ORIGINAL ARTICLE

A systematic literature review of empirical research on quality requirements

Thomas Olsson[1] · Séverine Sentilles[2] · Efi Papatheocharous[1]

Abstract

Quality requirements deal with how well a product should perform the intended functionality, such as start-up time and learnability. Researchers argue they are important and at the same time studies indicate there are deficiencies in practice. Our goal is to review the state of evidence for quality requirements. We want to understand the empirical research on quality requirements topics as well as evaluations of quality requirements solutions. We used a hybrid method for our systematic literature review. We defined a start set based on two literature reviews combined with a keyword-based search from selected publication venues. We snowballed based on the start set. We screened 530 papers and included 84 papers in our review. Case study method is the most common (43), followed by surveys (15) and tests (13). We found no replication studies. The two most commonly studied themes are (1) differentiating characteristics of quality requirements compared to other types of requirements, (2) the importance and prevalence of quality requirements. Quality models, QUPER, and the NFR method are evaluated in several studies, with positive indications. Goal modeling is the only modeling approach evaluated. However, all studies are small scale and long-term costs and impact are not studied. We conclude that more research is needed as empirical research on quality requirements is not increasing at the same rate as software engineering research in general. We see a gap between research and practice. The solutions proposed are usually evaluated in an academic context and surveys on quality requirements in industry indicate unsystematic handling of quality requirements.

Keywords Systematic literature review · Quality requirements · Non-functional requirements · Empirical evidence

1 Introduction

Quality requirements—also known as non-functional requirements—are requirements related to how well a product or service is supposed to perform the intended

✉ Thomas Olsson
thomas.olsson@ri.se
https://ri.se

Séverine Sentilles
severine.sentilles@mdh.se
https://www.mdh.se

Efi Papatheocharous
efi.papatheocharous@ri.se

1 RISE Research Institutes of Sweden AB, Gothenburg, Sweden

2 Mälardalen University, Västerås, Sweden

functionality [48]. Examples are start-up time, access control, and learnability [56]. Researchers have long argued the importance of quality requirements [39, 68, 82]. However, to what extent have problems and challenges with quality requirements been studied empirically? A recent systematic mapping study identified quality requirements as one of the emergent areas of empirical research [7]. There are several proposals over the years for how to deal with quality requirements, e.g., the NFR method [78], QUPER [87], quality models [37], and i* [107]. However, to what extent have they been empirically validated? We present a systematic literature review of empirical studies on problems and challenges as well as validated techniques and methods for quality requirements engineering.

Ambreen et al. conducted a systematic mapping study on empirical research in requirements engineering [7], published in 2018. They found 270 primary studies where 36 papers were categorized as research on quality requirements. They concluded that empirical research on quality requirements is an emerging area within requirements

engineering. Berntsson Svensson et al. carried out a systematic mapping study on empirical studies on quality requirements [15], published in 2010. They found 18 primary empirical studies on quality requirements. They concluded that there is a lack of unified view and reliable empirical evidence, for example, through replications and that there is a lack of empirical work on prioritization in particular. In our study, we follow up on the systematic mapping study om Ambreen et al. [7] by performing a systematic literature review in one of the highlighted areas. Our study complements Berntsson Svensson et al. study from 2010 by performing a similar systematic literature review 10 years later and by methodologically also using a snowball approach.

There exist several definitions of quality requirements as well as names [48]. Glinz defines a non-functional requirement as an attribute (such as performance or security) or a constraint on the system. The two prevalent terms are quality requirements and non-functional requirements. Both are used roughly as much and usually mean approximately the same thing. The ISO25010 defines quality in use as to whether the solution fulfills the goals with effectiveness, efficiency, freedom from risk, and satisfaction [56]. Eckhardt et al. analyzed 530 quality requirements and found that they described a behavior—essentially a function [40]. Hence, the term non-functional might be counter-intuitive. We use the term quality requirements in this paper. In layman's terms, we mean a quality requirement expresses *how well* a solution should execute an intended function, as opposed to functional requirements which express *what* the solution should perform. Furthermore, conceptually, we use the definition from Glinz [48] and the sub-characteristics of ISO25010 as the main refinement of quality requirements [56].

We want to understand from primary studies (1) what are the problems and challenges with quality requirements as identified through empirical studies, and (2) which quality requirements solutions have been empirically validated. We are motivated by addressing problems with quality requirements in practice and understanding why quality requirements is still, after decades of research, often reported as a troublesome area of software engineering in practice. Hence, we study which are the direct observations and experience with quality requirements. We define the following research questions for our systematic literature review:

RQ1   Which empirical methods are used to study quality requirements?
RQ2   What are the problems and challenges for quality requirements identified by empirical studies?
RQ3   Which quality requirements solution proposals have been empirically validated?

We study quality requirements in general and therefore exclude papers focusing on specific aspects, e.g., on safety or user experience.

We summarize the related literature reviews in Sect. 2. We describe the hybrid method we used for our systematic literature review in Sect. 3. Section 4 elaborates on the findings from screening of 530 papers to finally include 84 papers from the years 1995 to 2019. We discuss the results and threats to validity in Sect. 5; empirical studies on quality requirements are—in relative terms—less common than other types of requirements engineering papers, there is a lack of longitudinal studies of quality requirements topics, we found very few replications. We conclude the paper in Sect. 6 with a reflection that there seems to be a divide between solutions proposed in an academic setting and the challenges and needs of practitioners.

2 Related work

A recent systematic mapping study on empirical studies on requirements engineering states that quality requirements are "by far the most active among these emerging research areas" [7]. They classified 36 papers of the 270 they included as papers in the quality requirements area. In their mapping, they identify security and usability as the most common topics. These results are similar to that of Ouhbi et al. systematic mapping study from 2013 [81]. However, they had slightly different keywords in their search, including also studies on quality in the requirements engineering area, which is not necessarily the same as quality requirements. A systematic mapping study is suggested for a broader area whereas a systematic literature review for a narrower area which is studies in more depth [63]. To our knowledge, there are no recent systematic literature reviews on quality requirements.

Berntsson Svensson et al. performed a systematic literature review on empirical studies on managing quality requirements in 2010 [15]. They identified 18 primary studies. They classified 12 out of the 18 primary studies as case studies, three as experiments, two as surveys, and one as a mix of survey and experiment. They classified only four of the 18 studies as properly handling validity threats systematically. Their results indicate that there is a lack of replications and multiple studies on the same or similar phenomena. However, they identify a dichotomy between two views; those who argue that quality requirements need special treatment and others who argue quality requirements need to be handled at the same time as other requirements. Furthermore, they identify a lack of studies on prioritization of quality requirements. Berntsson Svensson et al. limited their systematic literature review to studies containing the keyword "software," whereas we

Fig. 1 We used two different approaches to create the start sets: Start set I is based on two other literature reviews, and Start set II is created through a keyword-based search in relevant publication venues. The two start sets are combined and snowballed on to arrive at the final set of included papers. The numbers between the steps in each set are the number of references within that set. The numbers between the sets are the total number of references included in the final set
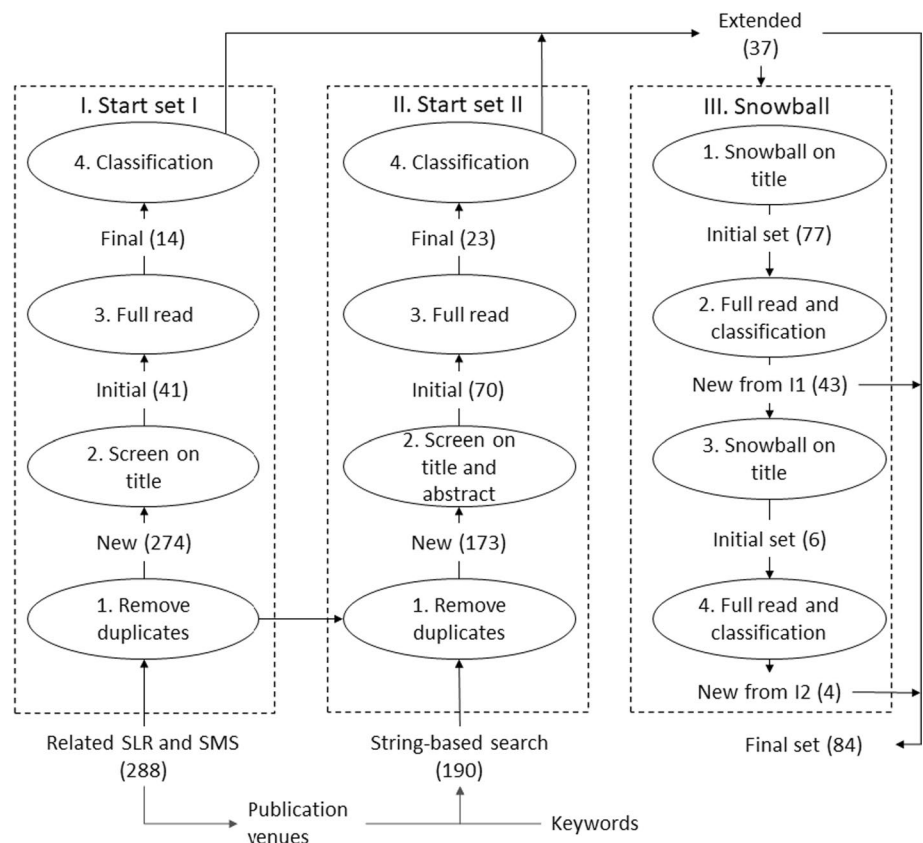
did not in our study. Furthermore, Berntsson Svensson et al. performed a keyword-based literature search with a number of keywords required to be present in the search set. We used a hybrid approach and relied on snowballing instead of strict keywords. Lastly, we used Ivarsson and Gorschek [58] for rigor, which entailed stricter inclusion criteria, i.e., as a result we did not include all studies from Berntsson Svensson et al. This, in combination with performing the study 10 years afterward, means we complement Berntsson Svensson both in terms of the method as well as studied period.

Alsaqaf et al. could not find any empirical studies on quality requirements in their 2017 systematic literature review on quality requirements in large-scale agile projects [5]. They included studies on agile practices and requirements in general. Hence, their scope does not overlap significantly with ours. They found, however, 12 challenges to quality requirements in an agile context. For example, a focus on delivering functionality at the expense of architecture flexibility, difficulties in documenting quality requirements in user stories, and late validation of quality requirements. We do not explicitly focus on agile practices. Hence, there is a small overlap between their study and ours.

3 Design

We designed a systematic literature review using a hybrid method [77]. The hybrid method combines a keyword-based search, typical of a systematic literature review [63], to define a start set and a snowball method [105] to systematically find relevant papers. We base our study on two literature reviews [7, 15], which we complement in a systematic way. The overall process is found in Fig. 1.

3.1 Approach

We decided to use a hybrid approach for our literature review [77]. A standard keyword-based systematic literature review [63] can result in a very large set of papers to review if keywords are not restrictive. On the other hand, having too restrictive keywords can result in a too-small set of papers. A snowball approach [105], on the other hand, is sensitive to the start set. If the studies are published in different communities not referencing each other, there is a risk of not finding relevant papers if the start set is limited to one community. Hence, we used a hybrid

Table 1 Journals and conferences included in finding the Start set II

Topic	Forum
Software Engineering	International conference on software engineering
	Asia-Pacific software engineering conference
	International computer software and applications conference
	Symposium on applied computing
	Transactions on software engineering
	Information and software technology
	Journal of systems and software
Requirements Engineering	International requirements engineering conference
	Requirements engineering: foundation for software quality[a]
	Requirements engineering journal
Empirical Software Engineering	Empirical software engineering symposium
	Empirical software engineering journal

[a]Before 2005, the REFSQ conference is not searchable in Scopus. As we are snowballing, we do not see this as a large threat to the validity

method where we combine the results from a systematic mapping study and a systematic literature review to give us one start set with a keyword-based search in the publication venues of the papers from the two review papers.

