and its applicability.

Input: Likert-Scala

⇒ Property: Validity

Acronym	CAM12
Evaluation type	automated and/or manually
Source	[KS]

Definition: The process model is a valid instance of a metamodel and is (self-)conclusive in itself.

Input: structural validity check (e.g. XML schema validation)

Appendix B

Übersicht

Autoren:	TU München	Dr. Marco Kuhrmann		Seite 1 von 7
	BVA/BIT 7	Dr. André Schnackenburg Dr. Christian Lange	Erstellt von: Erstelldatum: Zuletzt gespeichert am:	Christian Lange 11.06.14 23:38

Fragebogen (Entwurf)

Einsatz des V-Modell XT in der Behördenpraxis – Analyse, Auswertung, Verbesserung

Zielsetzung

Durch den Fragebogen sollen Daten erhoben werden, die Aufschluss über den Einsatz des V-Modell XT in der Behördenpraxis geben.

GQM-Paradigma

Das Goal-Question-Metric Paradigma (GQM) sieht eine Beschreibung einer empirischen Studie auf drei Abstraktionsstufen vor, wobei sich die Fragen der jeweils niedrigere Stufe von denen der jeweils höheren Stufe ableiten. Auf der höchsten Stufe werden die Kernfragen als Ziele (Goals) der Studie festgelegt. Davon leiten sich die zu beantwortenden Untersuchungsfragen (Questions) ab. Davon wiederum abgeleitet sind die konkreten Metrieken (Metrics), die es in der Studie zu messen gilt. Da es sich in der vorliegenden Studie um einen Fragebogen handelt, sind die Metrieken die eigentlichen Fragen, aus denen sich der Fragebogen zusammensetzt.

Kernfragen (Stufe 1)

Diese Studie betrachtet zwei Kernfragen, wobei die eine die Art und Weise der V-Modell XT Anwendung betrachtet und die andere die Auswirkungen der V-Modell XT Anwendung auf den Entwicklungsprozess, die Produkte und teilweise die Organisation. Die beiden Kernfragen werden wie folgt formuliert:

G1: Welche Kontext-Faktoren sind erforderlich für eine erfolgreiche V-Modell XT Anwendung?

G2: Welche Auswirkungen hat der Einsatz des VMXT?

Dabei betrachtet G1 die sog. „unabhängigen Variablen“ (Ursache, Vorbedingungen) und G2 betrachtet die „abhängigen Variablen“ (Wirkung, Effekt).

Untersuchungsfragen (Stufe 2)

Zu G1 (Ursache):

- Q1: Wie (konsequent) wird das V-Modell XT angewendet?
- Q2: Weshalb wird das V-Modell XT angewendet?
    - Auch: Wer ist der Driver für die Anwendung des V-Modell XT?
- Q3: Wie ist die Akzeptanz des V-Modell XT?
    - Sowohl bei dem Projekt-MA, als auch die Einbettung des V-Modell XT innerhalb der Organisation.
- Q4: In wiefern ist das V-Modell XT anwendbar auf Ihr Projekt?

Autoren:	TU München	Dr. Marco Kuhrmann		Seite 2 von 7
	BVA/BIT 7	Dr. André Schnackenburg Dr. Christian Lange	Erstellt von: Erstelldatum: Zuletzt gespeichert am:	Christian Lange 11.06.14 23:38

    - Bezüglich Tailorability, Verständlichkeit und eventuell erforderlicher Anpassung an organisationsspezifische Gegebenheiten.
- Q5: Welche Umstände haben zum Nicht-Einsatz des V-Modell XT geführt?
- Q6: Was war der Ausgangspunkt VOR der Nutzung des VMXT?
    - Wurde vorher bereits strukturiert vorgegangen?

Zu G2 (Wirkung):

- R1: Bringt der Einsatz des VMXT messbare Vorteile?
    - Insbesondere im Hinblick auf eine Verbesserung des Projektverlaufs und der Produktqualität.
- R2: Wird durch den Einsatz des VMXT Mehraufwand verursacht?
    - Hier gilt es zu hinterfragen, ob es sich um einmaligen Mehraufwand durch die Einführung des VMXT handelt oder um wiederkehrenden Mehraufwand durch Projektoverhead.
- R3: Bringt das VMXT nachweisbare Nachteile?
    - im Sinne von: Hat es Ihr Projekt gehemmt oder beschränkt? Ggf. auch andere Nachteile
- R4: Welche Folgen hat der Einsatz des VMXT nach dem Projekt bewirkt?
    - z.B.: auf Folgeprojekte, Anpassung der Organisation, Wissenstransfer…
- R5: Hat es weniger Kommunikationsprobleme gegeben als ohne VMXT-Einsatz?
    - CR’s. Missverständnisse,…
- R6: Wurde in Ihrer Behörde jemals ein VMXT-Projekt abgebrochen?

Metrieken / Konkrete Fragen (Stufe 3)

Hier listen wir die von den Untersuchungsfragen abgeleiteten Fragen auf, die konkret im Fragebogen vorkommen könnten. Diese Fragen sollten selektiert bzw. ergänzt werden, um den endgültigen Fragebogen zu erstellen. Möglicherweise werden einige Fragen nur im Interview, nicht aber im Fragebogen gestellt.

Zu Q1:

- Q1.1: Nutzen Sie eine Werkzeugunterstützung für die Nutzung/Anwendung des VMXT? Welche?
- Q1.2: Wie können Projektmitglieder auf das Projekthandbuch zugreifen?
- Q1.3: Haben Sie Einfluss auf das Projekthandbuch des Auftragnehmers gewonnen?
- Q1.4: Wie viele Iterationen hat Ihr Projekt durchlaufen?

Autoren:	TU München	Dr. Marco Kuhrmann		Seite 3 von 7
	BVA/BIT 7	Dr. André Schnackenburg	Erstellt von:	Christian Lange
		Dr. Christian Lange	Erstelldatum:	11.06.14 23:38
			Zuletzt gespeichert am:	

- Q1.5: Gegen die Anwendung welcher Regeln im VMXT haben Sie sich bewusst entschieden?
- Q1.6: Haben Sie andere Vorgehensmodelle in das V-Modell XT Projekt integriert? (oder: Kombinieren Sie ein anderes Vorgehensmodell mit dem VMXT?)
- Q1.7: Wurden alle Produkte erstellt, die im Projekthandbuch festgelegt wurden?
- Q1.8: Wurden Ergänzungen/Anpassungen der Projektdurchführungsstrategie vorgenommen, die über das Tailoring hinausgehen?
- Q1.9: Welche Vertragsart wurden mit dem AN abgeschlossen?
  - Werkvertrag oder Dienstvertrag
- Wie wird die Einhaltung des VMXT kontrolliert? (zum Beispiel Fragen aus der TUM (Fritzsche / Keil), Seite 47/48)
- **Fragen zum AN**
  - Hat der AN das V-Modell XT angewendet?
  - Wurden alle vom V-Modell XT vorgesehenen Produkte des Auftragnehmers an den Auftraggeber geliefert?
  - Welches Know-How zum V-Modell lag beim Projektteam des AN vor?
    - Alle MA zertifiziert / MA teilweise zertifiziert / keine Zertifikate, aber Know-How vorhanden / kein Know-How vorhanden

Zu Q2:

- Q2.1: Gibt es einen VMXT-Verantwortlichen in Ihrer Behörde?
  - Falls ja, was ist seine Aufgabe? (z.B. Coaching, Fortbildung, Koordinierung, Antrieb)
- Q2.2: Wurde die erstmalige Anwendung des VMXT durch einen Erlass in Ihrem Ressort ausgelöst?
- Q2.3: Gibt es in Ihrem Projekt einen oder mehrere Personen, die ohne formellen Auftrag Wissen über das VMXT erworben haben?
- Q2.4: Hat ihr Projektteam sich gezielt für den Einsatz des VMXT entschieden?
- Q2.5: Gab es besonders gewichtige Erwartungen, die sie veranlasst haben, das VMXT einzusetzen?

Zu Q3:

- Q3.1: Haben Sie das VMXT aus Anlass Ihres Projektes erlernt?
- Q3.2: Würden Sie das VMXT erneut verwenden?
  - Ja / Nein / in anderer Form

Autoren:	TU München	Dr. Marco Kuhrmann		Seite 4 von 7
	BVA/BIT 7	Dr. André Schnackenburg	Erstellt von:	Christian Lange
		Dr. Christian Lange	Erstelldatum:	11.06.14 23:38
			Zuletzt gespeichert am:	

- Q3.3: Haben Ihre Mitarbeiter das VMXT eigenverantwortlich genutzt?
- Q3.4: Sind inhaltliche Redundanzen zwischen den Produkten entstanden?

Zu Q4:

- Q4.1: War das VMXT geeignet ein Projekt Ihrer Größe durchzuführen?
- Q4.2: Konnten Sie in Ihrem Projektkontext unnötige VMXT-Produkte eindeutig identifizieren?
  - Nein,
  - Ja (im Interview: welche Produkte?)
- Q4.3: War das VMXT dazu geeignet, Ihren bevorzugten Software-Entwicklungsprozess abzubilden?
  - Ja
  - Nein (Interview: welchen?)
-

Zu Q5:

- Q5.1: Gab es ein oder wenige besonders gewichtige Kriterien, die zum Nicht-Einsatz geführt haben? Wenn ja, welche?
- Q5.2: Gab es ein oder mehrere Personen, die den Nicht-Einsatz gefordert haben?
- Q5.3: Welche Umstände haben zum Nicht-Einsatz des VMXT geführt?
  - Zeitdruck
  - Zu viel Overhead durch das VMXT
  - Mangel an Wissen bzw. geschultem Personal
  - Schlechte Erfahrung mit dem VMXT
  - Fehlende Werkzeugunterstützung
  - Das VMXT ist nicht geeignet für mein Projekt

Zu Q6:

- Q6.1: Haben Sie schon einmal nach einem anderen Vorgehensmodell als dem VMXT gearbeitet?
- Q6.2: Haben Sie ein Ingenieur-ähnliches Studium? (Informatik, Elektrotechnik, Mathematik, Physik,…)
- Q6.3: Hat Ihre Organisation ein eigenes oder angepasstes Vorgehensmodell?

- Q6.4: Haben Sie eigene Softwareentwickler im Projekt oder in Ihrer Organisation?

