Quality Assessment of Requirement Specifications using Metrics – A Research Proposal –

Jakob Mund
Technische Universität München
Boltzmannstr. 3
85748 Garching, Germany
mund@in.tum.de

ABSTRACT

The presented research investigates the possibilites and limitations of metrics in assessing the quality of (software) requirement specifications (SRS) and thereby aims to improve the understanding of the notion of *quality* for those specifications.

Essentially, the idea is to define quality depending on the activites the SRS is used for in the engineering process. Based upon a common metamodel to describe and evaluate individual metrics, both the quality model and a systematic literature review serve as sources for potential metrics. Individual measurements are combined according to an assessment model in order to obtain an adequate quality estimation of the SRS.

This paper therefore presents the research questions and the associated research plan together with open issues. The author would appreciate feedback.

Categories and Subject Descriptors

D.2.1 [**Software Engineering**]: Requirements/Specification; D.2.8 [**Software Engineering**]: Metrics—*product metrics*

Keywords

Empirical Software Engineering, Requirements Engineering, Quality Assessment, Requirements Metrics

1. INTRODUCTION

Requirements Engineering is a crucial activity for project success but at the same time suffers from many inherent difficulties[7], e.g., a high degree of uncertainty about the system under development and the inevitable collision of different disciplines. Therefore, quality assurance is central to requirements engineering, and metrics-based approaches are a promising means to this end when applied correctly[6].

IDoESE '13 Baltimore, Maryland USA

The present paper investigates the possibilites and limitations of metrics in assessing the quality of (software) requirement specifications (SRS) and thereby aims to improve the understanding of the notion of quality for those specifications.

For this purpose the notion of *quality* for SRS has to be defined precisely. Here, the idea is to model quality depending on concrete activites where the SRS is used in the engineering process, and to identify impact factors for those activites, as described in [12, 9] for software quality in general. This model is then used together with a common metamodel to describe and evaluate individual metrics. Ultimatively, the individual measurements are combined according to an assessment model in order to obtain an adequate quality estimation of the SRS.

The present paper is structured as follows. Section 2 gives a short overview of relevant prior work both from science and practice. The fundamental research objective and the four research questions are presented in Section 3, which in turn shall be answered given the research plan presented in Section 4. Section 5 sketches some already identified open research issues, and Section 6 concludes.

2. RELEVANT PRIOR WORK

In this section we give a brief overview of the scientific foundations and industrial experiences which has led to this research proposal.

2.1 Scientific Foundations

This research proposal follows the general idea of the Goal-Question-Metrics (GQM) approach [2, 11], because it shares the same fundamental idea to measure only what is relevant for the designated goals. However, as will be described subsequently in more detail, the goals and questions are covered by a quality model, and operationalization is not taken into account here. Concerning metrics this proposal follows the definitions and terminology of [4], and the evaluation design is based on [13] for empirical case studies and experiments in general and [10] for evaluating metrics in particular.

On the other hand the research is also influenced by the research activites of former and present collegues in our research group; when considering (software) requirement specifications, we have a common artefact-model in mind [5]

which allows us to refer to specific pieces of the specification by using a given terminology, without limiting ourselves to a specific representation respectively notation. The notion of quality we refer to here is based upon recent work in the *Quamoco* Project [12, 9]. In particular, we follow the idea to define quality based upon the degree to which the activities of the software engineering endeavor are supported.

2.2 Industrial Experiences

Without going into much detail, experience in practice was also a major motivational factor for this research proposal. During the author's work on two projects at large german industrial companies concerning software development (process) metrics and an exploratory endeavor on quality metrics for requirement specifications, respectively, several interesting observations were made.

First and foremost, there is much interest concerning quality in requirement specifications in general and to have an efficient way to measure it in particular. However, the author made the observation that understanding the notion of quality regarding requirement specifications is not well understood in practice. Regarding the use of metrics for quality assurance, early evaluation and interviews suggested the following: (i) metrics must be easy to understand[1], (ii) metrics must come with recommendation for actions in order to be useful, and (iii) presentation and tooling matters, and in most cases, is underwhelming to say the least.

Special care has to be taken when applying metrics: the learning curve[2] must be taken into account and special care has to be taken when using a set of metrics as an instrument to guide development or management. In this case it is important to select not only *good* metrics in the sense of precision or efficiency, but also to select the *right* ones to be able to derive the profound decisions from it[3] .