1. Start set I

We defined Start set I for our systematic literature review by using a systematic mapping study on empirical evidence for requirements engineering in general [7] from 2018 and a systematic literature review from 2010 [15] with similar research questions as in our paper.

The systematic literature review from 2010 by Berntsson Svensson et al. includes 18 primary studies [15]. However, we have different inclusion criteria (see Sect. 3.2). Hence, not all the references are included. In our final set, we included 10 of the 18 studies.

The systematic mapping by Ambreem et al. from 2018 looks at empirical evidence in general for requirements engineering [7]. They included 270 primary studies. However, there are some duplicates in their list. They classified 36 papers to be in the quality requirements area. However, there is an overlap with the Berntsson Svensson et al. review [15]. When we remove the already included papers from Berntsson Svensson et al., we reviewed 24 from Ambreem et al. and in the end included 4 of them.

2. Start set II

To complement start set I, we also performed a keyword-based search. We have slightly different research questions than the two papers in the Start set I. Therefore, our search string is slightly different than that of Ambreen et al. and Berntsson Svensson et al. Also, the most recent references in Start set I are from 2014, i.e., five years before we performed

our search. Hence, we also fill the gap of the papers published since 2014. We include all studies, not just studies from 2014 and onward, as our method and research questions are slightly different.

We used the most frequent and highly ranked publication venues from Start set I to limit our search but still have a relevant scope. Table 1 summarizes the included conferences and journals. Even though the publication venues included are not an exhaustive list of applicable venues, we believe they are representative venues that are likely to include most communities and thereby reducing the risk with a snowball approach of missing relevant publications, as intended with the hybrid method.

We used Scopus to search. The search was performed in September 2019. Table 2 outlines the components of the search string. The title and abstract were included in the search and only papers that include both the keyword for quality requirements and the keywords for empirical research.

Table 2 The components of the search string used in Scopus. Scopus handles stemming, variation such as "non-functional" and "non functional," etc. Hence, possible variations are handled

Topic	Keyword
Quality requirements	("Quality requirements" OR "Non-functional requirements" OR "Extra functional properties") AND
Empirical studies	("Empirical" OR "Survey" OR "Case study" OR "Experiment" OR "Interviews")

3. Snowballing

The last step in our hybrid systematic review [77] is the snowballing of the start set papers. We snowballed on the extended start set—the combination of Start set I and Start set II—to get to our final set of papers (cf. Fig. 1). In a snowball approach, both references in the paper (backward references) and papers referring to the paper (forward references) were screened [105]. We used Google Scholar to find forward references.

3.2 Planning

We arrived at the following inclusion criteria, in a discussion among the researchers and based on related work:

1. The paper should be on quality requirements or have quality requirements as a central result.
2. There should be empirical results with a well-defined method and validity section, not just an example or anecdotal experience.
3. Papers should be written in English.
4. The papers should be peer-reviewed.
5. The papers should be primary studies.
6. Conference or journal should be listed in reference ranking such as SJR.

Similarly, we defined our exclusion criteria as:

1. Literature reviews, meta-studies, etc.,—secondary studies—are excluded.
2. If a conference paper is extended into a journal version, we only include the journal version.
3. We include papers only once, i.e., duplicates are removed throughout the process.
4. Papers focusing on only specific aspect(s) (security, sustainability, etc.) are excluded.

All researchers were involved in the screening and classification process, even though the primary researcher performed the bulk of the work. The screening and classification were performed as follows:

1. Screen based on title and/or abstract.
2. We performed a full read when at least one researcher wanted to include the paper from the screening step.
3. Papers were classified according to the review protocol, see Sect. 3.3. This was performed by the primary researcher and validated by another researcher.

To ensure reliability in the inclusion of papers and coding, the process was performed iteratively according to the sets.

1. For Start set I, all references from the systematic literature review [15] and systematic mapping study [7] were screened by two or three researchers. We only used the title in the screening step for Start set I. Full read and classification were performed by two or three researchers.
2. For Start set II, the screening was primarily performed by the primary researcher but with frequent alignment with at least one more researcher to ensure consistent screening—both on title and abstract. Similarly for the full read and classification of the papers. Specifically, we paid extra attention to which papers to exclude to ensure we did not exclude relevant papers.
3. The primary researcher performed the snowballing. We screened on title only for backward and forward snowballing. We included borderline cases, to ensure we did not miss any relevant references.

The full read and classification were primarily performed by the primary researcher for Start set II and the Snowballing set. A sample of the papers was read by another researcher to improve validity in addition to those cases already reviewed by more than one researcher.

The combined number of primary studies from the systematic literature review [15] and systematic mapping study [7] are 288 in Start set I. However, there is an overlap between the two studies and there are some duplicates in the Ambreen et al. paper [7]. In the end, we had 274 unique papers in Start set I. After the screening, 41 papers remained. After the full read, additional papers were excluded resulting in 14 papers in the Start set I.

Our search in Start Set II resulted in 190 papers. A total of 173 papers remained after removing duplicate papers and papers already included in Start set I. After the screening and full read, the final Start set II was 23. Hence, the extended start set (combining Start set I and Start set II) together resulted in the screening of 447 papers and the inclusion of 37 papers.

The snowball process was repeated until no new papers are found. We iterated 2 times—denoted I1 and I2 in Fig. 1. In iteration 1, we reviewed 77 papers and included 43. In iteration 2, we reviewed 6 papers and included 4. This resulted in a total of 84 papers included and 530 papers reviewed.

3.3 Classification and review protocol

We developed the review protocol based on the systematic literature review [15] and systematic mapping study [7], the methodology papers [63, 105], and our research questions. The main items in our protocol are:

Fig. 2 An overview of papers included—publication type and year of publication
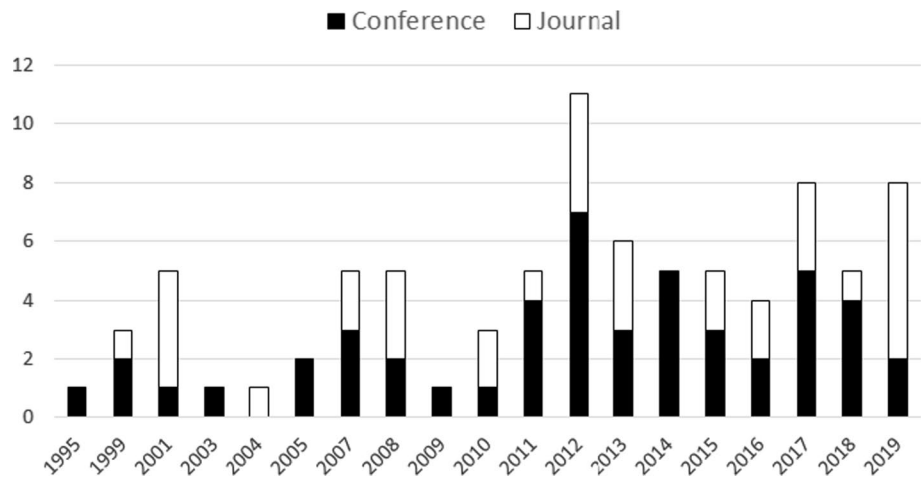
- Type of empirical study according to Wieringa et al. [104]. As we are focusing on empirical studies, we use the evaluation type—investigations of quality requirements practices in a real setting—,validation type—investigations of solution proposals before they are implemented in a real setting—,or experience type—studies where the researchers are taking a more part in the study, not just observing.
- Method used in the papers. We found the following primary methods used: Experiment, test, case study, survey, and action research.
- Analysis of rigor according to Ivarsson and Gorschek [58].
- Thematic analysis of the papers—in an initial analysis based on the author keyword and in later iterations further refined and grouped during the analysis process.

We used a spreadsheet for documentation of the classification and review notes. The classification scheme evolved iteratively (see Sect. 3.2) as we included more papers. The initial themes were documented in the review process. In the analysis phase, the initial themes were used for an initial grouping of the papers. The themes were aligned and grouped in the analysis process of the papers, which included a number of meetings and iterative reviews of the results. The final themes which we used for the papers are the results of the iterative analysis process, primarily performed by the first and third researcher.

3.4 Validity

All cases where there were uncertainties whether to include a paper—both in the screening step and the full read step—or on the classification were reviewed by at least two researchers. Furthermore, to ensure consistent

use of the inclusion and exclusion criterion as well as the classification we also sampled and reviewed papers that had only been screened or reviewed by only one researcher.

We used Scopus for Start set II. We confirmed that all journals and conferences selected from Start set I were found in Scopus. However, REFSQ was only indexed from 2005 and onward. However, we do not see this as a problem as we are snowballing and the papers that are missing from the Scopus search due to this, should appear in the results through the snowballing process.

We used Google Scholar in the snowballing. This is recommended [105] and usually gives the most complete results.

A hybrid search strategy can be sensitive to starting conditions, as pointed out by Mourão et al. [77]. However, their results indicate that the strategy can produce similar results as a standard systematic literature review. We carefully selected the systematic literature review and the systematic mapping study as Start set I and extended it with a keyword-based search for selected forums in Start set II. Hence, we believe the extended start set on which we snowballed is likely to be sufficient to ensure a good result when complemented with the snowball approach.

4 Analysis and results

The screening and reading of the papers in Start set I was performed in August and September 2019. The keyword-based search for Start set II was performed in September 2019. The snowballing was subsequently performed in October and November. In total, 530 papers are screened, of which 194 papers are read in full. This resulted in including 84 papers, from 1995 to 2019—see Fig. 2.

Table 3 Type of empirical study and context

Type method	Validation		Evaluation		Experience action research	Total
	Experiment	Test	Case study	Survey		
Academic	4	13	5		1	23
Industry		4	35	14	2	55
Mixed		2	2	1		5
Open source			1			1
Total	4	19	43	15	3	84

Fig. 3 An overview of papers included—method and accumulated number of publications per year
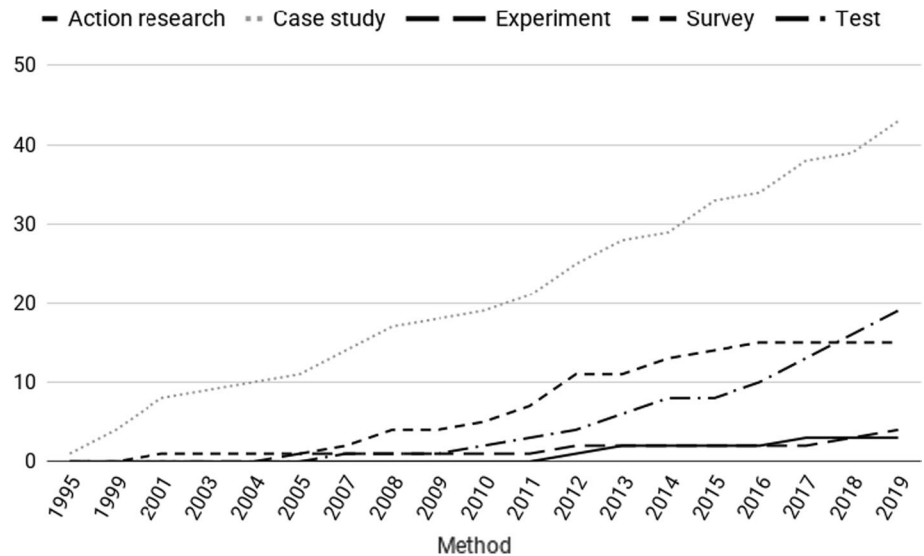
4.1 RQ1 Which empirical methods are used to study quality requirements?

The type of studies performed is found in Table 3—categorized according to Wieringa et al. [104]. We differentiate between two types of validations: experiments involving human subjects and tests of algorithms on a data set. For the latter, the authors either report experiment or case study, whereas we call them test. The evaluations we found are either performed as case studies or surveys. Lastly, we found three papers that used action research—categorized as experience in Table 3. It should be noted that the authors of the action research papers did not themselves explicitly say they performed an action research study. However, when we classified the papers, it is quite clear that, according to Wieringa et al. [104], they are in the action research category.