Zu R1:

- R1.1: Wurde das Projekt erfolgreich abgeschlossen?
- R1.2: Wurde das Projekt innerhalb der veranschlagten Zeit abgeschlossen?
- R1.3: Wurde das Softwareprodukt mit vollständigem Funktionsumfang vom Auftragnehmer geliefert?
- R1.4: Wurde das benötigte Wissen über das zu erstellte Softwareprodukt vom AN auf den AG übertragen?
- R1.5: Wie beurteilen Sie die Qualität des erstellten Softwareprodukts?

Zu R2:

- R2.1: Haben Sie Schulungen zum VMXT vor oder während Ihres Projektes genutzt?
- R2.2: Haben Sie spezielles VMXT-Coaching benötigt?
- R2.3: Konnten Sie Ihre Projektbibliothek im Allgemeinen ohne Mehraufwand pflegen?
- R2.4: Haben Sie Termine auf Grund erhöhter Dokumentationsaufwände nicht gehalten?

Zu R3:

- R3.1: Haben Sie eine von Ihnen favorisierte Entscheidung auf Grund der Vorgaben des VMXT während Ihres Projektes anders getroffen?
  - Wenn ja: Hätte Ihre ursprüngliche Entscheidung dem Projekt mehr genützt?

Zu R4:

- R4.1: Wurde nach Ihrem Projekt ein behördenscpezifisches VMXT entwickelt oder angepasst?
- R4.2: War Ihr Projekt Musterbeispiel für die Projektdurchführungsrichtlienien in Ihrem Hause?
- R4.3: Gibt es Projektleiter, die sich explizit an Sie mit Fragen zum VMXT gewandt haben?
- R4.4: Wurde während oder nach Projektabschluss ein VMXT-Schulungsbedarf identifiziert?
- R4.5: Hat die Bereitschaft zur Durchführung von Projekten nach VMXT nach dem Projekt eher zu- oder abgenommen?

Zu R5:

- R5.1: Schätzen Sie die Menge der CR's ein, die nicht aufgrund von veränderten Anforderungen entstanden sind.
  - Niedrig, Mittel, Hoch
- R5.2: Gab es in Ihrem Projekt zu einem Zeitpunkt eine Arbeit, die nicht hätte ausgeführt werden müssen, wenn man Informationen gekannt hätte, die an anderer Stelle im Projekt zum selben Zeitpunkt bekannt waren?
- R5.3: Sind in ihrem Projekt häufig oder schwerwiegende Missverständnisse aufgetreten?
- R5.4: Waren sie zu jedem Zeitpunkt des Projektes ausreichend über den Projektstatus informiert?
- R5.5: Gab es regelmäßige Abstimmungen zwischen Auftraggeber und Auftragnehmer?

Zu R6:

- R6.1: Wurde ihr Projekt abgebrochen?
- R6.2: Wurden in Ihrer Behörde jemals Projekte abgebrochen?
  - Falls ja, wurde in Ihrer Behörde ein VMXT-Projekt abgebrochen?

Sonstige Fragen:

- Wie beurteilen Sie den Einsatz des VMXT im Vergleich zu Ihren vorherigen Projekten bzgl. der folgenden Aspekte:
  - Flexibilität (1 – 5)
  - Mehraufwand (1 – 5)
  - Kommunikation (1 – 5)
- Welche Unterstützung zur Anwendung des VMXT schätzen Sie als hilfreich ein (z.B. Anwendungsleitfaden) [Vorschlag von M. Kuhrmann]

Allgemeine Fragen

- **Projektbeschreibung**
  - Wie viele VMXT-Projekte wurden in Ihrer Behörde/Org.-Einheit vor diesem Projekt bereits durchgeführt?
  - Projekttyp?
  - Größe des Projekts in #Personen?
  - Welche Laufzeit hatte ihr Projekt?
  - Wurde eine organisationsspezifische Anpassung des VMXT durchgeführt?
    - Falls ja, warum?

  - Durch wen?
- Wurde ein projektspezifisches Tailoring durchgeführt?
  - Durch wen?
- Wurden die Dokumentvorlagen angepasst?
- Stand ihr Projekt unter besonders hohem Zeitdruck?
- Teamaspekte:
  - Wie viele Jahre Erfahrung in IT-Projekten haben Ihre Projektmitarbeiter durchschnitttlich?
  - Wie hoch schätzen Sie die Arbeitsmotivation Ihrer Projektmitarbeiter ein? Gering, mittel, hoch
- [Filterfrage]: Haben Sie mit Ihrem VMXT-Projekt Software gebaut? Wenn nein, was war dann das Endprodukt Ihres Projektes? (z.B. Organisations-Projekte)

Gegenstand der Untersuchung

Gegenstand der Untersuchung ist das einzelne Projekt. Befragt werden sollte der jeweilige Projektleiter. In der Interviewphase können weitere Projektteilnehmer hinzugezogen werden.

Alternativ könnten Produkte, Organisationen oder Projektmitarbeiter Gegenstand der Untersuchung sein. In dieser Untersuchung sollen diese möglichen Gegenstände nur am Rande betrachtet werden. (z.B. wird der Gegenstand „Behörde" betrachtet, wenn nach einem organisationsspezifischen VMXT gefragt wird).

Weitere Schritte

- Akquise
- Vertrag
- Beschaffungsvermerk
- Termin: Workshop.

Notizen:

- Fragen zum Umfang des VMXT aufnehmen (z.B.: Sind die 800 Seiten eine Schwelle?)
- Hinterfragen:
  - Welche Stellung hat der Promotor?

Anwenderevaluation des V-Modell XT *

Dieser Fragebogen dient der Evaluierung der Anwendung des V-Modell XT *. Ziel dieses Fragebogens ist es, Probleme in der Anwendung des V-Modell XT * zu ermitteln und Verbesserungspotenziale zu ermitteln. Der Fragebogen setzt sich aus den folgenden Kategorien zusammen:

1. Metadaten (zur Person und zum Prozess)
2. Bereich Prozessmodell (dieser Bereich adressiert die exportierte V-Modell XT Dokumentation)
3. Bereich Produktvorlagen (dieser Bereich adressiert die exportierten Produktvorlagen, z.B. für Berichte)

Bitte beantworten Sie den Fragebogen so vollständig wie möglich! Sollten Sie eine Frage nicht beantworten können, lassen Sie den Antwortbereich bitte frei und setzen Sie keine Kreuze – in diesem Fall nutzen Sie bitte den Freitextbereich, um zu erläutern, warum Sie die Frage nicht beantworten können.

1) Bereich Metadaten

Teilnehmer (Name)	
Rolle im Projekt	
Wie viele Projekte nach V-Modell XT * haben Sie bereits durchgeführt?	

1. Welche Version des V-Modell XT * haben Sie verwendet?

   Versionsnummer: _____

2. Welchen Export der Prozessdokumentation haben Sie verwendet?

PDF: ☐	HTML: ☐	Sonstiges: ☐

3. Haben Sie die Produktvorlagen des V-Modell XT * verwendet?
   (*kreuzen Sie „Ja“ an, wenn Sie die Vorlagen aus dem V-Modell XT * verwendet haben, sonst „Nein“; kreuzen Sie „Tools“ an, wenn Sie die Vorlagen verwendet haben, welche aus den* ***-Werkzeugen (** *A, B, etc.) erzeugt wurden.*)

Ja: ☐	Nein: ☐	Tools: ☐

4. Haben Sie Schulungen zum V-Modell XT * erhalten?
   (*kreuzen Sie „Ja“ an, wenn Sie an einer Schulung teilgenommen haben, sonst „Nein“; kreuzen Sie „Coaching“ an, wenn Sie ein Coaching erhalten haben, unabhängig davon, ob Sie an einer Schulung teilgenommen haben oder nicht.*)

Ja: ☐	Nein: ☐	Coaching: ☐

2) Bereich Prozessmodell

1. Der Export ist übersichtlich und alle wesentliche Informationen sind leicht zugänglich:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

2. Der Export ist für mit dem Vorgehensmodell nicht vertraute („Außenstehende“) leicht zugänglich:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

3. Der Rahmenprozess wird übersichtlich dargestellt und gibt einen guten Überblick über alle wesentlichen Elemente:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

4. Der Export ermöglicht eine zielgerichtete Prüfung der Beziehungen zwischen den Prozesselementen und die Identifikation von Konsistenzverletzungen:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

5. Der Umsetzungsumfang ist zufriedenstellend und alle im modellierten Rahmenprozess konzipierten Elemente wurden vollständig umgesetzt:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

6. Alle konzipierten Rollen wurden umgesetzt:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

7. Alle konzipierten Artefakte wurden umgesetzt:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

8. Alle konzipierten Beziehungen zwischen den Artefakten und den Rollen wurden umsetzt:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

3) Bereich Produktvorlagen

1. Alle erforderlichen Produktvorlagen wurden vollständig umgesetzt:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

2. Die Produktvorlagen sind übersichtlich und alle wesentlichen Informationen sind leicht zugänglich:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

3. Die Produktvorlagen sind für mit dem Vorgehensmodell nicht vertraute („Außenstehende") leicht zugänglich:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

4. Die Produktvorlagen ermöglichen eine zielgerichtete Prüfung der Beziehungen zu weiteren Produkten und die Identifikation von Konsistenzverletzungen:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

5. Welche Anforderungen wurden nicht vollständig erfüllt?

6. Welche Handlungsmaßnahmen würden Sie empfehlen, um diese Anforderungen zu erfüllen?

15. Welche Schulnote geben Sie den Produktvorlagen (Gesamturteil)?

Sehr gut					Ungenügend
1: ☐	2: ☐	3: ☐	4: ☐	5: ☐	6: ☐

9. Alle konzipierten Aktivitäten wurden umsetzt

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

10. Der Rahmenprozess gemäß der Modellierung im Realisierungskonzept wurden vollständig umsetzt:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

11. Das Modell ist inhaltlich konsistent:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

12. Welche Anforderungen wurden nicht vollständig erfüllt?

13. Welche Handlungsmaßnahmen würden Sie empfehlen, um diese Anforderungen zu erfüllen?

14. Welche Schulnote geben Sie der Prozessdokumentation (Gesamturteil)?

Sehr gut					Ungenügend
1: ☐	2: ☐	3: ☐	4: ☐	5: ☐	6: ☐

Technische Universität München

Evaluation of Artefact-based RE Improvement and Quality Indicators

Project: <Text goes here>

Improvement Goals:

<Text goes here>

Date	*<Text goes here>*
Name	*<Text goes here>*
Role	*<Text goes here>*

Fehler! Kein Text mit angegebener Formatvorlage im Dokument.