3. RESEARCH QUESTIONS AND OBJECTIVES

The overall research objective is to investigate the role of metrics in requirements engineering to assess and indicate fields of improvement regarding the quality of requirement specifications in order to ultimatively improve cost and quality of software-intensive systems in its complete lifecycle, including system maintenance and use.

Following this general research direction four research questions shall be investigated in detail:

RQ1 How does the quality of requirement specifications influence the complete lifecycle of software-intensive systems?

RQ2 Which quality attributes of requirement specifications can be quantitatively measured and adequately assessed by the use of metrics?

RQ3 How and to what extent can a set of individual measurements be used to assess the quality of requirement metrics?

RQ4 How can metrics be employed for the purpose of a constructive quality assurence in the sense that it promotes quality improvement?

All research questions employ the notion of *quality* of requirement specifications. Unfortunately, despite several attempts (e.g., [1, 3]), this notion is still vague and poorly understood. The understanding of quality aims at seeing requirements engineering as a means for successfully engineering the (software-intensive) system under developement. As a consequence, the quality characteristics of requirement specifications have to be of significant use in the complete lifecycle, including the software development process. This working hypothesis shall be investigated in research question **RQ1**.

Given such quality characteristics the author is interested in factors, i.e. product or process properties, which significantly influence the quality of the requirement specification. Research question **RQ2** aims to identify those factors for which a quantiative measurement can deliver a *meaningful* result. Here, a meaningful result denotes a quantiative figure which can be used to associate a precise and valid interpretation to the quality attribute it shall describe. This utlimatively touches upon the question of which quality attributes shall be assessed using quantitative figures and for which qualitative means are more appropriate. Futhermore the author is well aware that not only properties of the quality attributes itself may be sufficient for the aforementioned distinction but other things have to be taken into account, e.g., the organizational context.

In general, research question **RQ3** is concerned with how to derive quality conclusions from individual measurements. This subsumes the questions how to derive more general assertions about the specification quality from measurements of quality attributes. This subsumes question of how to combine several measurements to derive more general assertions. Basically, the concern here is how to build and use a portfolio of several metrics in an efficient, effective and robust way, i.e., improve the trade-offs between the most accurate assessment results with minimal efforts which is resilient against disruptive influences such as data corruption or manipulation[4].

Finally, research question **RQ4** studies how those metrics identified as adequate and used accordingly to the previous research questions can be used in engineering software-intensive systems as an instrument to *constructively*, i.e. during requirement activities, improve the quality of require-

[1]Basically, the name must be self-explanatory to the degree that additional information is essentially not required.

[2]which, unfortunately, has positive and negative effects, e.g., regarding efficiency and exploitability

[3]An example often given are *Balanced Scorecards* [8], a strategic performance management approach in which reporting must occur in such a way that different but fixed perspectives on a given problem must always be considered in order to derive a decision which is good in a more holistic sense.

[4]One interesting idea might be to investigate if and to what degree the portfolio management of financial assets as known from applied mathematics can be applied. In particular, the quantitative handling of systematic and non-systematic risk may be applicable to threads to validity for metrics.

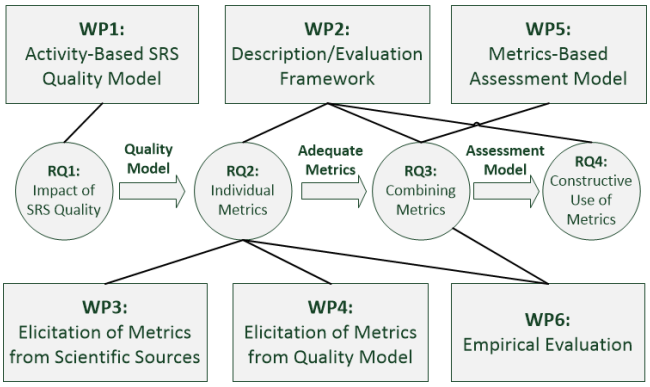

Figure 1: Overview of the research approach: the connection between research questions and work packages

ments specifications. Concerning metrics the goal is to employ metrics in such a way that it is very efficient to measure and the interpretation can be used to derive actions which effectively influence the quality of the requirements specification.