Case studies in an industry setting are the most common (35 of 84), followed by surveys in industry (14 of 84) and test in academic settings (13 of 84). This indicates that research on quality requirements is applied and evidence is primarily of individual case studies rather than through validation in laboratory settings. Case studies seem to have similar popularity over time, see Fig. 3. We speculate that since requirements engineering in general as well as quality requirements in particular is a human-intensive

activity, there are not so many clear cause–effect relationships to study in a rigorous experiment. Rather, it is more important to study practices in realistic settings. However, there are only three longitudinal studies.

Tests are, in contrast to the case studies, primarily performed in an academic setting, which is not necessarily representative in terms of scale and artifacts. The papers are published from 2010 to 2019—one exception, published in 2007, see Fig. 3. One explanation might be the developments in computing driving the trend to use large data sets.

We found only one study on open source, see Table 3, which is also longitudinal. We speculate that requirements engineering is sometimes seen as a business activity where someone other than the developers decides what should be implemented. In open-source projects, there is often a delegated culture where there is no clear product manager or similar deciding what to do, albeit there can be influential individuals such as the originator or core team member. We believe this entails that quality requirements engineering is different in open-source projects than when managed within an organization. It would be interesting to see if this hypothesis holds for requirements engineering in general and not just quality requirements. We believe, however, that by studying forums, issue management, and reviews that

Table 4 Overview of rigor in the studies, based on Ivarsson and Gorschek [58]

Design score	Validation score	Validation	Evaluation	Experience	Total
1	1	11	25	1	37
1	0.5	1	5	1	7
1	0	5	5		10
0.5	1	1	1		2
0.5	0.5		5		5
0.5	0	1	10		11
0	0	4	7	1	12
Total		23	58	3	84

We evaluate the design description—whether it would be possible to replicate the study—and validity description—whether the relevant validity threats are well described. A score of "1" means the design or validation is well described in sufficient detail. A score of "0.5" means the design or validation is briefly described but lacking relevant details. A score of "0" means there is no description or it is described in such a way that it is not possible to understand the details

open-source projects are an untapped resource for quality requirements research.

We classify the studies according to rigor, as proposed by Ivarsson and Gorschek [58]. We assess the design and validity rigor. Table 4 presents our evaluation of design and validity rigor in the papers. An inclusion criterion is that there should be an empirical study, not just examples or anecdotes. Hence, it is not surprising that overall studies score well in our rigor assessment.

High rigor is important for validations studies—to allow for replications—which is also the case for 11 out of 23 studies. The number increases to 13 if we include papers with a rigor score 0.5 for both design and validity and to 19 if we focus solely on the design rigor. Interestingly, we found no replication studies. Furthermore, the number of studies on a single (similar) approach or solution is in general low. We speculate that the originators of a solution have an interest in performing empirical studies on their solution. However, it seems unusual that practitioners or empiricists with no connection to the original solution or approach try to apply it. Furthermore, we also speculate that academic quality requirements research is not addressing critical topics for industry as there seems not to be an interest in applying and learning more about them. This implies that the research on quality requirements might need to better understand what are the real issues facing software developing organizations in terms of quality requirements.

The validity part of rigor is also important for evaluations and experience papers. Strict replications are typically not possible. However, understanding the contextual factors and validity are key in interpreting the results and assessing their applicability in other cases and contexts. 22 of the 58

Table 5 Overview of the experiment studies

Reference	Year	Theme	Subjects	Context
[3]	2005	ISO9126	158	Students
[99]	2012	CSRML (i*)	84	Students
[64]	2018	Templates	107	Students
[111]	2019	i*	32	Students

Table 6 Overview of the test studies. The data set is described in Table 7

Ref.	Year	Data set	Theme
[28]	2007	DePaul07, Siemens IET	Automatic analysis
[24]	2010	DePaul07	Automatic analysis
[109]	2011	DePaul07	Automatic analysis
[44]	2012	DePaul07, EU procurement	Tool
[95]	2013	DePaul07, CCHIT, iTrust, openEMR, DUA, RFP	Automatic analysis
[85]	2013	HWS, CRS (OSS projects)	Tool
[86]	2014	CCHIT, WorldVista	Automatic analysis
[91]	2014	DePaul07	Automatic analysis
[94]	2016	DePaul07, Concordia corpus	Automatic analysis
[74]	2016	Proprietary	Automatic analysis
[66]	2017	DePaul07	Automatic analysis
[72]	2017	Mobile app store reviews	Automatic analysis
[70]	2017	PTC, MIP, CCHIT	Automatic analysis
[6]	2018	DePaul07	Automatic analysis
[100]	2018	DePaul07	Automatic analysis
[42]	2018	TAS, deltaIoT	Runtime analysis
[60]	2019	Mobile app store reviews	Automatic analysis
[10]	2019	DePaul07, Predictor models	Automatic analysis
[106]	2019	DePaul07, CCHIT	Automatic analysis

evaluation and 1 of the 3 experience papers do not have a well-described validity section (rigor score 0), and 10 evaluation and 1 experience paper have a low score (rigor score 0.5). Hence, we conclude that the overall strength of evidence is weak.

4.1.1 Validations

Experiments are, in general, the most rigorous type of empirical study with the most control. However, it is difficult to scale to a realistic scenario. We found four experiments validating quality requirements with human subjects, see Table 5.

We note that all experiments are performed with students—at varying academic levels. This might very well be appropriate for experiments [54]. We notice that there are only four experiments, which might be justified by: (1) Experiments as a method is not well accepted nor understood

Table 7 Data sets. In most cases when they are openly accessible, they can, with some effort, be found online

Data set	Context	Access	Content
DePaul07	Academic	Open	Annotated requirements [27]
Siemens IET	Industry	Closed	Requirements document [28]
EU procurement	Public	Open	Requirements document [44]
CCHIT	Public	Open	Requirements document [95]
iTrust	?	?	Requirements documents [95]
OpenEMR	Public	Open	Manuals [95]
DUA	Public	?	Data Use Agreements [95]
RFP	Public	?	Request for Proposals [95]
HWS	?	?	? [85]
CRS	Example	Open	Use case documents [85]
WorldVistas	Public	Open	Requirements document [86]
Concordia RE corpus	Mixed (A+P)	Open	Requirements documents [94]
PTC	Industry	Closed	Project documents [70]
MIP	Academic	Closed	Project documents [70]
TAS	Example	Open	Project documents [42]
deltaIoT	Example	Open	Project documents [42]
Mobile app store reviews	Industry	Open	App reviews
Predictor models	?	?	?

Question marks indicate that we were unable to identify the necessary information, neither in the referred paper nor from the resources indicated in the papers

in the community. (2) Scale and context are key factors for applied fields such as requirements engineering, making it more challenging to design relevant experiments.

Several empirical studies study methods or tools by applying them to a data set or document set. We categorize those as tests, see Table 6. We found three themes for tests.

1. Automatic analysis—The aim is to evaluate an algorithm or statistical method to automatically analyze a text, usually a requirements document.
2. Tool—The aim is to evaluate a tool specifically.
3. Runtime analysis—Evaluating the degree of satisfaction of quality requirements during runtime.

Tests are also fairly rigorous in that it is possible to control the study parameters well. It can also be possible to perform realistic studies, representative of real scenarios. The challenge is often to attain data that is representative. The data set is described in Table 7.

The most commonly used data set is the DePaul07 data set [27]. It consists of 15 annotated specifications from student projects at DePaul University from 2007. This data set consists of requirements specification—annotated to functional and quality requirements as well as the type of quality requirement—from student projects.

There are few examples where data from commercial projects have been used. The data do not seem to be available for use by other researchers. There are examples where data from public organizations—such as government agencies—are available and used, e.g., the EU procurement specification, see Table 7.

The most common type of data is a traditional requirements document, written in structured text. There are also a couple of instances where use case documents are used. For non-requirements specific artifacts, manuals, data use agreements, request for proposals (RFPs), and app reviews are used. From the papers in this systematic literature view, artifacts such as backlogs, feature lists, roadmaps, pull requests, or test documents do not seem to have been included.

4.1.2 Evaluations

It is usually not possible to have the same rigorous control of all study parameters in case studies [38]. However, it is often easier to have realistic scenarios, more relevant for a practical setting. We found both case studies performed in an industry context with practitioners as well as in an academic context with primarily students at different academic levels. We found 43 papers presenting case study reports on quality requirements, see Fig. 4 (all details can be found in Table 10). We separate case studies that explicitly evaluate a specific tool, method, technique, framework, etc., and exploratory case studies aiming to understand a specific context rather than evaluating something specific.

Case studies sometimes study a specific object, e.g., a tool or method, see Table 10. We found 25 case studies explicitly studying a particular object. Two objects are evaluated
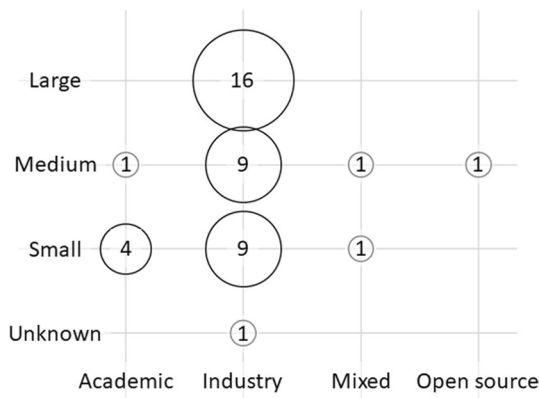
Fig. 4 Overview of the case studies on quality requirements—43 papers of the 84 included in this literature review. Scale refers to the context of the case study—small: sampling parts of the context (e.g., one part of a company) or is overall a smaller context (e.g., example system), medium: sampling a significant part of the context or a larger example system, large: sampling all significant parts of a context of an actual system (not example or made up). The context is also classified according to where the case studies are executed. Academic means primarily by students (at some academic level). Mixed means the case studies are executed in both an academic and industry context. For details, please see Table 10

We found 18 exploratory case studies on quality requirements where a specific object wasn't the focus, see Table 10. Rather, the goal is to understand a particular theme of quality requirements. Eight case studies want to understand details of quality requirements, e.g., the prevalence of a specific type of quality requirement or what happens in the lifecycle of a project. Five case studies have studied the process around quality requirements, two studies on sources of quality requirements (in particular app reviews), two studies in particular on developers' view on quality requirements (specifically using StackOverflow), and lastly one study on metric related to quality requirements. We found two longitudinal case studies.

The goal of a survey is to understand a broader context without performing any intervention [38]. Surveys can be used either very early in the research process before there is a theory to find interesting hypotheses or late in the process to understand the prevalence of a theme from a theory in a certain population. We found 15 surveys, see Table 8; 5 interviews, 9 questionnaires, one both. The goals of the surveys are a mix of understanding practices around the engineering of quality requirements and understanding actual quality requirements as such.

Table 8 Overview of the surveys found. The surveys in our study either use interviews or questionnaires—or both—to collect data. Size refers to how many participants—respondents to questionnaire or interviewees in interviews

Ref.	Year	Type	Size	Perspective	Theme
[108]	2008	Interviews	5	Development	Importance, Practices
[96]	2011	Interviews	11	Development	Prioritization
[9]	2012	Interviews	13	Architects	Importance, Practices
[97]	2012	Interviews	22	Development	Importance, Impact
[34]	2014	Interviews	16	Architects	Outsourcing, Alignment
[23]	2014	Both	48[a]	Development	Importance, Practices
[61]	2001	Questionnaire	34	Development	Alignment, Importance
[62]	2007	Questionnaire	75	Users	User satisfaction, Importance
[12]	2008	Questionnaire	133	Procurement	Importance
[51]	2010	Questionnaire	318	Development, Users	Importance
[35]	2011	Questionnaire	35	Development	Importance
[84]	2012	Questionnaire	39	Development	Importance
[11]	2012	Questionnaire	29	Development	Quality model
[47]	2015	Questionnaire	53	Development	Importance, Practices
[8]	2016	Questionnaire	56	Development	Importance

[a] 14 interviews and 34 questionnaire respondents

more than once, otherwise just one case study per object. We found no longitudinal cases; hence, the case studies are executed at one point in time and not followed up at a later time. The QUPER method is studied in several case studies in several different contexts (see Table 10). There are several case studies for the NFR method; however, it seems the context is similar or the same in most of the cases (row 2).