Evaluation of RE Improvement Approach from the Perspective of the Process Engineers

1.1 Rating of the RE Improvement approach

The following questions concern the evaluation of the overall artefact-based RE improvement approach ("ABREImp"), i.e. the workshops performed to define the new RE reference model of *.

Each of the closed questions considers the process (the workshops and the outcomes) in direct comparison to previously used approaches and / or the alternatives presented by partners.

1. Structuredness	The approach was systematic.								
ABREImp	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	I strongly agree
Previously experienced approach	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	I strongly agree
Please give a rationale for your rating (if possible):									

2. Simplicity	The approach was easy to use.								
ABREImp	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	I strongly agree
Previously experienced approach	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	I strongly agree
Please give a rationale for your rating (if possible):									

3. Goal-Orientation	The approach was goal-oriented and took into account problems and needs of all involved stakeholders.									
ABREImp	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree
Previously experienced approach	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree

Please give a rationale for your rating (if possible):

4. Experience-Orientation	The approach took into account the culture of the company and the experiences of all involved stakeholders.									
ABREImp	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree
Previously experienced approach	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree

Please give a rationale for your rating (if possible):

5. Sustainability	The decisions taken during the workshops were reproducible.									
ABREImp	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree
Previously experienced approach	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree

Please give a rationale for your rating (if possible):

6. Effectivity	The process analysis and the construction of the new RE reference model has lead to the desired results.									
ABREImp	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree
Previously experienced approach	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree

Please give a rationale for your rating (if possible):

7. Efficiency	I perceived the efficiency of the undertaking as high.									
ABREImp	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree
Previously experienced approach	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree

Please give a rationale for your rating (if possible):

8. Knowledge Transfer	The approach actively supported knowledge transfer and a learning curve at all participants.									
ABREImp	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree
Previously experienced approach	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree

Please give a rationale for your rating (if possible):

11 Overall attitude towards the approach / Reasoning for decision	Considering our demands for RE improvement, the chosen approach was better suited than available (known) approaches.								
I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree
Please give a rationale for your ratings (if possible):									
12. If I had to establish an artefact-based RE approach, I would follow again the same procedure Yes: ☐ No: ☐									
12 Final remarks	I want add following positive and negative aspects:								

1.2 General Questions

The following questions consider more general aspects of the approach taking also into account hypothetical situations given by available (known) prescriptive approaches that might not have been yet applied in same context; for example, taking into account prescriptive approaches like CMMI even if the participants have not yet conducted an RE improvement endeavour by using CMMI.

Hint: Prescriptive approaches constitute reference models against which companies are assessed/benchmarked and the improvement consists of adapting the company-specific RE reference to this norm. In contrast, problem-driven approaches follow the approach to adapt the company-specific RE approach against individual problems and needs.

9. RE Improvement Process	Regarding the RE improvement procedure, the following factors were critical for the overall improvement success.									
Goal orientation	Uncritical	☐	☐	☐	☐	☐	☐	☐	☐	Highly critical
Continuous stakeholder involvement	Uncritical	☐	☐	☐	☐	☐	☐	☐	☐	Highly critical
Fast improvement cycles	Uncritical	☐	☐	☐	☐	☐	☐	☐	☐	Highly critical
Establishment of an artefact-based RE reference	Uncritical	☐	☐	☐	☐	☐	☐	☐	☐	Highly critical
Certification	Uncritical	☐	☐	☐	☐	☐	☐	☐	☐	Highly critical
Please give a rationale for your ratings (if possible):										

10. RE Improvement Goals	Regarding the RE improvement outcome, the following factors are critical for the overall improvement success.									
Organization-specific results	Uncritical	☐	☐	☐	☐	☐	☐	☐	☐	Highly critical
Certifiable RE process	Uncritical	☐	☐	☐	☐	☐	☐	☐	☐	Highly critical
Improvements yield early results	Uncritical	☐	☐	☐	☐	☐	☐	☐	☐	Highly critical
Measurable improvement	Uncritical	☐	☐	☐	☐	☐	☐	☐	☐	Highly critical
Artefact-based RE reference	Uncritical	☐	☐	☐	☐	☐	☐	☐	☐	Highly critical
Please give a rationale for your ratings (if possible):										

Evaluation of resulting artefact-based RE Reference Model from the Perspective of the Project Participants

The following questions concern the evaluation of the artefact-based RE reference model ("ABRE") resulting from the previously rated workshops in direct comparison to the RE reference model previously used at the company.

Each of the closed questions considers the new artefact-based RE approach resulting from the workshops in direct comparison to previously used RE approach.

1. Flexibility of Reference Model	The RE reference model allows for flexibility in the process.								
ABRE	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree
Previously experienced approach	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree

Please give a rationale for your rating (if possible):

2. Ease of Use of Reference Model	The RE reference model is easy to understand.									
ABRE	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree
Previously experienced approach	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree

Please give a rationale for your rating (if possible):

3. Effectivity	The application of the RE reference model has lead to the desired results.									
ABREImp	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree
Previously experienced approach	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree

Please give a rationale for your rating (if possible):

4. Efficiency	When applying the RE reference model, I perceived the efficiency as high.									
ABREImp	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree
Previously experienced approach	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree

Please give a rationale for your rating (if possible):

5. Customisation of Reference Model	The RE reference model is tailorable according to project-specific situations of the company, i.e. the artefacts created according to the Tailoring Profile are useful.									
ABRE	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree
Previously experienced approach	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree

Please give a rationale for your rating (if possible):

6. Process Integration of Reference Model	The RE reference model is integrated into further development activities (e.g., testing) and within the line organisation.

ABRE	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree
Previously experienced approach	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree

Please give a rationale for your rating (if possible):

7. Structuredness of Artefacts	The specification documents are well-suited to be understood respecting their structure by people not being involved into the elaboration of the specifications.

ABRE	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree
Previously experienced approach	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree

Please give a rationale for your rating (if possible):

8. Syntactic Artefact Quality	The RE reference model supports a high syntactic quality in the created RE artefacts w.r.t. consistency and completeness. *Hint:* Syntactical quality means that all essential elements describing a requirement are given (e.g. all elements from a pattern based approach are available, this could be a process word or juristic dimension (shall-, should-, will-requirement type)).

ABRE	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree
Previously experienced approach	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree

Please give a rationale for your rating (if possible):

9. Traceability in Artefacts	The RE reference model supports traceability within RE (e.g., through rationales) and between RE and further disciplines. *Hint:* Traceability means both identifiable, unambiguous requirements and also the possibility of being traceable, e.g. from demand over business use case over system use case to test cases.

ABRE	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree
Previously experienced approach	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree

Please give a rationale for your rating (if possible):

10. Semantic Artefact Quality	The RE reference model supports semantically consistent and complete results. Hint: Semantic Completeness means that that the specified requirements cover all functional aspects from the original desired and needed customer requirements.

ABRE	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree
Previously experienced approach	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree

Please give a rationale for your rating (if possible):

11. Testability of Artefacts	The RE reference model supports testable RE artefacts.

ABRE	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree
Previously experienced approach	I strongly disagree	☐	☐	☐	☐	☐	☐	☐	☐	I strongly agree

Please give a rationale for your rating (if possible):

Evaluation of RE Quality Indicators from the Perspective of the Project Participants and Team Leaders

The following questions concern the evaluation of the RE quality indicators defined for the quality gates. As there does not exist a previous quality gate, the evaluation considers the investigation of potential improvements triggered during the RE process in response of applying the quality indicators. To this end, we distinguish two parts in the evaluation:

1. Constructive quality assurance: How did the quality indicators support project participants in their RE process?
2. Analytical quality assurance: How suitable are the indicators in general to support an adequate audit of RE as part of a quality gate?

In both categories, we define closed questions and open questions.

1.3 General Questions / Background and Expectations

1. Overall attitude towards indicators/metrics	In general, indicator measurements are well-suited for quality assessment.
I strongly disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ I strongly agree	
Please give a rationale for your ratings (if possible):	
2. Personal Experiences	My personal (past) experiences with indicator-based measurements are:
Strongly negative ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly positive	
Please give a rationale for your ratings (if possible):	
3. Involvement	My primary role regarding the use of the quality indicators in SWQM level 2 was (select only one):
Quality Gate Keeping assessing the quality of individual specification documents	☐
Requirements Engineer capturing and specifying the requirements using the artefact model	☐
Quality Manager developing quality indicators to improve specification quality in general	☐

4. Personal Influence	I would rate my personal influence on the selection and adaptation of the quality indicators as follows:
No influence ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strong influence	
Please give a rationale for your ratings (if possible):	

1.4 Metric Description

In this section the structured, holistic description of metrics shall be evaluated. It is important to know which information shall be part of a metric, and how these information must be structured in order to optimally support crucial activities when using metrics.

Note that only the description, not the metrics itself shall be evaluated here.

5. Personal Expectations	Rate the following information regarding their importance for successful implementation of quality indicators.									
Relation to quality criteria/goals	Unnecessary	☐	☐	☐	☐	☐	☐	☐	☐	Essential
Associated data collection method	Unnecessary	☐	☐	☐	☐	☐	☐	☐	☐	Essential
Reference value (range)	Unnecessary	☐	☐	☐	☐	☐	☐	☐	☐	Essential
Interpretation of reference value	Unnecessary	☐	☐	☐	☐	☐	☐	☐	☐	Essential
Threads to validity	Unnecessary	☐	☐	☐	☐	☐	☐	☐	☐	Essential
Recommendations for action	Unnecessary	☐	☐	☐	☐	☐	☐	☐	☐	Essential
In case information for the comprehensive description of quality indicators were missing, which one?										