4. RESEARCH APPROACH

The research approach can be clustered into the following work packages, and figure 1 illustrates which work packages answer which research questions:

WP1: Activity-Based SRS Quality Model As the notion of quality is still not as well-defined as needed for quantitative measurements, we develop a more precise quality model for (software) requirement specifications. As we see requirements engineering in general and requirement specifications in particular as an auxiliary activity in systems engineering, the quality model identifies, describes and structure quality attributes of requirement specifications which directly influence the outcome of other engineering activies. The quality model is fundamental for the overall research objective and affects all mentioned research questions.

WP2: Description and Evaluation Framework An important prerequisite to deal with metrics in a systematic and rigorous way is to establish a basic, structured understanding of the ingredients to measurement systems. Therefore, a common meta-model for dealing with metrics in requirements engineering needs to be established. Based upon this meta-model, a common evaluation model to evaluate metrics with respect to quality attributes of the aforementioned quality model shall be proposed. The outcome of the evaluation is the degree to what individual metrics are adequate to measure individual quality attributes, targeting research question **RQ2**.

WP3: SRS Quality Metrics in Scientific Literature In order to obtain comparable results, metrics proposed in scientific literature shall be expressed in terms of the aforementioned meta-model. The identification, extraction and unification in a common meta-model shall be conducted using a systematic literature review. The main outcome of this study is a structured set of metrics described in a common meta-model, enabling comparable evaluation. But as a further "side-product", the state-of-the-art of the description of SRS quality metrics can be made visible in a concise way by showing which attributes of metrics are important to the scientific community, and compare this with the expectations of metrics in (practical) use.

WP4: SRS Quality Metrics from the Quality Model Besides the systematic literature review, the quality model itself can be a valuable source for individual metrics. Therefore, the identified quality factors must be made measurable: In general, this mostly has to consider different (common) notations and representations for the artifacts of the SRS, and is not an easy task by any means. The elicitation of metrics is planned to be conducted by an argumentative approach; to the author, this seems a reasonable way given that the outcome will still be evaluated.

WP5: Metric-Based Assessment Model As part of the research approach an assessment model using individual metrics and measurements shall be constructed (reserach question **RQ3**). Therefore, as a first step relations between metrics and its modalities shall be identified and integrated into the metrics meta-model, based on an argumentative approach (more likely) or empirical evidence (less likely). Fundamental characteristics for the relations and its modalities are for example costs and threats to validity of the individual metrics. Then, the applicability of measurement composition and portfolio models to requirement specification metrics shall be studied, and consequently adapted to propose the metric-based assessment model.

WP6: Empirical Evaluations Given the metrics extracted from scientific literature and derived from the quality model, which are described in a unified meta-model and an evaluation framework, the individual metrics shall be empirically evaluated using case studies or controlled experiments. This allows to assess the applicability of the individual metrics itself on the one hand and to sharpen the interpretations (e.g., thresholds) associated with the metrics based on empirical evidence. Futhermore, the proposed metric-based assessment model can be evaluated by combining the measurement results already present according to the model. Hence, this work package addresses research questions **RQ2** and **RQ3**.

At the point in time of writing the research proposal, the author is not sure whether and to what degree the constructive usage of metrics (research question **RQ4**) in concrete project settings shall be evaluated empirically. Basically, the idea is to define what "constructive use" of metrics really is, identify those attributes required from metrics to be usable for this purposes and give an idea about how to integrate metrics into requirements engineering activites. However, the research questions seems to require long-term studies with realistic baseline approaches for quality assessment for the empirical evaluation.

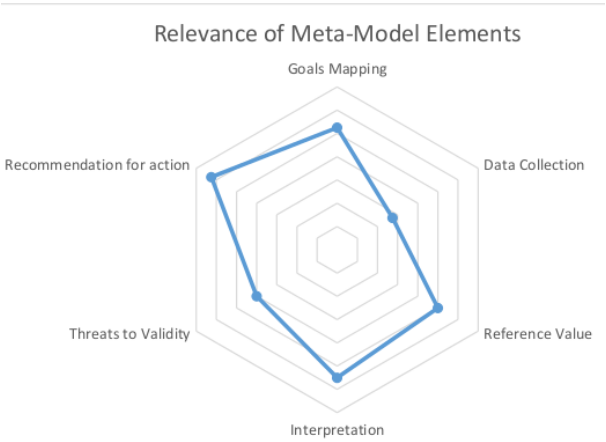

Figure 3: Preliminary results of the metrics meta-model from four process engineers of Wacker Chemie AG