Overall, the surveys we found are small in terms of the sample of the population. In most cases, they do not report from which population they sample. The most common theme is the importance of quality requirements and specific sub-characteristics—typically according to ISO9126 [57] or ISO25010 [56]. However, we cannot draw any conclusions as sampling is not systematic and the population unclear. We believe it is not realistic to systematically sample any

Table 9 Action research

Reference	Year	Context	Theme
[71]	2012	Industry (Siemens)	Quality models and ISO9126.
[2]	2013	Public (University)	Using ElicitO to help intranet development at the university.
[75]	2017	Public (gov ag)	Working several years to improve processes, using ISO9126.

population and achieve a statistically significant result on how important quality requirements are nor which sub-characteristics are more or less important. We speculate that, besides the sampling challenge, the variance among organizations and point in time will likely be large, making the practical implications of such studies of questionable value.

4.1.3 Experience

In action research, the researchers are more active and part of the work than, e.g., in a case study [38]. Whereas a case study does not necessarily evaluate a specific object, action research typically reports some kind of intervention where something is changed or performed. We found 3 papers we classify as action research types of experience papers [104], see Table 9.

Two of the studies are performed at one point in time [2, 71]. One study is longitudinal, describing the changes to the processes and practices around quality requirements over several years [75]. Interestingly, all three studies directly refer to ISO9126 [57] or ISO25010 [56].

4.2 RQ2 What are the problems and challenges for quality requirements identified by empirical studies?

We have grouped the studies on quality requirements themes thematically to analyze the problems and challenges that are identified in the empirical studies. The groups are developed iteratively among the researchers, initially from the author keywords in the included papers and then iteratively refined.

4.2.1 Quality requirements and other requirements

There is an academic debate on what quality requirements are and what they should be called [48]. We found a study indicating that quality requirements—sometimes called non-functional requirements—are functional or behavioral [41]. This is in line with other studies that report that a mix of requirements type is common [14]. Two studies find that architects address quality requirements the same way as other requirements [33, 84], also confirmed in other

surveys [83]. However, there are also research studies indicating a varying prevalence and explicitness than other requirements [14, 41, 49, 80, 90]. We interpret the current state of evidence to be unclear on the handling of quality requirements. We speculate that the answers to opinion surveys might be biased towards the expected "right" answer—as expected by the researcher—rather than the actual viewpoint of the respondent.

4.2.2 Importance and prevalence of quality requirements

Many papers present results related to the importance and prevalence of quality requirements—or sub-characteristics of quality requirements. Four papers present results from artifact analysis [14, 22, 80, 90]. We found eight personal opinion survey papers [8, 9, 12, 23, 34, 35, 47, 97]. Similarly, a list of quality requirements types is developed through a survey for service-oriented applications [11]. Furthermore, we found three papers analyzing app store reviews [50, 60, 72] and two papers developer's discussions on StackOverflow [1, 110] and one paper studying 8 open-source projects communication [43]. The individual papers do not present statistical tests or variance measures. Furthermore, we found no papers elaborating on a rationale for why the distribution of sub-characteristics of quality requirements are more or less prevalent or seems as important by the subjects. We hypothesize that the importance of different quality requirements types varies over time, domain, and with personal opinion. This implies that there is no general answer to the importance of different quality requirements types. Rather, we believe it is important to adapt the quality requirements activities—such as planning and prioritization—to the specific context rather than to use predefined lists.

4.2.3 Specification of quality requirements

We found three case study papers reporting on artifact analysis of realistic requirements documents [14, 41, 90]. The practice seems to vary in how quality requirements are written; quantification, style, etc. One paper reporting on a scope decision database analysis [80]. The prevalence of quality requirements features is low and varies over time. Two interview surveys, furthermore, find quantification varies for the different cases as well as for the different quality requirements types [33, 97]. Four surveys indicate that quality requirements are often poorly documented and without template [9, 20, 23, 108]. Overall, the studies mostly report the usage of informal specification techniques (structured text) rather than specific modeling notations.

Table 10 All case study papers, with context, scale, whether they are exploratory or studying a specific solution, and the main theme.

Reference	Year	Context	Scale	Exploratory	Theme
[30]	1999	Academic	Medium	N	NFR method
[55]	2001	Academic	Small	N	QUARCC, S-COST
[31]	2004	Academic	Small	N	NFR method
[65]	2014	Academic	Small	N	SeNOR, NoRT
[29]	2015	Academic	Small	Y	Understanding QRs
[26]	1995	Industry	Small	N	NFR method
[59]	1999	Industry	Large	N	Gilb style
[67]	1999	Industry	Large	N	Quality profile
[69]	2001	Industry	Medium	Y	Metrics
[49]	2001	Industry	Medium	Y	Process evaluation
[20]	2003	Industry	Small	Y	Process evaluation
[36]	2005	Industry	Medium	N	IESE NFR Method
[92]	2007	Industry	Medium	N	Quality model
[88]	2007	Industry	Large	N	QUPER
[52]	2008	Industry	Medium	N	MOQARE
[16]	2008	Industry	Large	N	QUPER
[53]	2008	Industry	Large	Y	Understanding QRs
[101]	2009	Industry	Small	N	Light-weight process
[19]	2011	Industry	Small	N	Business process and i*
[17]	2011	Industry	Small	N	QUPER
[102]	2012	Industry	Small	N	QUAMOCO
[13]	2012	Industry	Large	N	QUPER
[83]	2012	Industry	Small	Y	Process evaluation
[89]	2012	Industry	Medium	N	NFR evaluation model
[25]	2013	Industry	Small	N	QAT Framework
[33]	2013	Industry	Large	Y	Process evaluation
[90]	2013	Industry	Medium	Y	Understanding QRs
[18]	2015	Industry	Large	N	QUPER
[46]	2015	Industry	Small	N	SQIMF
[14]	2015	Industry	Medium	Y	Understanding QRs
[41]	2016	Industry	Large	Y	Understanding QRs
[21]	2017	Industry	?	N	MERLiNN
[93]	2017	Industry	Large	N	ADEG-NFR
[50]	2017	Industry	Medium	Y	Sources
[110]	2017	Industry	Large	Y	Developers view
[103]	2018	Industry	Large	Y	Sources
[22]	2019	Industry	Large	Y	Understanding QRs
[1]	2019	Industry	Large	Y	Developers view
[80]	2019	Industry	Large	Y	Understanding QRs
[4]	2019	Industry	Large	Y	Process evaluation
[32]	2001	Mixed	Medium	N	NFR method
[79]	2007	Mixed	Small	N	QRF
[43]	2010	Open source	Medium	Y	Understanding QRs

4.2.4 Roles perspective

Different roles—for example, project manager, architect, product manager—view and work with quality requirements differently. Two interview surveys report that architects are often involved in the elicitation and definition of quality requirements [9, 33]. Furthermore, the clients or customers—in a bespoken context—are not explicit nor active in the elicitation and definition of quality requirements [33]. We found six papers collecting opinion data on the priority of quality requirements types from a role perspective [8, 9, 35, 51, 61, 97]. We did not find any particular trend nor

general view for different roles, except that when asked subjects tend to answer that quality requirements as a topic is important and explicitly handled—albeit that there are improvement potentials. Hence, it seems to us that, again, there might not be a general answer to the importance of different quality requirements types.

One study found that architects—despite being one source of quality requirements—are not involved in the scoping [34]. Another study found that relying on external stakeholders might lead to long lead-times and incomplete quality requirements [80]. We found one study on quality requirements engineering in an agile context. They report that communication and unstated assumptions are major challenges for quality requirements [4]. Even though opinion surveys indicate that subjects—independent of roles—claim to prioritize and explicitly work with quality requirements, there are indications that implicit quality requirements engineering is common and this leads to misalignment.

We find evidence of how different roles perceive and handle quality requirements to be insufficient to draw any particular conclusions.

4.2.5 Lifecycle perspective

We found two papers presenting results related to changes over time for the prevalence of different quality requirements types. Ernst and Mylopoulos study 8 open-source project [43] and Olsson et al. study scope decisions from a company [80]. Ernst and Mylopoulos did not find any specific pattern across the 8 open-source project in terms of prioritization or scoping of quality requirements. Olsson et al. concluded that there was an increase in the number of quality-oriented features and the acceptance of quality requirements in the scope decision process later in the product lifecycle compared to early in the product lifecycle. We found one student experiment on the stability of prioritization within the release of a smaller project [29]. They conclude that interoperability and reliability are more stable in terms of project priority whereas usability and security changed priority more in the release cycle. Lastly, we found a paper presenting a study on the presence of "Not a Problem" issue reports in the defect flow compared to how precise quality requirements are written [53]. The main result is that the more precise quality requirements are written, the lower the amount of "Not a Problem" issue reports.

The number of studies is small, which makes it difficult to draw any conclusions. However, we speculate that what happens over time is also likely to vary and be context specific. We hypothesize that there might be general patterns that, for example, products early in the lifecycle tend to overlook quality requirements, whereas products later in the lifecycle tend to focus more on quality requirements. Furthermore, it might also be differences in the handling of

quality requirements depending on how close to release the project is. We see these topics as relevant to study in more detail. Longitudinal studies, involving different artifacts and sources information, e.g., issue report systems, can be an interesting way forward.

4.2.6 Prioritization

We found two case studies on quality requirements prioritization [33, 96]. Berntsson Svensson et al. conducted an interview study with product managers and project leaders [96]. They found that ad hoc prioritization and priority grouping of quality requirements are the most common. Furthermore, they found that project leaders are more systematic (55% prioritize ad hoc) compared to the product managers (73% prioritize ad hoc). Daneva et al. found in their interview study that architects are commonly involved in prioritization of quality requirements [33]. They identified ad hoc and priority grouping as the most common approach to prioritization. Daneva et al., furthermore, found that 7 out of the 20 architects they interviewed considered themselves the role that sets the priority for quality requirements.

In summary, we find there is overall a lack of understanding of quality requirements prioritization. The studies indicate the involvement of different roles, which we believe warrants further research. Furthermore, the lack of systematic prioritization seems to be in line with requirements in general and not just for quality requirements.

4.2.7 Sources of quality requirements

There can be several sources of requirements, both roles as well as artifacts. As reported before, architects are sometimes involved in the elicitation and definition of quality requirements [9, 33]. Three studies have identified user reviews on mobile app markets as a potential source of quality requirements [50, 60, 103]. One study found that users are not sufficiently involved in the elicitation [49]. However, we did not find studies on, for example, usage data or customer services data as a means to elicit and analyze quality requirements.

4.3 RQ3 Which quality requirements solution proposals have been empirically validated?

Several techniques, tools, etc., have been proposed to address problem and challenges with quality requirements. The results of the evaluations or validation of different quality requirements solutions are grouped after similarity. The sections are ordered according to the number of studies.

4.3.1 Automatic analysis

One research direction which has gained popularity is different forms of automatic analysis. The idea is that a tool can be developed to support human engineers in different aspects of quality requirements engineering. All studies we found reported positive results.

We found a number of papers investigating automatic identification and classification of quality requirements from different types of sources [6, 10, 24, 28, 66, 70, 74, 86, 91, 94, 95, 100, 106, 109]. The different papers test different algorithms and approaches on different data sets. The most commonly used data set is from a project course at DePaul University from 2007. That data set has annotated requirements (functional or quality requirements) as well as quality requirements types, see Tables 6 and 7. Overall, the studies are executed in an academic context (11 out of 14) and all at a small scale which might not be representative for realistic commercial cases. Furthermore, it is often assumed the presence of requirements documents, which might not be the case for agile contexts.