Usefulness of Metric Description	I would rate the usefulness of the metric description itself (not the individual metrics, just the way they are descriped) was helpful to gain a deeper, holistic understanding of the employed metrics.
Strongly disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly agree	
Please give a rationale for your ratings (if possible):	

Please give a rationale for your ratings (if possible):	
2. Adequacy	The reference values (resp. value ranges) adequately partitioned the quality indicator results into the classes of good performance respectively minor/major deviation from it.
Conformity to Artefact Model	I strongly disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ I strongly agree
Scenario/Test Coverage	I strongly disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ I strongly agree
Measurability of Non-functional Req.	I strongly disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ I strongly agree
Diversity of Priorities	I strongly disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ I strongly agree
Clone-Detection	I strongly disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ I strongly agree
Please give a rationale for your ratings (if possible):	
3. Summary	In general, the application of the quality indicators was helpful to assess the quality of the specification.
	I strongly disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ I strongly agree
Please give a rationale for your ratings (if possible):	

Threads to Validity	The threads to validity and needs for actions (Interpretationsrisiken und Handlungsempfehlungen) support a more sensitive use of metrics.
	Strongly disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly agree
Please give a rationale for your ratings (if possible):	
Analystical Use of Metrics	The structured description of metrics supports, especially the interpretation and reference values, is helpful to assess the quality of artefacts.
	Strongly disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly agree
Please give a rationale for your ratings (if possible):	
Need for action	In general, need for actions (Handlungsempfehlungen) of the structured metric description are helpful to constructively employ metrics.
	Strongly disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Strongly agree
Please give a rationale for your ratings (if possible):	

1.5 Analytical Quality Assurance

1. Correctness of Quality Indicators	Improvement respectively degradation (regarding quality criteria) in the specification documents correctly impose positive respectively negative changes in corresponding quality indicators (independent of actual results and reference values).
Conformity to Artefact Model	I strongly disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ I strongly agree
Scenario/Test Coverage	I strongly disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ I strongly agree
Measurability of Non-functional Req.	I strongly disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ I strongly agree
Diversity of Priorities	I strongly disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ I strongly agree
Clone-Detection	I strongly disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ I strongly agree

1.6 Constructive Quality Assurance

1. Suitability	The application of quality indicators (results, interpretation and recommendations for action) provided valuable feedback and guidance during the requirements specification phase.
Conformity to Artefact Model	I strongly disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ I strongly agree
Scenario/Test Coverage	I strongly disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ I strongly agree
Measurability of Non-functional Req.	I strongly disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ I strongly agree
Diversity of Priorities	I strongly disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ I strongly agree
Clone-Detection	I strongly disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ I strongly agree
Please give a rationale for your ratings (if possible):	
2. Summary	In general, the application of quality indicators was helpful to constructively improve the quality of the specification during the requirements elicitation phase.
	I strongly disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ I strongly agree
Please give a rationale for your ratings (if possible):	

1.7 Operationalisation

Here, we evaluate how well the indicator concepts were operationalized.

Tool Support	The used tools (e.g., Excel checklist & metrics) were helpful in assessing the quality (in terms of SWQM Quality Gate I)
	I strongly disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ I strongly agree
Please give a rationale for your ratings (if possible):	

Process Integration	The integration of the metrics into the processes was clearly defined and supported the quality assessment adequately.
	I strongly disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ I strongly agree
Please give a rationale for your ratings (if possible):	

1.8 Finding Requirements Defects

Assessment of poor quality req. documents	When reviewing requirements documents, I can easily decide if a requirement document (as a whole) is **bad**.
	I strongly disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ I strongly agree
Please give a rationale for your ratings (if possible):	
Assessment of high quality req. documents	When reviewing requirements documents, I can easily decide if a requirement document (as a whole) is **good**.
	I strongly disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ I strongly agree
Please give a rationale for your ratings (if possible):	
Automatic support	In my opinion, an automatic support tool for reviewing textual requirements yields the potential to considerably support the assessment of requirements specification documents.
	I strongly disagree ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ I strongly agree
Please give a rationale for your ratings (if possible):	

Fragebogen

Bereich Metadaten

Diese Informationen werden nur für statistische Zwecke erfasst. Sie werden vertraulich behandelt und es erfolgt keine personenbezogene Auswertung der Angaben.

optional		Teilnehmer (Name)			
	1	Projektgegenstand		*falls sonstiges:*	
	2	Projektart		*falls sonstiges:*	
	3	Projektdurchführung			
	4	Umfang des Projekts (in PJ geschätzt)			
	5	Dauer des Projekts (in Jahren geschätzt)			
	6	Ihre (V-Modell-)Rolle im Projekt			
	7	Wie viele Projekte nach V-Modell XT BNetzA haben Sie bereits durchgeführt?		*Anzahl der Projekte hier angeben*	
	8	Haben Sie vor dem V-Modell XT BNetzA bereits mit Vorgänger-Versionen (VM-BNetzA) gearbeitet?			

Informationen zum bewerteten Prozess

1	Welche Version des V-Modell XT BNetzA haben Sie verwendet?		*Versionsnummer hier angeben*	
2	Welchen Export der Prozessdokumentation haben Sie verwendet?		*falls sonstiges:*	
3	Haben Sie die Produktvorlagen des V-Modell XT BNetzA verwendet?			
4	Haben Sie Schulungen zum V-Modell XT BNetzA erhalten?			

„Ja“, wenn Sie die Vorlagen aus dem V-Modell XT BNetzA verwendet haben, sonst „Nein“; „Tools“, wenn Sie Vorlage
V-Modell-Inhalte ‚ersetzen' und die aus den Projektwerkzeugen (z.B. TFS-Work Items) erzeugt wurden.

„Ja“, wenn Sie an einer Schulung teilgenommen haben, sonst „Nein“; „Coaching“, wenn Sie ein (Projekt-)Coaching e
davon, ob Sie an einer Schulung teilgenommen haben oder nicht.

Bereich Prozessmodell und Prozessdokumentation

Kreuzen Sie zu den folgenden Fragen jeweils genau eine Zelle mit "X" an. Können Sie eine Frage nicht beantworten, lassen Sie die Zeile leer.

			Stimme nicht zu							Stimme zu
	1	Der Export ist übersichtlich und alle wesentlichen Informationen sind leicht zugänglich:								
	2	Der Export ist für mit dem Vorgehensmodell nicht vertraute ("Außenstehende") leicht zugänglich:								
	3	Der Rahmenprozess wird übersichtlich dargestellt und gibt einen guten Überblick über alle wesentlichen Elemente:								
	4	Der Export ermöglicht eine zielgerichtete Prüfung der Beziehungen zwischen den Prozesselementen und die Identifikation von Konsistenzverletzungen:								
	5	Der Umfang des V-Modell XT BNetzA ist zufriedenstellend:								
	6	Der Umfang der Rollen ist zufriedenstellend:								
	7	Der Umfang der V-Modell-Produkte ist zufriedenstellend:								
	8	Die Darstellungstiefe (der Detaillierungsgrad) der V-Modell-Produkte ist zufriedenstellend:								
	9	Die Beziehungen zwischen den V-Modell-Produkten und den Rollen sind klar dargestellt:								
	10	Der Umfang der Aktivitäten ist zufriedenstellend:								
	11	Die Darstellungstiefe (der Detaillierungsgrad) der Aktivitäten ist zufriedenstellend:								
	12	Das Modell ist inhaltlich konsistent:								
appropriate	13	Welche Anteile der Prozessdokumentation würden Sie erweitern?								
appropriate	14	Welche Anteile der Prozessdokumentation würden Sie reduzieren?								
appropriate	15	Welche zusätzlichen Inhalte würden Sie dem V-Modell XT BNetzA noch hinzufügen?								
appropriate	16	Welche Handlungsmaßnahmen würden Sie empfehlen, um diese Anforderungen zu erfüllen?								
	17	Welche Schulnote geben Sie der Prozessdokumentation (Gesamturteil)?								

Bereich Produktvorlagen

Kreuzen Sie zu den folgenden Fragen jeweils genau eine Zelle mit "X" an. Können Sie eine Frage nicht beantworten, lassen Sie die Zeile leer.

			Stimme nicht zu							Stimme zu
	1	Alle erforderlichen Produktvorlagen sind vollständig umgesetzt:								
	2	Die Produktvorlagen sind übersichtlich und alle wesentlichen Informationen sind leicht zugänglich:								
	3	Die Produktvorlagen sind für mit dem Vorgehensmodell nicht vertraute („Außenstehende“) leicht zugänglich:								
	4	Die Produktvorlagen ermöglichen eine zielgerichtete Prüfung der Beziehungen zu weiteren Produkten und die Identifikation von Konsistenzverletzungen:								
freitext	5	Welche Anteile der Produktvorlagen würden Sie erweitern?								
freitext	6	Welche Anteile der Produktvorlagen würden Sie reduzieren?								
freitext	7	Welche zusätzlichen Inhalte würden Sie den Produktvorlagen des V-Modell XT BNetzA noch hinzufügen?								
freitext	8	Welche Handlungsmaßnahmen würden Sie empfehlen, um diese Anforderungen zu erfüllen?								
	9	Welche Schulnote geben Sie den Produktvorlagen (Gesamturteil)?								