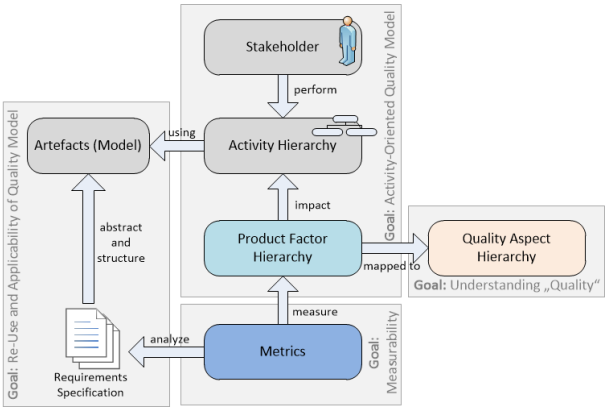

Figure 5: The Activity-Based SRS Quality Model

4.1 Current Research, Preliminary Results and Planned Next Steps

Regarding current research progress, a meta-model for structuring metrics has already been developed (cf. figure 2). The meta-model follows the idea of GQM+Strategies [2] applied to requirements engineering in the sense that the general goals become quality goals (of requirement specifications) and the questions become quality factors, all according to the aforementioned activity-based SRS quality model. Furthermore, we structure the data collection (measurement), data analysis (interpretation) and improvement forecast (recommended actions) in more detail. Here, we introduce the notion of threats to validity known from empirical software engineering [13] to allow for a more sensitive handling of metrics. We were already able to rudimentary evaluate the description in practice at Wacker Chemie AG (using structured interviews with four process engineers), where preliminary results suggested a more structured view on metrics helped in implementing and using metrics, and the importance of the metrics components for its use can be rated as seen in figure 3.

Currently, the meta-model is to be supplemented by an evaluation model based upon a taxonomical definition of metrics and the criteria defined in [10]. In its current form inter-metric relations are limitied: an explicit "refines" relation expressing that one metric is "more precise" than another, where this refinement presumes that the metrics measure the same quality factor or at least the same quality attribute. For a more rigorous approach considering **RQ3**, this refinement relation has to be defined much more precisely, and further factors, threats to validity in particular, have to be taken into account.

The fundamental idea of the activity-based SRS quality model, inspired by the *Quamoco Project*[12, 9], is illustrated in figure 5. So far, we identified first stakeholders and their activities, and instanciated the model to show the level of abstraction on quality and the potential outcome of the model. Figure 4 shows an example for the testing activitiy *Perform system test*, in which we identified two factors, i.e., the delimitability and the unambigiouty of the order of the individual steps, which have a (positive) impact on the sub-activity to create the test sequence. To gain a deeper understanding of the more common notions of quality (e.g., [1]) we associate those quality attributes with the factors of the activity-based quality model.

When done with the quality model and metrics description and evaluation framework, the next step would be to elicitate metrics: the systematic literature review and derivation of metrics from the quality model.

4.2 When am I done?

Considering the depth and size of each of the aforementioned activities the author's approach is to bring each of it to the level that saturation sets in, in the sense that more details or more amount would not introduce significantly new insights. In order to reach this goal a substantial number of metrics must be elicitated, an the empirical evaluation has to at least incorporate two case studies to clear out the most generous threats to external validity. Additionally, small experiements shall be applied to fill potential gaps in the industrial case studies due to, e.g., specific parameters not present in the specific project setting.

As the research questions are quite large in scope the research plan proposed in this paper can only provide and evaluate a limited number of answers to them. It is out of scope to evaluate which is the best answer to the reseach questions, e.g., what is the *definitive* definition of quality, but the aim of this research proposal is to provide hopefully valuable insights and empirical evidence to the state of the art in metrics-based quality assessment of requirement specifications.