We found two papers presenting studies of user reviews in app stores, e.g., Apple app store or Google play [60, 72], both rigorous. Similar to other work on automatic classification, the two studies evaluated different algorithms to identify and classify quality requirements in app reviews.

We found one paper on early aspect mining [85]. The study evaluated a tool to detect quality requirements aspects in a requirements document. Based on the detection, quality requirements are suggested to the requirements engineer. Another study evaluated a use case tool with explicit quality requirements for an agile context [44]. Both studies imply feasibility but cost or amount of effort of using them in large-scale realistic cases are not studied.

We found one study on runtime adaptations of quality requirements for self-managing systems [42]. Rather than defining fixed values for trade-offs among quality requirements, quality requirements are defined as intervals that can be optimized in runtime, depending on the specific operational conditions. They test their approach on two example systems, which show better compliance when using their approach than not.

We summarize that, while there are many tests and experiments, there are few studies of realistic scale and with realistic artifacts on automatic analysis in an quality requirements context. We also find that there is a lack of costs of running the automatic analysis, such as preparation of data, needs in terms of hardware and software, and knowledge needed by an analyst. We conclude that automatic analysis shows promise in an academic setting but has yet to be studied in a realistic scale case study or action research.

4.3.2 Goal modeling

We found two experiments on different extensions of i*, validating usefulness and correctness of the extensions compared to the original i* approach [99, 111]. The experiments are conducted in an academic setting. Both experiments conclude that the extensions are better. Based on these experiments, we cannot say anything in general about modeling of quality requirements and usefulness of i* in general.

Researchers have performed several case studies [26, 30–32]. The researchers and case context are similar and all present how the NFR method and goal modeling can work in different situations. One case study evaluated a process where business models and a quality requirements catalog are used to finally build a goal model for relevant quality requirements [19]. We found one case study using goal modeling to support product family quality requirements using goal modeling [79]. All of these case studies are of low rigor both in terms of design and validity. We have found one paper describing a test of generating goal graphs from textual requirements documents [86], and another paper testing a tool for goal modeling in an agile context [44]. Both papers indicate feasibility, i.e., the techniques seem to work in their respective context.

Overall, the evidence point to that goal modeling—in various forms—can be used and does add benefits in terms of visualization and systematic reasoning. However, we have not found any realistic scale case studies on quality requirements, nor any data on effort or impact on other parts of the development. We have not found any surveys on modeling techniques used for quality requirements. Hence, we have not found evidence of the use of goal modeling in industry specifically for quality requirements. We judge the collected evidence that goal modeling does have potential benefits but they have not been evaluated in realistic scale projects with a systematic evaluation of the whole quality requirements process.

4.3.3 Quality models and ISO9126 / ISO25010

ISO9126 [57]—and the updated version in ISO25010 [56]—is used in many of the papers we found. Al-Kilidar et al. validated the usefulness of ISO9126 in an experiment with students [3]. They conclude that the standard is difficult to interpret and too general to be useful. However, the experiment has a low rigor score—0.5 for design and 0 for validity—and lacks relevant validity description.

ADEG-NFR use ISO25010 [56] as catalogue of quality requirements types [93]. The IESE NFR method also uses the ISO standard as the basis for creating checklists and quality models [36]. Sibisi and van Waveren proposes a similar approach as the IESE NFR method, using the ISO standard as a starting point in customizing checklists and

quality model [92]. Two papers present evaluations of two approaches combining ISO9126 quality models with goal modeling [2, 19]. Another paper evaluated an approach where a checklist was derived from ISO9126 to guide the elicitation [67]. Similarly, two papers evaluate workshop and brainstorming approaches to elicitation and analysis based on ISO9126, which they propose to complement with multi-stakeholder workshops [65, 101]. Lastly, the Quamoco approach suggests connecting the abstract quality model in the ISO standard with concrete measurements [102].

We found one paper defining a catalog of quality requirements for service-oriented applications [11]. They proposed an initial list of quality requirements which was evaluated with practitioners using a questionnaire-based approach. Mohagheghi and Aparicio conducted a three-year-long project at the Norwegian Labour and Welfare Administration [75]. The aim was to improve the quality requirements. Lochmann et al. conducted a study at Siemens, Germany [71]. The business unit in question develops a traffic control system. They introduced a quality model approach to the requirements process for quality requirements. The approach is based on ISO9126 [57].

All of the approaches suggest incorporating the quality requirements specific parts into the overall requirements process as well as tailoring to the needs of the specific organization. The different approaches seem to be recognized as useful in realistic settings, leading to a more complete understanding of the quality requirements scope with a reasonable effort. It further seems as if the tailoring part is important to gain relevance and acceptance from the development organizations—especially when considering the experiment from Al-Kilidar et al. [3].

4.3.4 Prioritization and release planning

We found one method—QUPER—focused explicitly on prioritization and release planning [88]. The researchers behind the method have performed several case studies to evaluate it [13, 16–18]. We also found a prototype tool evaluation [17]. This is the single most evaluated approach. QUPER is the only approach we found which is explicitly focused on prioritization and release planning.

We found one paper proposing and evaluating an approach to handle interaction and potentially conflicting priorities among quality requirements [46]. This is similar to QUARCC, which is tested in a tool evaluation [55].

We summarize that QUPER has been evaluated both with academics and with practitioners in realistic settings. However, the long-term impact of using QUPER seems not to have been studied. However, other than QUPER, we conclude that there is no strong evidence for other solutions for prioritization and release planning of quality requirements.

4.3.5 Metrics and quality models

We found two papers evaluating the connection between key metrics to measure quality requirements types and user satisfaction [62, 69]. The results imply that quality requirements metrics—measuring the presence of quality requirements types according to ISO9126 [57] in the specifications—is correlated with user satisfaction. Hence, even though this is not a longitudinal study, there are implications that good quality requirements engineering practices might increase user satisfaction. Both studies are personal opinion surveys, which makes it difficult to evaluate causality and root-cause. Furthermore, they measured at one point in time.

We found one paper proposing and evaluating an approach to create metrics to evaluate the responses to a request for proposal (RFP) [89]. The metrics are based on recommendations from authorities and focused on process metrics rather than product metrics. They report in their case study that they could identify specifications with deficiencies in quality requirements.

We summarize that there is not a lot of evidence on the usage of metrics in connection to quality requirements. We find that the studies have identified interesting hypotheses that can be evaluated both in an academic setting through experiments or case studies as well as in real settings through case studies or action research.

4.3.6 Knowledge management

We found three papers on different aspects of knowledge management to address quality requirements engineering. Balushi et al. report on a study at the University of Manchester [2]. They applied the ElicitO framework on a project to enhance the university website. The ontology in ElicitO implements ISO9126 [57]. The MERliNN framework suggests procedures to identify and manage knowledge flows in the elicitation and analysis process [21]. One paper evaluating a tool for QUARCC and S-Cost—knowledge-based tools for handling inter-relationships among quality requirements and stakeholders [55]. All three approaches report improved completeness of quality requirements and aligned terminology among stakeholders.

We summarize that knowledge management solutions are not well studied in the quality requirements context. We also note that there are similarities between knowledge management and quality models—which is also evident as one of the studies used ISO 9126 [57].

4.3.7 Others

We found one paper evaluating MOQARE, a misuse oriented approach to find adversarial quality requirements [52]. Another paper found that creating a clearer template and

instructions for how to write a specification improved not only the quality requirements but also the attitude towards quality requirements [59]. One paper evaluates a method for how to select an appropriate technique depending on the relevant quality requirements and context factors such as lifecycle phase, etc [25]. Kopczyńska et al. experimented on a template approach for eliciting quality requirements [64]. They define a template as a regular expression. The experiment is performed with students in the third year at university. They find that using templates improved completeness and overall quality of the quality requirements. However, using templates did not speed up the elicitation process. As this is a small-scale validation, more research is needed to understand how it performs in a realistic setting.

5 Discussion

The results of our systematic literature review indicate that there are many quality requirements engineering aspects that warrant further research. The small number of studies found—84 papers over 30 years—point to a lack of studies. Furthermore, it seems to us that there is a divide between academically proposed solutions and needs of practitioners.

5.1 RQ1 Which empirical methods are used to study quality requirements?

A central research question when performing a systematic literature review is to try and answer a specific question through empirical evidence from several studies [63]. There is a tendency towards more empirical studies on quality requirements in Fig. 2. We found 1–5 papers per year in the 1990s and 4–11 papers per year in the 2010s. However, considering that the scientific community is producing more and more papers every year, the tendency to more empirical studies in quality requirements might be smaller than that of the empirical software engineering community as a whole. A naive search on Scopus for "empirical software engineering" resulted in 50–100 papers per year during the 1990s and 450–800 papers per year in the 2010s. When we further limit the result to "requirements," we end up with 1–20 in the 1990s and 100–250 papers per year in the 2010s. Ambreen et al. argued that quality requirements research is one of the emerging areas in their mapping study [7]. In their mapping study, they found 1–7 papers per year in the 1990s and 17–32 papers 2005–2012[1]. In a recent paper in IEEE Software, quality requirements or non-functional requirements occurred in less than 1% as a keyword in papers in the

Requirements Engineering Journal and at the REFSQ conference and not at all in the top-ten list for the Requirements Engineering conference [98]. This is an indication that the statement that quality requirements being an emerging area of research needs to be nuanced. We believe there is a need for further research on quality requirements. However, with the results we got from our study, it seems that the research on quality requirements might be less directed towards the practical challenges facing industry. This, however, is something that needs more research to be confirmed.

It should be noted that we have not included studies where specific sub-characteristics of quality requirements, such as security or usability, are the focus. Ambreen et al., however, included also those [7]. Hence, the figures we presented might be on the lower side regarding the number of empirical studies. At the same time, we included 84 papers, whereas Ambreen et al. found 36 papers. If we limit our results to papers before 2013—i.e., to the same period as Ambreen et al.—we found 43 papers.

In terms of the type of research performed, we see a similar distribution among validation, evaluation, and experience studies as Ambreen et al. [7]—see Table 3. Evaluation research is the most common and case study in industry the largest category. As noted in Sect. 4.1, however, the rigor overall is weak. Furthermore, replications, longitudinal, and evaluation studies of a particular solution are rare. Similarly, we found only 4 experiments that can infer that the field could benefit from larger research initiatives planning several studies over many years. This would enable researchers to plan multiple studies, combining different research methods and larger sampling of the relevant population.

5.2 RQ2 What are the problems and challenges for quality requirements identified by empirical studies?

We found studies claiming quality requirements are treated the same way as other types of requirements. We also found studies claiming companies do prioritize quality requirements and other studies claiming quality requirements are not handled properly. Furthermore, we found several studies attempting to discern which of the sub-characteristics—such as security or performance—are more important than others in a certain context. The studies we found are conducted in different contexts, domains, and different research methods. We interpret the empirical data that, on the one hand, different sub-characteristics might warrant individual attention, on the other, different research questions and methods are needed to understand and address industry-relevant challenges. We hypothesize that the requirements engineering community has not yet found a good way to analyze the quality requirements practices and challenges. We believe longitudinal studies is one way forward to deepen the

---

[1] Their results do not include papers after 2012. Hence, we choose a slightly different interval.

understanding of quality requirements over the lifecycle of a product family rather than individual products or even just one point in time. This, of course, is not isolated to quality requirements and would entail changes to how projects are funded to allow for bigger projects.

Overall, we summarize that quality requirements are written informally without a specific notation or modeling approach, there is no clear industry practice, and documentation of quality requirements seems to be performed in the same way as other requirements. Firstly, we speculate that there is a lack of understanding from practitioners of the specific needs for quality requirements, as they do not seem to prioritize separate handling. We believe the key to improving the understanding of the importance of quality requirements is to better understand the consequences and implications of quality requirements. Secondly, we believe the results imply that quality requirements engineering need to align well with other requirements topics as the cost of separate handling might deter usage.