Nr.		Bereich	Frage	Metrik	
		Fragebogen	Gegenstand dieser Studie sind die Erfahrungen, die im Umgang mit dem V-Modell XT im Projektalltag gemacht wurden. Bitte legen Sie für die Beantwortung der Fragen ein (beliebiges) V-Modell-XT-Projekt zugrunde. Sollten Sie bereits mehrere Projekte durchgeführt haben, möchten wir Sie bitten, den Fragebogen auch mehrmals (einmal für jedes Projekt) auszufüllen. Vielen Dank!	Ggf. sollten wir noch ein Feld vorsehen, in dem die Leute kenntlich machen können, ob Sie den Fragebogen mehrfach ausfüllen... Vielleicht eine Projektnummer oder so?	
		Metadaten	Wurde ein V-Modell-Projekt durchgeführt?	Ja/Nein	
			Welchem Projekttyp entsprach das durchgeführte Projekt?	AG/AN/AG-AN/ORG	
			Was war der Projektgegenstand?	SW/HW/SW-HW/Prozess/Sonst.	Sonst.: z.B. eine Studie
			Was war die Laufzeit des Projekts in Personenmonaten (PM)?	Zahl	
			In welcher Behörde wurde das Projekt durchgeführt?	Text	
			Welche Position nehmen Sie in Ihrer Behörde ein?	Text	
			Welche Position nahmen/nehmen Sie im hier betrachteten Projekt ein?	Text	
			Bitte geben Sie uns Ihre Kontaktinformationen (optional), sodass wir Sie über das Ergebnis der Studie informieren und ggf. Rückfragen stellen können.	3x Text (Name, Email, Telefon)	
		Bereich	*Frage*		
Nr.	1	Organisatorisches		Metrik	Nein: Jeweils + Begründungsoption (Text)
1	1.1		Gibt es einen V-Modell-Verantwortlichen in Ihrer Behörde?	Ja/Nein	
2	1.2		Ist das V-Modell für Sie verpflichtend anzuwenden?	Ja/Nein	
	2	Effizienz		Metrik	
3	2.1		Waren die Vorgaben des V-Modells ausreichend für Ihr Projekt?	Ja/Nein	
4	2.2		Wurde Ihr Projekt in der veranschlagten Zeit abgeschlossen?	Ja/Nein	
5	2.3		Wurde Ihr Projekt erfolgreich abgeschlossen?	Ja/Nein	
6	2.4		Ist die Projektdokumentation ein Grund für Verzögerungen?	Ja/Nein	
7	2.5		Wie beurteilen Sie den Einsatz des V-Modell XT im Vergleich zu Ihren vorherigen Projekten ohne V-Modell-XT-Einsatz bzgl. des Aspekts Flexibilität?	1..5	1 = schlecht, 3 = neutral, 5 = deutlich besser
8	2.6		Wie beurteilen Sie den Einsatz des V-Modell XT im Vergleich zu Ihren vorherigen Projekten ohne V-Modell-XT-Einsatz bzgl. des Aspekts Mehraufwand?	1..5	
9	2.7		Wie beurteilen Sie den Einsatz des V-Modell XT im Vergleich zu Ihren vorherigen Projekten ohne V-Modell-XT-Einsatz bzgl. des Aspekts Kommunikation?	1..5	
	3	Wissensstand		Metrik	
10	3.1		Haben Sie oder Ihre Mitarbeiter V-Modell-Schulungen in Anspruch genommen?	Ja/Nein	
11	3.2		Hat Ihr Projekt ein spezielles V-Modell-Coaching erhalten?	Ja/Nein	
	4	Dürchführung		Metrik	
12	4.1		Nutzen Sie Werkzeuge, um die Projektdurchführung zu unterstützen?	Ja/Nein	Hier auch bei Ja ein Textfeld mit der Frage: Welche?
13	4.2		Waren die Produktvorlagen des V-Modells für Ihr Projekt geeignet?	Ja/Nein	
14	4.3		Wie wird die Einhaltung des V-Modells kontrolliert?	Ja/Nein	
15	4.4		Wurden regelmäßige Abstimmungen zwischen Auftraggeber und Auftragnehmer einberufen?	Ja/Nein	
16	4.5		Haben Sie sich bewusst gegen die Anwendung von Regeln im V-Modell entschieden?	Ja/Nein	
17	4.6		Wurde für Ihr Projekt ein Tailoring durchgeführt?	Ja/Nein	
18	4.7		War die Qualität des Projektergebnisses voll zufrieden stellend?	Ja/Nein	
19	4.8		Haben Sie sich bewusst gegen die Anwendung von Regeln im V-Modell XT entschieden?	Ja/Nein	
	5	Motivation		Metrik	
20	5.1		Hat sich Ihr Projektteam eigenständig für den Einsatz des V-Modells entschieden?	Ja/Nein	
21	5.2		Haben Sie von Ihnen favorisierte Entscheidungen im Projekt wegen der Vorgaben des V-Modells verworfen und eine andere getroffen?	Ja/Nein	
22	5.3		Ist das V-Modell dazu geeignet, Ihren etablierten Entwicklungsprozess abzubilden?	Ja/Nein	
23	5.4		Würden Sie das V-Modell erneut anwenden?	Ja/Nein	

B.6 Artifact Model Evaluation (GloBuS)

Fragebogen

Bereich Metadaten (siehe Auswahloptionen für Antworten)

Diese Informationen werden nur für statistische Zwecke erfasst. Sie werden vertraulich behandelt und es erfolgt keine personenbezogene Auswertung der Angaben.

	Teilnehmer (Name)			
1	Üblicher Projektgegenstand (Software, Hardware, Dienstleistung, Sonstiges)		*falls sonstiges:*	
2	Übliche Projektart (Neuentwicklung, Weiterentwicklung, Sonstiges)		*falls sonstiges:*	
3	Übliche Projektdurchführung (Vergabe, In-House, Gemischt)			
4	Üblicher Umfang des Projekts (in PJ geschätzt)			
5	Üblicher Dauer des Projekts (in Jahren geschätzt)			
6	Vorgehensmodelle (z.B. V-Modell XT, Scrum, RUP, …)			
7	Wie viele Projekte haben Sie bereits durchgeführt?		*Anzahl der Projekte hier angeben*	

Modellbewertung (absolut)

	Kreuzen Sie zu den folgenden Fragen jeweils genau eine Zelle mit "X" an. Können Sie eine Frage nicht beantworten, lassen Sie die Zeile leer.	Stimme nicht zu							Stimme zu
1	Das Modell ist vollständig.								
2	Der Modell ist verständlich.								
3	Das Modell ist präzise.								
4	Das Modell gibt einen guten Überblick über alle wesentlichen Elemente und deren Zusammenhänge.								
5	Die Darstellungstiefe (der Detaillierungsgrad) des Modells ist zufriedenstellend.								
6	Das Modell ist inhaltlich konsistent (widerspruchsfrei).								
7	Das Modell ermöglicht Flexibilität.								
8	Das Modell ist eindeutig.								
9	Das Modell sind inhaltlich richtig.								
10	Die Benennung der Klassen ist verständlich.								
11	Die Beziehungen zwischen den Klassen sind inhaltlich richtig.								

12	Welche Teile des Modells würden Sie erweitern? Welche zusätzlichen Inhalte würden Sie dem Modell noch hinzufügen?	
13	Welche Teile des Modells würden Sie reduzieren?	
14	Begründungen zu den oben stehenden Bewertungen (Freitext)	
15	Welche Schulnote geben Sie dem Modell (Gesamturteil)?	

Modellbewertung (relativ zu bislang eingesetzten Verfahren)

	Kreuzen Sie zu den folgenden Fragen jeweils genau eine Zelle mit "X" an. Können Sie eine Frage nicht beantworten, lassen Sie die Zeile leer.	Stimme nicht zu							Stimme zu
16	Das Modell vereinfacht die Kommunikation zwischen Projektpartnern und Vertragsparteien.								
17	Das Modell vereinfacht die Kollaboration zwischen Projektpartnern und Vertragsparteien.								
18	Das Modell vereinfacht den Datenauschtausch zwischen Projektpertnern und Vertragsparteien.								
19	Das Modell ist flexibler als bisherige Ansätze.								
20	Das Modell ist besser strukturiert als bisherige Ansätze.								
21	Das Modell ist besser für eine Werkzeugunterstützung geeignet als bisherige Ansätze.								
22	Das Modell ist vollständiger als bisherige Ansätze.								
23	Der Detaillierungsgrad des Modells ist höher als in bisherigen Ansätzen.								
24	Der Detaillierungsgrad des Modells hat insgesamt Vorteile gegenüber bisherigen Ansätzen.								
25	Die verwendete Terminologie ist konsistenter als in bisherigen Ansätzen.								
26	Der Detaillierungsgrad der Beziehungen ist höher als in bisherigen Ansätzen.								
27	Der Detaillierungsgrad der Beziehungen hat insgesamt Vorteilte gegenüber bisherigen Ansätzen.								

28	Welche Schulnote geben Sie dem Modell im Vergleich zu bisherigen Ansätzen (Referenz bisheriger Ansatz: 3,0)?	

Übungsblatt 7.1: Fragebogen für Aufgabe 12

Gruppenteilnehmer	
Verwendetes Prozessframework	

1. Warum haben Sie sich (zu Beginn) für das verwendete Prozessframework entschieden?

2. Das verwendete Werkzeug war intuitiv anwendbar und das Vorgehen zur Prozessmodellierung war jederzeit klar:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

3. Ich fand mich schnell im Werkzeug zurecht und konnte alle notwendigen Funktionen zur Umsetzung der konzipierten Prozesselemente schnell verwenden.

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

4. War das Werkzeug Ihrer Meinung nach geeignet, um das konzipierte Modell umzusetzen? Wenn nicht, warum?

5. Das verwendete Werkzeug war geeignet, um das konzipierte Modell mit wenig Overhead umzusetzen:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

Übung zur Vorlesung: Vorgehensmodelle im Software Engineering
Wintersemester 2011/2012
Dr. Marco Kuhrmann
Georg Kalus, Dr. Daniel Méndez Fernández
12. Juni 2014

Technische Universität München
Fakultät für Informatik

6. Ich war mir jederzeit über die Konsequenzen der getroffenen Realisierungsentscheidungen bewusst:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

7. Die Prozessdokumentation konnte bedarfsgerecht erzeugt werden:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

8. Der Umsetzungsumfang war zufriedenstellend und ich konnte alle im modellierten Rahmenprozess konzipierten Elemente, wenn auch nicht in geplanter Detailtiefe, vollständig umsetzen:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

9. Ich konnte alle konzipierten Rollen umsetzen:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

10. Ich konnte alle konzipierten Artefakte umsetzen:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

11. Ich konnte alle konzipierten Beziehungen zwischen den Artefakten und den Rollen umsetzen:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

12. Ich konnte alle konzipierten Aktivitäten umsetzen

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

13. Ich konnte den Rahmenprozess gemäß der Modellierung im Realisierungskonzept vollständig umsetzen:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

14. Ich konnte jederzeit mein Modell exportieren:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

15. Das eingesetzte Werkzeug war bei der Umsetzung zielführend und ich konnte alle gestellten Anforderungen damit erfüllen:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

16. Ich konnte Konsistenzprobleme jederzeit identifizieren / einfach prüfen:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

17. Wenn Sie die in den vergangenen Workshops erstellten Konzepte erneut umsetzen müssten, würden Sie sich für das gleiche Werkzeug entscheiden? Warum?