5. OPEN RESEARCH ISSUES

The following threats to validity and open reseach issues were identified by the author:

Validity of Hypothesis The metric-based quality assessment is based on a number of fundamental hypothesis such as:

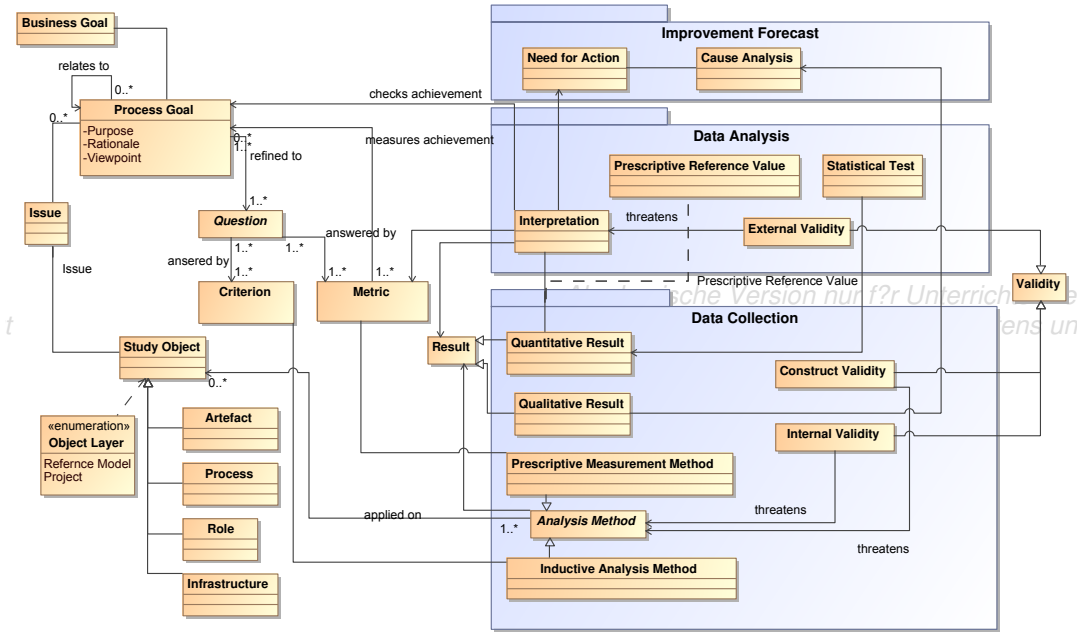

Figure 2: The SRS quality metrics meta-model

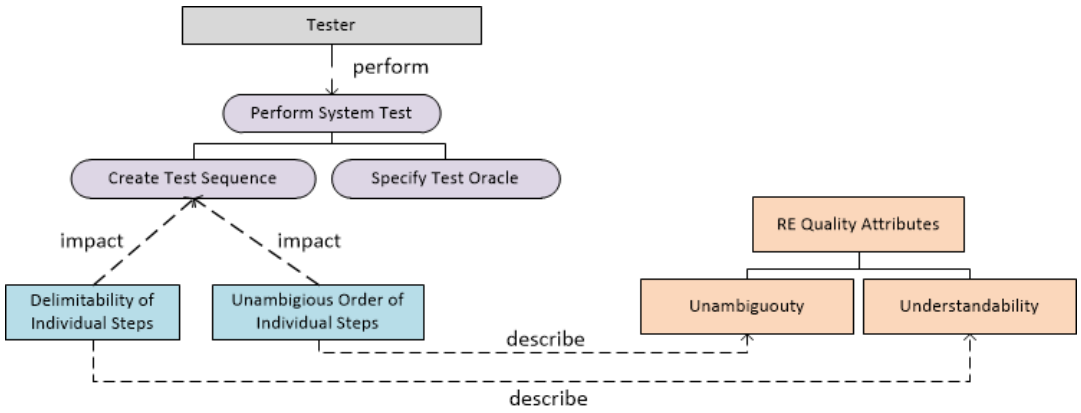

Figure 4: Quality model example regarding the engineering activity "testing"

- Quality of requirement specifications has significant impact on the cost and/or quality of the system under development
- Quality can be made an explicit, objective and measureable criterion of requirement specifications
- Metrics are capable of adequately measuring quality of requirement specifications
- Quality goals can be "translated" into useful prescriptive reference values
- Metrics can be used to efficiently and effectively improve the quality of requirement specifications

Whether those hypothesis hold is on the one hand part of the research outcome, on the other hand fundamental for the some of the planned activities.

Empirical Flaws Subtle flaws in empirical studies can easily lead to data suggesting flawed theories, especially for the author given few experience in empirical software engineering.