We did not find any studies connecting quality requirements to business value nor success criteria such as timely delivery or increased sales. There are some attempts to connect user satisfaction to quality requirements and defect flows to quality requirements. There are also some studies trying to discern perspectives internally—e.g., that of architects. We also found some studies trying to understand app reviews as a source of requirements. However, we did not find any other studies attempting to identify other sources or ways of eliciting or analyzing new requirements. Interestingly, we found only one study explicitly on open-source software. There are some studies on agile methods, but we did not find any studies on DevOps, nor any studies bridging the gap between engineering and business. We propose to study quality requirements in more contexts, especially software ecosystems where commercial organizations cooperate both on development as well as operations. Furthermore, open-source is increasingly common. Hence, we believe these areas warrant more research.

We believe—at least for certain quality requirements sub-characteristics—it is key to understand the actual usage by actual users to discern which quality requirements to address. We propose data-driven approaches as an important trend for quality requirements, seen with different automatic analysis techniques of app stores. Another example of unexploited potential is customer service data—whether through, e.g., issues or reviews. We lack a clear data-driven perspective, where usage data, as one example, is studied with a quality requirements perspective. Furthermore, we believe the community needs to understand and address the connection between quality requirements and external factors such as business value or project success. An often-cited study by Finkelstein et al. claim quality requirements as one source of problem [45]. We recognize, though, that

this type of research is both challenging to design, expensive to perform, and difficult to get rigorous and relevant results. We see a need for future research in quality requirements to study not quality requirements in isolation but as part of larger studies where quality requirements are one of the research questions.

5.3 RQ3 Which quality requirements solution proposals have been empirically validated?

Quality models—as an over-arching principle—is the most common approach to elicitation. The different approaches typically propose tailoring of a generic quality model—often ISO9126 [57] or ISO25010 [56]—and a process with workshops to elicit and analyze quality requirements. The studies overwhelmingly report success with such strategies, albeit few have concrete and validated numbers for effort nor lead-time. However, it seems to us that there is no solid evidence of the cost-effectiveness of quality models, a lack of evidence of adherence over time, nor clear data for success factors from a more complete scope when using quality models. We see an opportunity for companies already using—or willing to introduce—quality models which should make it reasonable to conduct longitudinal studies on quality models as future work.

We only found one modeling approach—goal modeling. Goal modeling is—similar to quality models—well researched. There are experiments and case studies. However, we could not find any surveys nor action research. Goal modeling is, just as quality models, integrated often integrated into a process encompassing all aspects of quality requirements. We interpret the lack of quality requirements modeling studies as follows: (1) Modeling of quality requirements cannot be separated from the modeling of other requirements. This is in line with quality requirements artifact studies, which often reports similar or identical specification for quality requirements and other requirements. (2) The modeling solutions proposed by academia are not something practitioners see as applicable or relevant.

We found very few studies on data-driven requirements engineering [73] in the context of quality requirements. Rather, there seems to be a focus on the requirements specification, e.g., with quality models, goal modeling, and with the automatic analysis studies on mining specifications for quality requirements. Furthermore, we could not find any approaches integrating quality requirements engineering with, for example, DevOps and continuous experimentation. We see a gap for studies on how user feedback (reviews, customer service data, etc.) and usage data (measurements when using a software product or service) can be used for quality requirements prioritization, elicitation, and release planning. Furthermore, we propose that evaluating the trend over time as a means to better understand the connection between

quality requirements and user satisfaction. Especially, we see an opportunity with the lead-time from an improvement in the quality requirements—through the implementation—and the lead-time from downward-trending user satisfaction to actions taken—using various measurements—are important to develop relevant early forecasting metrics and improved prioritization mechanisms which consider estimations of the user satisfaction.

We note that the validation studies we found tend to report an improvement and rarely conclude that the proposals as a whole do not work. The explanations can be many, but publication bias can be one, another might be confirmation bias. We see this as an indication that empirical software engineering as an area is still developing and maturing.

5.4 Limitations and threats to validity

5.4.1 Construct validity

Construct validity refers to the decisions on method and tools, and whether they are appropriate for the research questions. We utilized a hybrid search strategy. The risk is if the start set for the snowballing approach is insufficient, all papers will not be found when snowballing. However, Mourao et al. recently published an evaluation that a hybrid strategy is an appropriate alternative [76]. Hence, the hybrid method is considered to be appropriate for our research questions.

Start set I includes 274 papers between the years 1991 and 2014. Start set II includes 173 papers between 1990 and 2019—of which 70 are from the period 2015–2019. The snowballing iterations included 83 papers from 1976–2019. 447 of the 530 (84%) of papers screened come from Start set I and the Start set II of papers. Furthermore, based on our experience, we believe we have included several key papers on empirical evidence on quality requirements. Hence, we believe that the fact that most of the included papers come from Start set I and the Start set II indicates that we have likely found the majority of all relevant papers. This implies that our method selection is appropriate.

5.4.2 Internal validity

A validity threat is if papers are excluded even though they should have been included or erroneous classification. We ensured that all border-line cases were screened by at least two researchers and a sample of all papers was also screened by at least two researchers to mitigate this, see Sect. 3.4. We also followed a pre-defined method thoroughly. In the end, the process resulted in:

- For Start set I, all papers were screened by at least 2 researchers.

- For Start set II, 61% of the papers were screened by at least two researchers.
- For the two snowballing iterations, 29% and 60% of the papers respectively were screened by at least two researchers.
- 66% of the 193 papers included in the full read were read by at least two researchers, including reviewing the classification.
- 83% of excluded papers in the screening step and 77% of the excluded papers in the full read step were read by at least two researchers.

Excluded papers were reviewed more often by at least two researchers than included papers, mitigating the threat of excluding papers that should be included. We argue that threats to internal validity are low.

There is a risk that we missed relevant papers are we excluded relevant papers as we excluded papers focusing on a specific type of quality requirements, e.g., performance or security. The risk is related to, on the one hand, terminology and, on the other, that the specific empirical study is on a specific type but the method or phenomena applies to quality requirements in general. For the former, we believe that our selection of search terms in the extended step is by far the most prevalent, hence, it should not be a big problem. Furthermore, since we also use a snowballing approach, this threat is further minimized. For the latter, we cannot completely dismiss the threat as some empirical evidence might not be presented on other papers even though the results might be applicable. We excluded 5 papers based on these criteria. Hence, we conclude that even though this is a threat, it is not likely to largely impact the internal from our paper.

5.4.3 Conclusion validity

We followed a systematic process to address threats to the conclusion validity. Furthermore, we report the steps and results in such a way that it should be possible to replicate them. The threat to conclusion is the inclusion/exclusion primarily, which entails a human judgment and thereby susceptible for errors. However, as mentioned for internal validity threats, we used a peer review process among the authors to minimize the threats of human errors.

5.4.4 External validity

External validity concerns the applicability of the results of our study. We believe the systematic hybrid process limits this threat as we do not exclude research communities nor do we exclude studies even if particular keywords are missing. However, we did not analyze different domains in detail as that information was not available in sufficient detail in enough studies. It might be that different domains exhibit

different characteristics in terms of quality requirements engineering. Hence, the results should be applied after careful consideration.

6 Conclusion

The results of our systematic literature review indicate that there are many quality requirements engineering aspects that warrant further research. We judge that 84 papers over 30 years point to a lack of studies. However, this is something that should be studied in more detail to be confirmed. Furthermore, it seems to us that there is a divide between academically proposed solutions accepted by practitioners. The proposed solutions are rarely evaluated in realistic settings—and replications are non-existent. Furthermore, practitioners rarely report using any specific approach for quality requirements. At the same time, the existing surveys are small, have an unclear sample and population, and are rarely connected to any theory. We, therefore, hypothesize that overall, there is a lack of clear empirical evidence for what software developing organizations should adopt. This, again, is something that warrants further research to understand the needs of practitioners and their relation to proposed solutions found in the literature.

For practitioners, there are some recommendations of what has worked in realistic contexts. Quality models with the associated processes, QUPER, and the NFR method have been reported as useful in several studies. However, it is not clear what the return of investment is nor the long-term effect. Still, we believe our results indicate those to be a good starting point if an organization should improve their quality requirements practices. Furthermore, goal modeling has been evaluated in academic settings with positive results. However, we could not find any evaluations in a realistic setting specifically for quality requirements. In the context where a document or specification is received, different automatic analysis approaches seem to be able to help in identifying quality requirements. However, we could not find any available tools nor clear integration in the overall software engineering process. Hence, even though these solutions show potential, the effort needed to apply them in practice is unclear.

For researchers, we see a need for longitudinal studies on quality requirements. There are examples of solutions evaluated at one point in time. However, we could not find any studies on the long-term effect and costs of changing how companies work with quality requirements. We believe that the product or portfolio lifecycle is particularly underresearched. Furthermore, we believe there is a lack of understanding of the challenges and needs in realistic settings, as the solutions proposed by researchers seem to fail in getting acceptance from practitioners. This is a rather difficult issue

for individuals to address, rather the requirements engineering community should try to establish a new way of performing research where larger and longer studies are viable.

Furthermore, there are only a few studies on sources of quality requirements in general and data-driven alternatives specifically. We believe there is potential in sources such as usage data, customer service data, and continuous experimentation to complement stakeholder analysis, expert input, and focus groups. The former has the potential to take in a breadth of input closer to the actual users while the latter will focus on fewer persons' opinions or experiences which will be less representative of the actual usage.

We limited our systematic literature review to quality requirements in general and excluding sub-categories such as security or usability. We believe it would be interesting to perform a similar study on the different sub-categories. For one, there might be differences in the sub-categories both regarding the strength of evidence and the types of solutions proposed. On the other, it might be that it does not make sense to have one solution for all types of quality requirements categories.

References

1. Ahmad A, Feng C, Li K, Asim SM, Sun T (2019) Toward empirically investigating non-functional requirements of ios developers on stack overflow. IEEE Access 7:61145–61169. https://doi.org/10.1109/access.2019.2914429
2. Al Balushi TH, Sampaio PRF, Loucopoulos P (2013) Eliciting and prioritizing quality requirements supported by ontologies: a case study using the e licit o framework and tool. Expert Syst 30(2):129–151
3. Al-Kilidar H, Cox K, Kitchenham B (2005) The use and usefulness of the iso/iec 9126 quality standard. In: International symposium on empirical software engineering. IEEE, pp 126–132. https://doi.org/10.1109/ISESE.2005.1541821
4. Alsaqaf W, Daneva M, Wieringa R (2019) Quality requirements challenges in the context of large-scale distributed agile: an empirical study. Inf Softw Technol 110:39–55. https://doi.org/10.1016/j.infsof.2019.01.009

5. Alsaqaf W, Daneva M, Wieringa R (2017) Quality requirements in large-scale distributed agile projects–a systematic literature review. In: International working conference on requirements engineering: foundation for software quality. Springer, pp 219–234. https://doi.org/10.1007/978-3-319-54045-0_17

6. Amasaki S, Leelaprute P (2018) The effects of vectorization methods on non-functional requirements classification. In: 2018 44th euromicro conference on software engineering and advanced applications (SEAA). IEEE, pp. 175–182. https://doi.org/10.1109/seaa.2018.00036

7. Ambreen T, Ikram N, Usman M, Niazi M (2018) Empirical research in requirements engineering: trends and opportunities. Requir Eng 23(1):63–95. https://doi.org/10.1007/s00766-016-0258-2

8. Ameller D, Galster M, Avgeriou P, Franch X (2016) A survey on quality attributes in service-based systems. Softw Qual J 24(2):271–299

9. Ameller D, Ayala C, Cabot J, Franch X (2012) How do software architects consider non-functional requirements: an exploratory study. In: 2012 20th IEEE international requirements engineering conference (RE). IEEE, pp 41–50. https://doi.org/10.1109/re.2012.6345838