18. Welche Aspekte hätten Sie im Rahmen dieses Audits Ihrer Meinung nach berücksichtigt werden müssen?

Übung zur Vorlesung: Vorgehensmodelle im Software Engineering
Wintersemester 2011/2012
Dr. Marco Kuhrmann
Georg Kalus, Dr. Daniel Méndez Fernández
12. Juni 2014

Technische Universität München
Fakultät für Informatik

Übungsblatt 7.2: Fragebogen für Aufgabe 13 und 14

Gruppenteilnehmer	
Analysiertes Modell	

1. Der HTML-Export ist übersichtlich und alle wesentliche Informationen sind leicht zugänglich:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

2. Der HTML-Export ist für mit dem Vorgehensmodell nicht vertraute („Außenstehende") leicht zugänglich:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

3. Der Rahmenprozess wird übersichtlich dargestellt und gibt einen guten Überblick über alle wesentlichen Elemente:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

4. Der Export ermöglicht eine zielgerichtete Prüfung der Beziehungen zwischen den Prozesselementen und die Identifikation von Konsistenzverletzungen:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

5. Der Umsetzungsumfang ist zufriedenstellend und alle im modellierten Rahmenprozess konzipierten Elemente wurden vollständig umgesetzt:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

6. Alle konzipierten Rollen wurden umgesetzt:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

Übung zur Vorlesung: Vorgehensmodelle im Software Engineering
Wintersemester 2011/2012
Dr. Marco Kuhrmann
Georg Kalus, Dr. Daniel Méndez Fernández
12. Juni 2014

Technische Universität München
Fakultät für Informatik

7. Alle konzipierten Artefakte wurden umgesetzt:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

8. Alle konzipierten Beziehungen zwischen den Artefakten und den Rollen wurden umsetzt:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

9. Alle konzipierten Aktivitäten wurden umsetzt

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

10. Der Rahmenprozess gemäß der Modellierung im Realisierungskonzept wurden vollständig umsetzt:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

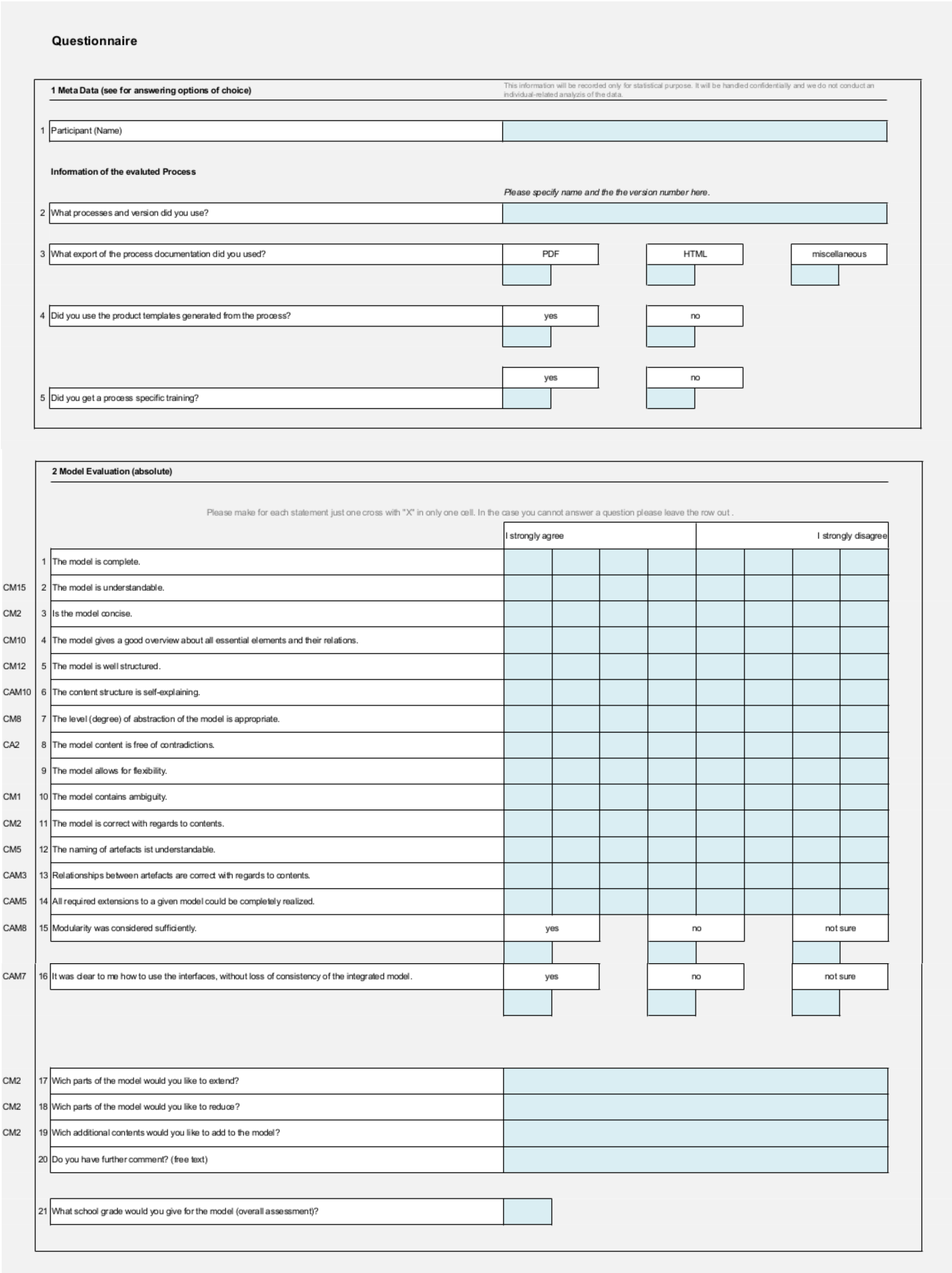
11. Das Modell ist inhaltlich konsistent:

Stimme nicht zu							Stimme zu
☐	☐	☐	☐	☐	☐	☐	☐

12. Welche Anforderungen wurden nicht vollständig erfüllt?

13. Welche Handlungsmaßnahmen würden Sie empfehlen, um diese Anforderungen zu erfüllen?

Appendix C



Questionnaire

1 Meta Data (see for answering options of choice)

This information will be recorded only for statistical purpose. It will be handled confidentially and we do not conduct an individual-related analyzis of the data.

1	Participant (Name)	

Information of the evaluted Process

Please specify name and the the version number here.

2	What processes and version did you use?			
3	What export of the process documentation did you used?	PDF	HTML	miscellaneous
4	Did you use the product templates generated from the process?	yes	no	
5	Did you get a process specific training?	yes	no	

2 Model Evaluation (absolute)

Please make for each statement just one cross with "X" in only one cell. In the case you cannot answer a question please leave the row out .

			I strongly agree							I strongly disagree
	1	The model is complete.								
CM15	2	The model is understandable.								
CM2	3	Is the model concise.								
CM10	4	The model gives a good overview about all essential elements and their relations.								
CM12	5	The model is well structured.								
CAM10	6	The content structure is self-explaining.								
CM8	7	The level (degree) of abstraction of the model is appropriate.								
CA2	8	The model content is free of contradictions.								
	9	The model allows for flexibility.								
CM1	10	The model contains ambiguity.								
CM2	11	The model is correct with regards to contents.								
CM5	12	The naming of artefacts ist understandable.								
CAM3	13	Relationships between artefacts are correct with regards to contents.								
CAM5	14	All required extensions to a given model could be completely realized.								

CAM8	15	Modularity was considered sufficiently.	yes	no	not sure
CAM7	16	It was clear to me how to use the interfaces, without loss of consistency of the integrated model.	yes	no	not sure

CM2	17	Wich parts of the model would you like to extend?	
CM2	18	Wich parts of the model would you like to reduce?	
CM2	19	Wich additional contents would you like to add to the model?	
	20	Do you have further comment? (free text)	
	21	What school grade would you give for the model (overall assessment)?	

Figure C.1: Meta Data and Model Evaluation (absolute)

3 Process Model and Process Documentation

Please make for each question just one cross with "X" in only one cell. In the case you cannot answer a question please leave the row out .

			I strongly agree							I strongly disagree
	1	The export is clear an all essential information is easily accessible.								
CM7	2	The export is easily accessible for persons not familiar with the process model ("outsiders").								
CM12	3	The overall process is clearly represented and gives a good overview of all essential elements.								
	4	The export allows for straightforward inspection of releationships between process elements and the identification of consistency violations.								
CM9	5	The scope of the process model is satisfying.								
CM2	6	The number of the roles is appropriate.								
	7	The scope of the process artifacts is satisfying.								
CM8	8	The level (degree) of abstraction of the process artifacts is appropriate.								
	9	Relationships between process artifacts and roles are clearly represented.								
	10	The scope of the activities (tasks) is satisfying.								
	11	The level (degree) of abstraction of the activities (tasks) is appropriate.								
CM4	12	The process documentation is consistent with the process model.								
CAM9	13	Replaceable process parts could be properly designed and realized.								
	14	The process model is a meaningful representation of the target domain and fulfills the functional requirement.	yes			no			not sure	
CM2	15	Wich parts of the process documentation would you like to extend?								
CM2	16	Wich parts of the process documentation would you like to reduce?								
CM2	17	Wich additional contents would you like to add to the process model?								
CM2	18	What actions would recommend to meet these demands?								
	19	What school grade would you give for the process model (overall assessment)?								

Figure C.2: Process Model and Process Documentation

4 Tool Support

Please make for each question just one cross with "X" in only one cell. In the case you cannot answer a question please leave the row out .

			I strongly agree							I strongly disagree
CM11	1	The used tool is intuitively applicable and the approach of process modeling is straightforward.								
CM14	2	I was immediately familiar with the tool and could instantly use all necessary functionality for the realization of the designed process elements.								
CM14	3	The tool was useful to realize the designed model.								
	4	If not, why?								
CM14	5	The used tool was useful to realize the model with little overhead.								
	6	I was aware about the consequences of the taken realization decision at any time.								
CM13	7	All required tailoring profiles could be completely realized.								
	8	The process documentation could be created for the respective context.								
CM3	9	I could completely realize all designed elements of the modeled overall process.								
CM3	10	I could realize all roles.								
CM3	11	I could realize all artefacts.								
CM3	12	I could realize all realationsships between artefacts and roles.								
CM3	13	I could realize all activities (tasks).								
CM6	14	According to the designs in the realization concept, I could completely realize the overall process.								
	15	I could export the model at any time.								
CM14	16	The used tool was straightforward and all requirements could be realized.								
CAM1	17	I could identify/easily check consitency problems at any time.								
	18	If I had the choice I would coose the same tool.	yes			no			not sure	
	19	Why?								
	20	I could realize all process requirements.								
	21	What process requirements were not completely realized?								
	22	What would be necessary to realize missing/incomplete requirements?								