Unawareness of Related Work Metrics and its application is a topic widely discussed in literature not only restricted to software engineering but many other and quite different disciplines too. Therefore the potential exists to get inspiration from other disciplines, e.g. mathematics (measurement theory in particular) or economics, but also to be unaware of publications in fields which the author is not primarly involved in.

The Top-Down vs. Bottom Up Spectrum Thinking of the topic of RE metrics, the proposed approach can be considered "top-down". Although we want to avoid

the "measure what we can measure for measurement's sake" pitfall, however, there are some associated risks: (i) the author gets feedback about the approaches applicability quite late, hence (ii) early results may (and probably will) be at least partly invalidated, and (iii) the late results are hard to get you projects and fundings in industrial projects.

Operationalization as a Risk in Evaluation The research work targets two distinct directions concerning the requirements metrics: conceptual work on the quality model and framework for metrics in requirements engineering on the one hand and the empirical evaluation of individual metrics in practice on the other hand, with the goal to answer research questions on a conceptual level. However the research plan (intentionally) ignores how to bridge this gap

6. CONCLUSIONS

Here, we presented a research proposal to investigate how and to what degree metrics can be used to assess the quality of (software) requirement specifications. The research joins argumentative and conceptual approaches (e.g., constructing the assessment model) with empirical methods (e.g., elicitating metrics from scientific sources and case studies for evaluation purposes). The author would greatly appreciate feedback on this research plan, especially but not limited to the open research issues presented in the last section.

7. ACKNOWLEDGMENTS

I would like to thank Manfred Broy for advising my PhD thesis and the opportunity to apply for the IDoESE doctoral symposium. Futhermore I want to thank my collegues at my research group for the joint work on some of the more fundamental topics of requirements engineering, and in particular Daniel Mendez and Henning Femmer for fruitful discussions about initial drafts of this research proposal.

8. REFERENCES

[1] Ieee guide to software requirements specifications. *IEEE Std 830-1984*, 1984.

[2] V. R. Basili, G. Caldiera, and H. D. Rombach. The goal question metric approach. *Encyclopedia of software engineering*, 2:528–532, 1994.

[3] A. Davis, S. Overmyer, K. Jordan, J. Caruso, F. Dandashi, A. Dinh, G. Kincaid, G. Ledeboer, P. Reynolds, P. Sitaram, et al. Identifying and measuring quality in a software requirements specification. In *Software Metrics Symposium, 1993. Proceedings., First International*, pages 141–152. IEEE, 1993.

[4] N. E. Fenton and S. L. Pfleeger. *Software metrics: a rigorous and practical approach*. PWS Publishing Co., 1998.

[5] D. M. Fernández, B. Penzenstadler, M. Kuhrmann, and M. Broy. A meta model for artefact-orientation: fundamentals and lessons learned in requirements engineering. In *Model Driven Engineering Languages and Systems*, pages 183–197. Springer, 2010.

[6] S. Fickas and M. S. Feather. Requirements monitoring in dynamic environments. In *Requirements Engineering, 1995., Proceedings of the Second IEEE International Symposium on*, pages 140–147. IEEE, 1995.

[7] N. Juristo, A. M. Moreno, and A. Silva. Is the european industry moving toward solving requirements engineering problems? *Software, IEEE*, 19(6):70–77, 2002.

[8] R. S. Kaplan and D. P. Norton. Using the balanced scorecard as a strategic management system. *Harvard business review*, 74(1):75–85, 1996.

[9] K. Lochmann, D. M. Fernández, and S. Wagner. A case study on specifying quality requirements using a quality model. In *APSEC*, pages 577–582, 2012.

[10] A. Meneely, B. Smith, and L. Williams. Validating software metrics: A spectrum of philosophies. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 21(4):24, 2012.

[11] R. Van Solingen and E. Berghout. Integrating goal-oriented measurement in industrial software engineering: industrial experiences with and additions to the goal/question/metric method (gqm). In *Software Metrics Symposium, 2001. METRICS 2001. Proceedings. Seventh International*, pages 246–258. IEEE, 2001.

[12] S. Wagner, K. Lochmann, L. Heinemann, M. Kläs, A. Trendowicz, R. Plösch, A. Seidl, A. Goeb, and J. Streit. The quamoco product quality modelling and assessment approach. In *Proceedings of the 2012 International Conference on Software Engineering*, pages 1133–1142. IEEE Press, 2012.

[13] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén. *Experimentation in software engineering*. Springer, 2012.