10. Baker C, Deng L, Chakraborty S, Dehlinger J (2019) Automatic multi-class non-functional software requirements classification using neural networks. In: 2019 IEEE 43rd annual computer software and applications conference (COMPSAC), vol 2. IEEE, pp 610–615. https://doi.org/10.1109/compsac.2019.10275

11. Becha H, Amyot D (2012) Non-functional properties in service oriented architecture - a consumer's perspective. J Softw 7(3):575–587. https://doi.org/10.4304/jsw.7.3.575-587

12. Benslimane Y, Cysneiros LM, Bahli B (2007) Assessing critical functional and non-functional requirements for web-based procurement systems: a comprehensive survey. Requir Eng 12(3):191–198. https://doi.org/10.1007/s00766-007-0050-4

13. Berntsson Svensson R, Sprockel Y, Regnell B, Brinkkemper S (2012) Setting quality targets for coming releases with quper: an industrial case study. Requir Eng 17(4):283–298. https://doi.org/10.1007/s00766-011-0125-0

14. Berntsson Svensson R, Olsson T, Regnell B (2013) An investigation of how quality requirements are specified in industrial practice. Inf Softw Technol 55(7):1224–1236. https://doi.org/10.1016/j.infsof.2013.01.006

15. Berntsson Svensson R, Host M, Regnell B (2010) Managing quality requirements: a systematic review. In: 2010 36th EUROMICRO conference on software engineering and advanced applications. IEEE, pp 261–268. https://doi.org/10.1109/seaa.2010.55

16. Berntsson Svensson R, Olsson T, Regnell B (2008) Introducing support for release planning of quality requirements-an industrial evaluation of the quper model. In: Second international workshop on software product management. IEEE, pp 18–26

17. Berntsson Svensson R, Parker PL, Regnell B (2011) A prototype tool for quper to support release planning of quality requirements. In: Fifth international workshop on software product management (IWSPM). IEEE, pp 57–66. https://doi.org/10.1109/iwspm.2011.6046204

18. Berntsson Svensson R, Regnell B (2015) A case study evaluation of the guideline-supported quper model for elicitation of quality requirements. In: International working conference on requirements engineering: foundation for software quality. Springer, pp 230–246. https://doi.org/10.1007/978-3-319-16101-3_15

19. Bittencourt RV, De Araujo RM (2011) Using business models to identify expected quality for information systems. Int J Bus Process Integr Manag 5(2):159–172. https://doi.org/10.1504/ijbpim.2011.040207

20. Borg A, Yong A, Carlshamre P, Sandahl K (2003) The bad conscience of requirements engineering: an investigation in real-world treatment of non-functional requirements. In: Software engineering research and practice in Sweden

21. Buitr SL, Pino FJ, Flores-Rios BL, Ibarra-Esquer JE, et al. (2017) A model for enhancing tacit knowledge flow in non-functional requirements elicitation. In: 2017 5th international conference in software engineering research and innovation (CONISOFT). IEEE, pp 25–33. https://doi.org/10.1109/conisoft.2017.00011

22. Calazans TA, Paldês RÁ, Canedo ED, Masson TE, Guimares FdA, Rezende MK, Braosi E, Kosloski RAD (2019) Quality requirements: analysis of utilization in the systems of a financial institution. In: ACM/IEEE international symposium on empirical software engineering and measurement (ESEM). IEEE, pp 1–6. https://doi.org/10.1109/esem.2019.8870151

23. Caracciolo A, Lungu MF, Nierstrasz O (2014) How do software architects specify and validate quality requirements?. In: European conference on software architecture. Springer, pp 374–389. https://doi.org/10.1007/978-3-319-09970-5_32

24. Casamayor A, Godoy D, Campo M (2010) Identification of non-functional requirements in textual specifications: a semi-supervised learning approach. Inf Softw Technol 52(4):436–445. https://doi.org/10.1016/j.infsof.2009.10.010

25. Chiam YK, Staples M, Ye X, Zhu L (2013) Applying a selection method to choose quality attribute techniques. Inf Softw Technol 55(8):1419–1436. https://doi.org/10.1016/j.infsof.2013.02.001

26. Chung L, Nixon BA (1995) Dealing with non-functional requirements: three experimental studies of a process-oriented approach. In: 17th international conference on software engineering. IEEE, pp 25–25. https://doi.org/10.1145/225014.225017

27. Cleland-Huang, J., Mazrouee S, Liguo H, Port D (2007) nfr. https://doi.org/10.5281/zenodo.268542

28. Cleland-Huang J, Settimi R, Zou X, Solc P (2007) Automated classification of non-functional requirements. Requir Eng 12(2):103–120. https://doi.org/10.1007/s00766-007-0045-1

29. Condori-Fernandez N, Lago P (2015) Can we know upfront how to prioritize quality requirements? In: 2015 IEEE fifth international workshop on empirical requirements engineering (EmpiRE). IEEE, pp 33–40. https://doi.org/10.1109/empire.2015.7431305

30. Cysneiros LM, do Prado Leite JS (1999) Integrating non-functional requirements into data modeling. In: Proceedings IEEE international symposium on requirements engineering. IEEE, pp 162–171

31. Cysneiros LM, do Prado Leite JCS (2004) Nonfunctional requirements: from elicitation to conceptual models. IEEE Trans Software Eng 30(5):328–350. https://doi.org/10.1109/tse.2004.10

32. Cysneiros LM, do Prado Leite JCS, Neto JDMS (2001) A framework for integrating non-functional requirements into conceptual models. Requir Eng 6(2):97–115. https://doi.org/10.1007/s007660170008

33. Daneva M, Buglione L, Herrmann A (2013) Software architects' experiences of quality requirements: what we know and what we do not know? In: international working conference on requirements engineering: foundation for software quality. Springer, pp 1–17. https://doi.org/10.1007/978-3-642-37422-7_1

34. Daneva M, Marczak S, Herrmann A (2014) Engineering of quality requirements as perceived by near-shore development centers' architects in eastern europe: the hole in the whole. In: Proceedings of the 8th ACM/IEEE international symposium on empirical software engineering and measurement. ACM, p 19. https://doi.org/10.1145/2652524.2652534

35. De La Vara JL, Wnuk K, Berntsson-Svensson R, Sánchez J, Regnell B (2011) An empirical study on the importance of quality requirements in industry. In: Proceedings of the 23rd international conference on software engineering and knowledge engineering. pp 438–443. https://doi.org/10.1007/s10270-008-0108-x

36. Doerr J, Kerkow D, Koenig T, Olsson T, Suzuki T (2005) Non-functional requirements in industry-three case studies adopting an experience-based nfr method. In: 13th IEEE international conference on requirements engineering (RE'05). IEEE, pp 373–382. https://doi.org/10.1109/re.2005.47

37. Dörr J, Kerkow D, Von Knethen A, Paech B (2003) Eliciting efficiency requirements with use cases. In: Ninth international workshop on requirements engineering: foundation for software quality. In conjunction with CAiSE, vol 3

38. Easterbrook S, Singer J, Storey MA, Damian D (2008) Selecting empirical methods for software engineering research. In: Guide to advanced empirical software engineering. Springer, pp 285–311. https://doi.org/10.1007/978-1-84800-044-5_11

39. Ebert C (1998) Putting requirement management into praxis: dealing with nonfunctional requirements. Inf Softw Technol 40(3):175–185. https://doi.org/10.1016/s0950-5849(98)00049-4

40. Eckhardt J, Vogelsang A, Fernández DM (2016) Are "non-functional" requirements really non-functional? an investigation of non-functional requirements in practice. In: 2016 IEEE/ACM 38th international conference on software engineering (ICSE). IEEE, pp 832–842. https://doi.org/10.1145/2884781.2884788

41. Eckhardt J, Vogelsang A, Fernández DM (2016) Are "non-functional" requirements really non-functional? an investigation of non-functional requirements in practice. In: IEEE/ACM 38th international conference on software engineering (ICSE). IEEE, pp 832–842. https://doi.org/10.1145/2884781.2884788

42. Edwards R, Bencomo N (2018) Desire: further understanding nuances of degrees of satisfaction of non-functional requirements trade-off. In: Proceedings of the 13th international conference on software engineering for adaptive and self-managing systems. ACM, pp 12–18. https://doi.org/10.1145/3194133.3194142

43. Ernst NA, Mylopoulos J (2010) On the perception of software quality requirements during the project lifecycle. In: International working conference on requirements engineering: foundation for software quality. Springer, pp 143–157. https://doi.org/10.1007/978-3-642-14192-8_15

44. Farid WM, Mitropoulos FJ (2012) Normatic: A visual tool for modeling non-functional requirements in agile processes. In: Proceedings of IEEE southeastcon. IEEE, pp 1–8 https://doi.org/10.1109/secon.2012.6196989

45. Finkelstein A, Dowell J (1996) A comedy of errors: the london ambulance service case study. In: Proceedings of the 8th International Workshop on Software Specification and Design. pp. 2–4. IEEE

46. García-Mireles GA, De La Rubia MÁM, García F, Piattini M (2015) Identifying quality characteristic interactions during software development. In: International conference on evaluation of novel approaches to software engineering (ENASE). IEEE, pp 78–89. https://doi.org/10.5220/0005351800780089

47. García-Mireles GA (2016) Identifying relevant product quality characteristics in the context of very small organizations. Comput Sci Inf Syst 13(3):875–900. https://doi.org/10.2298/csis160809034g

48. Glinz M (2007) On non-functional requirements. In: 15th IEEE international requirements engineering conference (RE 2007). IEEE, pp 21–26. https://doi.org/10.1109/re.2007.45

49. Grimshaw DJ, Draper GW (2001) Non-functional requirements analysis: deficiencies in structured methods. Inf Softw Technol 43(11):629–634. https://doi.org/10.1016/s0950-5849(01)00171-9

50. Groen EC, Kopczyńska S, Hauer MP, Krafft TD, Doerr J (2017) Users-the hidden software product quality experts?: a study on how app users report quality aspects in online reviews. In: IEEE 25th international requirements engineering conference (RE). IEEE, pp 80–89. https://doi.org/10.1109/re.2017.73

51. Haigh M (2010) Software quality, non-functional software requirements and it-business alignment. Software Qual J 18(3):361–385. https://doi.org/10.1007/s11219-010-9098-3

52. Herrmann A, Paech B (2008) Moqare: misuse-oriented quality requirements engineering. Requir Eng 13(1):73–86

53. Ho CW, Williams L, Robinson B (2008) Examining the relationships between performance requirements and "not a problem" defect reports. In: 16th IEEE international requirements engineering conference. IEEE, pp 135–144

54. Höst M, Regnell B, Wohlin C (2000) Using students as subjects-a comparative study of students and professionals in lead-time impact assessment. Empir Softw Eng 5(3):201–214. https://doi.org/10.1023/A:1026586415054

55. In H, Boehm BW (2001) Using winwin quality requirements management tools: a case study. Ann Softw Eng 11(1):141–174

56. ISO/IEC 25010:2011(E) (2011) Systems and software engineering - systems and software quality requirements and evaluation (SquaRE)

57. ISO/IEC 9126-1:2001 (2001) Software engineering - product quality - part 1: quality model

58. Ivarsson M, Gorschek T (2011) A method for evaluating rigor and industrial relevance of technology evaluations. Empir Softw Eng 16(3):365–395. https://doi.org/10.1007/s10664-010-9146-4

59. Jacobs S (1999) Introducing measurable quality requirements: a case study. In: Proceedings IEEE international symposium on requirements engineering (Cat. No. PR00188). IEEE, pp 172–179

60. Jha N, Mahmoud A, Mining non-functional requirements from app store reviews. Empir Softw Eng 1–37. https://doi.org/10.1007/s10664-019-09716-7