5 Product Templates

Please make for each question just one cross with "X" in only one cell. In the case you cannot answer a question please leave the row out .

			I strongly agree							I strongly disagree
	1	All product templates are completely realized.								
	2	The product templates are clear and all essential information is easily accessible.								
CM7	3	The product templates are easily accessible for persons not familiar with the process model ("outsiders").								
CAM1	4	The product templates allow for direct inspection of releationships to further products and the identification of consistency violations.								
CM2	5	Which parts of the product templates would you like to extend?								
CM2	6	Which parts of the product templates would you like to reduce?								
CM2	7	Which additional contents from the process would you like to add to product templates?								
	8	What would be necessary to realize missing/incomplete requirements?								
	9	What school grade would you give for the product templates (overall assessment)?								

Figure C.3: Tool Support and Product Templates

CAM4 **1 Model Evaluation (relatively to the approaches used so far - comparative evaluation only)**

Please make for each question just one cross with "X" in only one cell. In the case you cannot answer a question please leave the row out .

			I strongly agree							I strongly disagree
	1	The model simplyfies the communication between stakeholders (project partners and contracting party).								
	2	The model simplyfies the collaboration between stakeholders (project partners and contracting party).								
	3	The model simplyfies the data exchange between stakeholders (project partners and contracting party).								
	4	The model is more felxible than approaches used so far.								
	5	The model is better structured than approaches used so far.								
	6	The model is better suitable for tool support than approaches used so far.								
	7	The model is more complete than approaches used so far.								
	8	The level of abstraction in this model is higher than in approaches used so far.								
	9	Taken as a whole, the level of abstraction in this model has advantages compared to approaches used so far.								
CAM2	10	The used terminology is more consistent than in approaches used so far.								
	11	The level of abstraction of relationships is higher than in approaches used so far.								
	12	Taken as a whole, the level of abstraction of relationships has advantages compared to approaches used so far.								

13	What school grade would you give for the model in comparison with approaches used so far (overall assessment)?	1 (very good)

Figure C.4: Model Evaluation (relatively to the approaches used so far - comparative evaluation only)

Bibliography

[Abr00] ABRAHAMSSON, PEKKA: *Measuring the Success of Software Process Improvement : The Dimensions*. 2000.

[Ara94] ARANGO, GUILLERMO: *A Brief Introduction to Domain Analysis*. In *Proceedings of the 1994 ACM Symposium on Applied Computing*, SAC '94, pages 42–46, New York, NY, USA, 1994. ACM.

[Bas] BASILI, V. R.: *Models and Metrics for Software Management and Engineering.pdf*.

[BD] BHAT, JYOTI M and NIVEDITA DESHMUKH: *Methods for Modeling Flexibility in Business Processes*.

[BDE+05] BERANDER, PATRIK, LARS-OLA DAMM, JEANETTE ERIKSSON, TONY GORSCHEK, KENNET HENNINGSSON, PER JÖNSSON, SIMON KÅ GSTRÖM, DRAZEN MILICIC, FRANS MÅ RTENSSON, KARI RÖNKKÖ, PIOTR TOMASZEWSKI, LARS LUNDBERG, MICHAEL MATTSSON and CLAES WOHLIN: *Software quality attributes and trade-offs*. (June), 2005.

[Ber] BERINGER, FRED: *Software quality metrics and model*.

[Ber00] BERRY, MICHAEL: *An Instrument for Assessing Software Measurement Programs*. 200:183–200, 2000.

[BF] BAND, SCHRIFTENREIHE and GUIDO FISCHERMANNS: *Praxishandbuch Prozessmanagement*.

[BGM87] BENBASAT, IZAK, DAVID K. GOLDSTEIN and MELISSA MEAD: *The Case Research Strategy in Studies of Information Systems*. MIS Quarterly, (September):369—-386, 1987.

[BH09] BUBER, RENATE and HARTMUT H. HOLZMÜLLER (editors): *Qualitative Marktforschung*. Gabler-Lehrbuch. Gabler, Wiesbaden, 2., überarb. Aufl. edition, 2009.

[Bro12] BROY, MANFRED: *Models and Metrics for Software Management and Engineering.pdf*, 2012.

[CGM98] CUGOLA, GIANPAOLO, CARLO GHEZZI and POLITECNICO MILANO: *Software Processes : a Retrospective and a Path to the Future 1*. (June):14–17, 1998.

[CGP+06] CANFORA, GERARDO, FÉLIX GARCÍA, MARIO PIATTINI, FRANCISCO RUIZ and C. A. VISAGGIO: *Applying a framework for the improvement of software process maturity*. Software: Practice and Experience, 36(3):283–304, March 2006.

[CHWL] COPLIEN, JAMES, DANIEL HOFFMAN, DAVID WEISS and BELL LABS: *Commonality and Variability in Software Engineering*. (December 1998):37–45.

[CMU95] CARNEGIE MELLON UNIVERSITY, CORPORATE: *The Capability Maturity Model: Guidelines for Improving the Software Process*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1995.

[CS] CORBIN, JULIET and ANSELM STRAUSS.

[CTBP07] CRESWELL, JOHN W., ABBAS TASHAKKORI, PATRICIA BAZELEY and VICKI L. PLANO CLARK: *Call for Papers of the Journal of Mixed Methods Research*. Journal of Mixed Methods Research, 2007.

[DG81] DENICOFF, MARVIN and ROBERT GRAFTON: *Software Metrics : A Research Initiative*. 1981.

[DHL] DIFFERDING, CHRISTIANE, BARBARA HOISL and CHRISTOPHER M LOTT: *Technology Package for the Goal Question Metric Paradigm*. (April 1996).

[DTD12] DIKICI, AHMET, OKTAY TURETKEN and ONUR DEMIRORS: *A Case Study on Measuring Process Quality: Lessons Learned*. 2012 38th Euromicro Conference on Software Engineering and Advanced Applications, pages 294–297, September 2012.

[Dun95] DUNCAN, N.B.: *Capturing flexibility of information technology infrastructure: A study of resource characteristics and their measure*. Journal of Management Information Systems, 12(2):37–57, 1995.

[Dyb05] DYBA, TORE: *An Empirical Investigation of the Key Factors for Success in Software Process Improvement*. 31(5):410–424, 2005.

[ELPC12] EDUARDO, JAVIER, LÓPEZ LÓPEZ, ANGELA PAOLA and GARAY CIFUENTES: *Software process improvement as an organizational self-production process*. pages 1–10, 2012.

[ESSD08] EASTERBROOK, STEVE, JANICE SINGER, MARGARET-ANNE STOREY and DANIELA DAMIAN: *Selecting Empirical Methods for Software Engineering Research*. In SHULL, FORREST, JANICE SINGER and DAGI.K. SJ¿BERG (editors): *Guide to Advanced Empirical Software Engineering*, pages 285–311. Springer London, 2008.

[FFM88] FUGGEFFA, ALFONSO, ALFONSO FUGGETTA and POLITECNICO MILANO: *Software Process : A Roadmap Software Process : A Roadmap*. 97, 1988.

[Fin03] FINK, ARLENE: *The Survey Handbook*. Sage Books, 2 edition, 2003. Volume 1 of the Survey Kit.

[Fle] FLEMING, IAN: *Software Metrics*.

[FM07] FOSTER, TOM and MARCELLO MAMINO: *An introduction to o-minimality*. (September):1–13, 2007.

[Fra] FRANCINO, YVETTE: *Software quality attributes and their rankings*.

[Fra86] FRANZ, BY CHARIES R: *User Response to an Online Information System : A Field Experiment*. (March):29–43, 1986.

[GA06] GALIN, DANIEL and MOTTI AVRAHAMI: *feature Analyzing Past Studies*. (December), 2006.

[GRCP03] GARCÍA, FÉLIX, FRANCISCO RUIZ, JOSÉ ANTONIO CRUZ and MARIO PIATTINI: *Integrated Measurement for the Evaluation and Improvement of Software Processes*. In OQUENDO, FLAVIO (editor): *Software Process Technology*, volume 2786 of *Lecture Notes in Computer Science*, pages 94–111. Springer Berlin Heidelberg, 2003.

[Gro07] GROUP, SOFTWARE ENGINEERING: *Guidelines for performing Systematic Literature Reviews in Software Engineering*. 2007.

[GS99] GLASER, BARNEY G. and ANSELM L. STRAUSS: *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine de Gruyter, New York, NY, 1999.

[HLM11] HIHN, JAIRUS, SCOTT LEWICKI and SCOTT MORGAN: *Bootstrapping Process Improvement Metrics: CMMI Level 4 Process Improvement Metrics in a Level 3 World*. 2011 44th Hawaii International Conference on System Sciences, pages 1–10, January 2011.

[Hum89] HUMPHREY, WATTS S.: *Managing the Software Process*. Number ISBN: 0-201-18095-2 in *Software Engineering*. Addison-Wesley Publishing Company, 1989.

[IN06] IVERSEN, JAKOB and OJELANKI NGWENYAMA: *Problems in measuring effectiveness in software process improvement: A longitudinal study of organizational change at Danske Data*. International Journal of Information Management, 26(1):30–43, February 2006.

[JB] JONES, CAPERS and OLIVIER BONSIGNOUR: *The Economics of Software Quality*.

[Jes] JESSE, ROLAND: *Script-based Presentation of Simulation Results*.

[JJ] JUERGENS, ELMAR and ELMAR JUERGENS: *Why and How to Control Cloning in Software Artifacts Institut für Informatik Why and How to Control Cloning in Software Artifacts*.

[JP] JAYASWAL, BIJAY and PETER PATTON: *Software Quality Metrics - Developer.com - Developer.com*.

[Kan02] KAN, STEPHEN H.: *Metrics and Models in Software Quality Engineering , Second Edition*. 2002.

[KDK13] KUHRMANN, MARCO, M DANIEL and ALEXANDER KNAPP: *Who Cares About Software Process Modelling ? A First Investigation About the Perceived Value of Process Engineering and Process Consumption*. pages 138–152, 2013.

[Kit] KITCHENHAM, BARBARA: *Procedures for Performing Systematic Reviews*.

[KP08] KITCHENHAM, BARBARAA. and SHARIL. PFLEEGER: *Personal Opinion Surveys*. In SHULL, FORREST, JANICE SINGER and DAGI.K. SJ¿BERG (editors): *Guide to Advanced Empirical Software Engineering*, pages 63–92. Springer London, 2008.

[KPB+09] KITCHENHAM, BARBARA, O. PEARL BRERETON, DAVID BUDGEN, MARK TURNER, JOHN BAILEY and STEPHEN LINKMAN: *Systematic literature reviews in software engineering Ð A systematic literature review*. Information and Software Technology, 51(1):7–15, January 2009.

[KS] KRITERIEN, RAHMENKONZEPT and BRUNO SCHIENMANN: *Evaluation von Objektmodellen*.

[KTF09] KUHRMANN, MARCO, THOMAS TERNITÈ and JAN FRIEDRICH: *Das V-Modell XT anpassen*. Informatik im Fokus. Springer, 2009. (to appear).

[Kuh13] KUHRMANN, MARCO: *ArSPI: An Artifact Model for Software Process Improvement and Management*. Technical Report TUM-I1337, 2013.

[Kur] KURTEV, IVAN: *Adaptability of Model Transformations*.

[Mac00] MACPHERSON, DUGALD: *Notes on o-Minimality and Variations*. 39:97–130, 2000.

[MAKS12] MÜNCH, JÜRGEN, OVE ARMBRUST, MARTIN KOWALCZYK and MARTÍN SOTO: *Software Process Definition and Management*. The Fraunhofer IESE Series on Software and Systems Engineering. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[MCWO00] MAYRHAUSER, A V O N, FORT COLLINS, C WOHLIN and M C OHLSSON: *Assessing and Understanding Efficiency and Success of Software Production*. 154:125–154, 2000.

[MFHT02] MAHDY, AHMED, MOHAMED E. FAYAD, HAITHAM HAMZA and PEEYUSH TUGNAWAT: *Stable and Reusable Model-Based Architectures*. In *Proceedings of the ECOOP 2002 12th Workshop on Model-based Software Reuse*, 2002.

[MFS] MAHDY, AHMED, MOHAMED E FAYAD and COMPUTER SCIENCE: *A Software Stability Model Pattern*.

[MMACM12] MESQUIDA, ANTONI LLUÍS, ANTONIA MAS, ESPERANÇA AMENGUAL and JOSE A. CALVO-MANZANO: *IT Service Management Process Improvement based on ISO/IEC 15504: A systematic review*. Information and Software Technology, 54(3):239–247, March 2012.

[Obj08] OBJECT MANAGEMENT GROUP INC.: *Software & Systems Process Engineering Meta-Model Specification*. (April), 2008.

[OHK+89] OLSON, TIMOTHY G, WATTS S HUMPHREY, DAVID KITSON, TIMOTHY G OLSON, WATTS S HUMPHREY and DAVE KITSON: *Conducting SEI-Assisted Software Process Assessments Conducting SEI-Assisted*. 1989.

[OR12] O'LEARY, P. and I. RICHARDSON: *Process reference model construction: implementing an evolutionary multi-method research approach*. IET Software, 6(5):423, 2012.

[PGP07] PINO, FRANCISCO J., FÉLIX GARCÍA and MARIO PIATTINI: *Software process improvement in small and medium software enterprises: a systematic review*. Software Quality Journal, 16(2):237–261, November 2007.

[PK01] PFLEEGER, SHAD LAWRENCE and BARBARA A KITCHENHAM: *Principles of Survey Research Part 1 " Turning Lemons into Lemonade Systems / Software , I n c . What our series will discuss Description of the Lethbridge survey*. 26(6):16–18, 2001.

[PlGFF] PERNSTÅ L, JOAKIM, TONY GORSCHEK, ROBERT FELDT and DAN FLORÉN: *Software Process Improvement in Inter-departmental Development of Software-Intensive Automotive Systems Ð A Case Study*. pages 93–107.

[PMMP08] PITT, MARK A, JAY I MYUNG, MAXIMILIANO MONTENEGRO and JAMES POOLEY: *Measuring model flexibility with parameter space partitioning: an introduction and application example.* Cognitive science, 32(8):1285–303, December 2008.

[Pra97] PRAKASH, NAVEEN: *Towards a formal definition of methods*. Requirements Engineering, 2(1):23–50, March 1997.

[Pre] PRESSMAN, ROGER S.: *R.S. Pressman & Associates, Inc.*

[PW10] PETERSEN, KAI and CLAES WOHLIN: *Electronic Research Archive of Blekinge Institute of Technology Title : The Effect of Moving from a Plan-Driven to an Incremental and Agile Development Approach : An Industrial Case Study Author : Year : Issue : Pagination :*. pages 654–693, 2010.

[Rei07] REICH, K. (HRSG.): *Brainstorming 1.* pages 1–9, 2007.

[RH01] RAINER, AUSTEN and TRACY HALL: *An analysis of some ?core studies? of software process improvement*. Software Process: Improvement and Practice, 6(4):169–187, December 2001.

[RH08] RUNESON, PER and MARTIN HÖST: *Guidelines for conducting and reporting case study research in software engineering*. Empirical Software Engineering, 14(2):131–164, December 2008.

[Rob02] ROBSON, COLIN: *Real World Research - A Resource for Social Scientists and Practitioner-Researchers*, 2002.

[RR03] RAVICHANDRAN, T. and A. RAI: *Structural analysis of the impact of knowledge creation and knowledge embedding on software process capability*. IEEE Transactions on Engineering Management, 50(3):270–284, August 2003.

[Sea] SEAMAN, CAROLYN B: *Qualitative Methods in Empirical Studies of Software Engineering*. pages 1–48.

[SM10] SULAYMAN, MUHAMMAD and EMILIA MENDES: *Quantitative assessments of key success factors in software process improvement for small and medium web companies*. Proceedings of the 2010 ACM Symposium on Applied Computing - SAC '10, page 2319, 2010.

[SNJ+07] STAPLES, MARK, MAHMOOD NIAZI, ROSS JEFFERY, ALAN ABRAHAMS, PAUL BYATT and RUSSELL MURPHY: *An exploratory study of why organizations do not adopt CMMI*. Journal of Systems and Software, 80(6):883–895, June 2007.

[Soc] SOCK H. CHUNG, R. KELLY RAINER, JR., BRUCE R. LEWIS: *The Impact of Information Technology Infrastructure Flexibility*.

[TKM+] TSIEN, CHRISTINE L, ISAAC S KOHANE, NEIL MCINTOSH, HOSPITAL INFORMATICS PROGRAM, BOSTON MA and ROYAL INFIRMARY: *Building ICU Artifact Detection Models With More Data in Less Time BWH / MGH Harvard Affiliated Emergency Medicine Residency , Boston MA . b Children Õ s*.

[TT06] TEDDLIE, CHARLES and ABBAS TASHAKKORI: *A General Typology of Research Designs Featuring Mixed Methods 1*. 13(1):12–28, 2006.

[UMG+12] UNTERKALMSTEINER, MICHAEL, STUDENT MEMBER, TONY GORSCHEK, A K M MOINUL ISLAM, CHOW KIAN CHENG, RAHADIAN BAYU PERMADI and ROBERT FELDT: *Evaluation and Measurement of Software Process Improvement Ñ A Systematic Literature Review*. 38(2):398–424, 2012.

[Ver] VERHOEF, C: *Quantifying Software Process Improvement*. pages 1–148.

[vSB99] SOLINGEN, RINI VAN and EGON BERGHOUT: *The Goal/Question/Metric Method: a practical guide for quality improvement of software development*. McGraw-Hill Publishing Company, 1999.

[Wag13] WAGNER, STEFAN: *Software Product Quality Control*. Springer, 2013.

[Wal07a] WALLMÜLLER, ERNEST: *SPI - Software Process Improvement mit CMMI, PSP/TSP und ISO 15504*. Number ISBN: 978-3-446-40492-2. Carl Hanser, 2007.

[Wal07b] WALLMÜLLER, ERNEST: *SPI - Software Process Improvement mit CMMI, PSP/TSP und ISO 15504*. Hanser, München, 2007.

[WGW+93] WEBER, CHARLES V, SUZANNE M GARCIA, CHARLES V WEBER, SUZANNE M GARCIA and MARILYN BUSH: *Key Practices of the Capability Maturity Model SM , Version 1 . 1 Software Engineering Institute*. (February), 1993.

[WK] WANG, YINGXU and GRAHAM KING: *Philosophies and Approaches to Software Process Improvement*.

[WM06] WALKER, DIANE and FLORENCE MYRICK: *Grounded theory: an exploration of process and procedure*. Qualitative health research, 16(4):547–59, April 2006.

[Woh14] WOHLIN, CLAES: *Guidelines for Snowballing in Systematic Literature Studies and a Replication in Software Engineering*. 2014.

[WRH+00] WOHLIN, CLAES, PER RUNESON, MARTIN HÖST, MAGNUS C. OHLSSON, BJÖORN REGNELL and ANDERS WESSLÉN: *Experimentation in Software Engineering: An Introduction*. Kluwer Academic Publishers, Norwell, MA, USA, 2000.

[Xt] XT, V-MODELL: *Part 1 : Fundamentals of the V-Modell*.

[YC79] YOURDON, E. and L. CONSTANTINE: *Structured Design: Fundamentals of a Discipline of Computer Program and Systems Design*. Prentice Hall, 1979.

[Yin03] YIN, ROBERT K.: *Case study research - Design and methods*. Applied Social Research Methods Series, 5, 2003.

[ZP] ZAHRAN, SAMI and RICHINGS PARK: *Business and cost justification of software process improvement Ò ROI from SPI Ó*.

[ZW97] ZELKOWITZAYB, MARVIN V and DOLORES WALLACE: *Experimental validation in software engineering Õ*. 39:735–743, 1997.

[ZW98] ZELKOWITZ, MARVIN V. and DOLORES R. WALLACE: *Experimental Models for Validating Technology*. (May):23–31, 1998.