61. Johansson E, Wesslén A, Bratthall L, Host M (2001) The importance of quality requirements in software platform development-a survey. In: Proceedings of the 34th annual hawaii international conference on system sciences. IEEE, p 10. https://doi.org/10.1109/HICSS.2001.927252

62. Jung HW (2007) Validating the external quality subcharacteristics of software products according to iso/iec 9126. Comput Stand Interfaces 29(6):653–661. https://doi.org/10.1016/j.csi.2007.03.004

63. Kitchenham B, Charters S (2007) Guidelines for performing systematic literature reviews in software engineering

64. Kopczyńska S, Nawrocki J, Ochodek M (2018) An empirical study on catalog of non-functional requirement templates: usefulness and maintenance issues. Inf Softw Technol 103:75–91. https://doi.org/10.1016/j.infsof.2018.06.009

65. Kopczyńska S, Nawrocki J (2014) Using non-functional requirements templates for elicitation: a case study. In: 2014 IEEE 4th international workshop on requirements patterns (RePa). IEEE, pp 47–54. https://doi.org/10.1109/repa.2014.6894844

66. Kurtanović Z, Maalej W (2017) Automatically classifying functional and non-functional requirements using supervised machine learning. In: 2017 IEEE 25th international requirements engineering conference (RE). IEEE, pp 490–495. https://doi.org/10.1109/re.2017.82

67. Kusters RJ, van Solingen R, Trienekens JJ (1999) Identifying embedded software quality: two approaches. Qual Reliab Eng Int 15(6):485–492

68. Lawrence B, Wiegers K, Ebert C (2001) The top risk of requirements engineering. IEEE Softw 18(6):62–63. https://doi.org/10.1109/52.965804

69. Leung HK (2001) Quality metrics for intranet applications. Inf Manag 38(3):137–152. https://doi.org/10.1016/s0378-7206(00)00060-4

70. Lian X, Cleland-Huang J, Zhang L (2017) Mining associations between quality concerns and functional requirements. In: 2017

IEEE 25th international requirements engineering conference (RE). IEEE, pp 292–301. https://doi.org/10.1109/re.2017.68

71. Lochmann K, Fernández DM, Wagner S (2012) A case study on specifying quality requirements using a quality model. In: 19th Asia-Pacific software engineering conference, vol 1. IEEE, pp. 577–582. https://doi.org/10.1109/apsec.2012.57

72. Lu M, Liang P (2017) Automatic classification of non-functional requirements from augmented app user reviews. In: Proceedings of the 21st international conference on evaluation and assessment in software engineering. ACM, pp 344–353. https://doi.org/10.1145/3084226.3084241

73. Maalej W, Nayebi M, Johann T, Ruhe G (2015) Toward data-driven requirements engineering. IEEE Softw 33(1):48–54. https://doi.org/10.1109/MS.2015.153

74. Mahmoud A, Williams G (2016) Detecting, classifying, and tracing non-functional software requirements. Requir Eng 21(3):357–381. https://doi.org/10.1007/s00766-016-0252-8

75. Mohagheghi P, Aparicio ME (2017) An industry experience report on managing product quality requirements in a large organization. Inf Softw Technol 88:96–109. https://doi.org/10.1016/j.infsof.2017.04.002

76. Mourão E, Pimentel JF, Murta L, Kalinowski M, Mendes E, Wohlin C (2020) On the performance of hybrid search strategies for systematic literature reviews in software engineering. Inf Softw Technol. https://doi.org/10.1016/j.infsof.2020.106294

77. Mourão E, Kalinowski M, Murta L, Mendes E, Wohlin C (2017) Investigating the use of a hybrid search strategy for systematic reviews. In: International symposium on empirical software engineering and measurement (ESEM). IEEE, pp 193–198. https://doi.org/10.1109/esem.2017.30

78. Mylopoulos J, Chung L, Nixon B (1992) Representing and using nonfunctional requirements: a process-oriented approach. IEEE Trans Softw Eng 6:483–497. https://doi.org/10.1109/32.142871

79. Niemelä E, Immonen A (2007) Capturing quality requirements of product family architecture. Inf Softw Technol 49(11–12):1107–1120. https://doi.org/10.1016/j.infsof.2006.11.003

80. Olsson T, Wnuk K, Gorschek T (2019) An empirical study on decision making for quality requirements. J Syst Softw 149:217–233. https://doi.org/10.1016/j.jss.2018.12.002

81. Ouhbi S, Idri A, Fernández-Alemán JL, Toval A (2013) Software quality requirements: a systematic mapping study. In: 20th Asia-Pacific software engineering conference (APSEC), vol 1. IEEE, pp 231–238. https://doi.org/10.1109/apsec.2013.40

82. Paech B, Kerkow D (2004) Non-functional requirements engineering-quality is essential. In: 10th international workshop on requirements engineering foundation for software quality

83. Phillips LB, Aurum A, Svensson RB (2012) Managing software quality requirements. In: 38th euromicro conference on software engineering and advanced applications. IEEE, pp 349–356. https://doi.org/10.1109/seaa.2012.65

84. Poort ER, Martens N, Van De Weerd I, Van Vliet H (2012) How architects see non-functional requirements: beware of modifiability. In: International working conference on requirements engineering: foundation for software quality. Springer, pp 37–51. https://doi.org/10.1007/978-3-642-28714-5_4

85. Rago A, Marcos C, Diaz-Pace JA (2013) Uncovering quality-attribute concerns in use case specifications via early aspect mining. Requir Eng 18(1):67–84

86. Rahimi M, Mirakhorli M, Cleland-Huang J (2014) Automated extraction and visualization of quality concerns from requirements specifications. In: IEEE 22nd international requirements engineering conference (RE). IEEE, pp 253–262. https://doi.org/10.1109/re.2014.6912267

87. Regnell B, Svensson RB, Olsson T (2008) Supporting roadmapping of quality requirements. IEEE Softw 25(2):42–47. https://doi.org/10.1109/ms.2008.48

88. Regnell B, Höst M, Svensson RB (2007) A quality performance model for cost-benefit analysis of non-functional requirements applied to the mobile handset domain. In: International working conference on requirements engineering: foundation for software quality. Springer, pp 277–291. https://doi.org/10.1007/978-3-540-73031-6_20

89. Saito Y, Monden A, Matsumoto K (2012) Evaluation of non functional requirements in a request for proposal (rfp). In: Joint conference of the 22nd international workshop on software measurement and the 2012 seventh international conference on software process and product measurement. IEEE, pp 106–111. https://doi.org/10.1109/iwsm-mensura.2012.23

90. Shahrokni A, Feldt R (2013) Industrial challenges with quality requirements in safety critical software systems. In: 39th Euromicro conference on software engineering and advanced applications. IEEE, pp 78–81. https://doi.org/10.1109/seaa.2013.57

91. Sharma VS, Ramnani RR, Sengupta S (2014) A framework for identifying and analyzing non-functional requirements from text. In: Proceedings of the 4th international workshop on twin peaks of requirements and architecture. ACM, pp 1–8. https://doi.org/10.1145/2593861.2593862

92. Sibisi M, Van Waveren CC (2007) A process framework for customising software quality models. In: AFRICON. IEEE, pp 1–8. https://doi.org/10.1109/afrcon.2007.4401495

93. Silva A, Pinheiro PR, Albuquerque A, Barroso J (2017) Evaluation of an approach to define elicitation guides of non-functional requirements. IET Softw 11(5):221–228. https://doi.org/10.1049/iet-sen.2016.0302

94. Singh P, Singh D, Sharma A (2016) Rule-based system for automated classification of non-functional requirements from requirement specifications. In: 2016 International conference on advances in computing, communications and informatics (ICACCI). IEEE, pp 620–626. https://doi.org/10.1109/icacci.2016.7732115

95. Slankas J, Williams L (2013) Automated extraction of non-functional requirements in available documentation. In: 2013 1st international workshop on natural language analysis in software engineering (NaturaLiSE). IEEE, pp 9–16. https://doi.org/10.1109/naturalise.2013.6611715

96. Svensson RB, Gorschek T, Regnell B, Torkar R, Shahrokni A, Feldt R, Aurum A (2011) Prioritization of quality requirements: state of practice in eleven companies. In: IEEE 19th international requirements engineering conference. IEEE, pp 69–78. https://doi.org/10.1109/re.2011.6051652

97. Svensson RB, Gorschek T, Regnell B, Torkar R, Shahrokni A, Feldt R (2011) Quality requirements in industrial practice-an extended interview study at eleven companies. IEEE Trans Softw Eng 38(4):923–935. https://doi.org/10.1109/tse.2011.47

98. Tenbergen B, Daun M (2019) Is requirements-engineering research delivering what it promised?: a review of its accomplishments and opportunities after 10 years. IEEE Softw 36(4):6–11. https://doi.org/10.1109/ms.2019.2909127

99. Teruel MA, Navarro E, López-Jaquero V, Montero F, Jaen J, González P (2012) Analyzing the understandability of requirements engineering languages for cscw systems: a family of experiments. Inf Softw Technol 54(11):1215–1228. https://doi.org/10.1016/j.infsof.2012.06.001

100. Tóth L, Vidács L (2018) Study of various classifiers for identification and classification of non-functional requirements. In: International conference on computational science and its applications. Springer, pp 492–503. https://doi.org/10.1007/978-3-319-95174-4_39

101. Vanhanen J, Mäntylä MV, Itkonen J (2009) Lightweight elicitation and analysis of software product quality goals: a multiple industrial case study. In: 2009 third international workshop on

software product management. IEEE, pp 42–52. https://doi.org/10.1109/iwspm.2009.5

102. Wagner S, Lochmann K, Heinemann L, Kläs M, Trendowicz A, Plösch R, Seidl A, Goeb A, Streit J (2012) The quamoco product quality modelling and assessment approach. In: Proceedings of the 34th international conference on software engineering. IEEE Press, pp 1133–1142. https://doi.org/10.1109/icse.2012.6227106

103. Wang T, Liang P, Lu M (2018) What aspects do non-functional requirements in app user reviews describe? an exploratory and comparative study. In: 2018 25th Asia-pacific software engineering conference (APSEC). IEEE, pp 494–503. https://doi.org/10.1109/apsec.2018.00064

104. Wieringa R, Maiden N, Mead N, Rolland C (2006) Requirements engineering paper classification and evaluation criteria: a proposal and a discussion. Requir Eng 11(1):102–107. https://doi.org/10.1007/s00766-005-0021-6

105. Wohlin C (2014) Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: International conference on evaluation and assessment in software engineering. p 38. https://doi.org/10.1145/2601248.2601268

106. Younas M, Jawawi D, Ghani I, Shah MA Extraction of non-functional requirement using semantic similarity distance. Neural Comput Appl 1–15. https://doi.org/10.1007/s00521-019-04226-5

107. Yu E (1997) Towards modelling and reasoning support for early-phase requirements engineering. In: International symposium on requirements engineering. pp 226–35. https://doi.org/10.1109/ISRE.1997.566873

108. Yusop N, Zowghi D, Lowe D (2008) The impacts of non-functional requirements in web system projects. Int J Value Chain Manag. https://doi.org/10.1504/ijvcm.2008.016116

109. Zhang W, Yang Y, Wang Q, Shu F (2011) An empirical study on classification of non-functional requirements. In: The twenty-third international conference on software engineering and knowledge engineering (SEKE 2011). pp 190–195

110. Zou J, Xu L, Yang M, Zhang X, Yang D (2017) Towards comprehending the non-functional requirements through developers eyes: an exploration of stack overflow using topic analysis. Inf Softw Technol 84:19–32. https://doi.org/10.1016/j.infsof.2016.12.003

111. Zubcoff J, Garrigós I, Casteleyn S, Mazón JN, Aguilar JA, Gomariz-Castillo F (2019) Evaluating different i*-based approaches for selecting functional requirements while balancing and optimizing non-functional requirements: a controlled experiment. Inf Softw Technol 106:68–84. https://doi.org/10.1016/j.infsof.2018.09.004

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.