Friedrich-Alexander-Universität Erlangen-Nürnberg
Technische Fakultät, Department Informatik

Katsiaryna KRAUCHANKA

MASTER THESIS

Validating Requirements Specification Using Surveys

Submitted on 14.09.2016

Supervisor: Prof. Dr. Dirk Riehle, M.B.A.
Andreas Kaufmann, M.Sc.
Professur für Open-Source-Software
Department Informatik, Technische Fakultät
Friedrich-Alexander University Erlangen-Nürnberg

Versicherung

Ich versichere, dass ich die Arbeit ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen angefertigt habe und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen wurde. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.

[CITY], [DATE]

Abstract

The quality of the software requirements is vital to a project's success, and the means of assessing the software requirements specifications quality are of high importance. This thesis presents a novel method of validation large requirements specifications using surveys. We developed a detailed method description, sketching a draft at the beginning of the research and constantly improving it after each method application on different industrial projects at Siemens AG. Moreover, the results were compared with the linguistic analysis of ambiguity indicators that present in the requirements and with the expert assessment on the requirements clearness. The result of method application is the metric, which shows the portion of non ambiguous requirements in the software requirements specification under study.

Keywords: Software Requirement Specification, Requirements Engineering, Requirements Validation

Table of contents

Table of Contents

List of figures

List of tables

List of abbreviations

NL - Natural Language

SDL - Software Development Lifecycle

SR - Software Requirement

SRS - Software Requirement Specification

TBD - to be determined

1. Introduction

1.1 Original Thesis Goals

One of the original thesis goals was to document the method of validation large requirements specifications using surveys and develop the metric in order that the method could be correctly applied by everyone interested to determine the quality of the SRS. Besides, we aimed at finding out how efficient is the method in revealing ambiguities comparing with already known methods.

1.2 Changes to Thesis Goals

The thesis goals remained unchanged.

2. Research Chapter

2.1 Introduction

This thesis presents a novel metric for evaluating the quality of the software requirements specifications written in Natural Languages (NL). The method represents a high interest, due to the fact that currently around 71 per cent of the requirement specifications are written in natural languages (Umber, Ashfa, & Bajwa, 2011). However, the NL are inherently ambiguous and may lead to different interpretations of the same requirement (Berry & Erik, 2004). The attempts to use semi-formal and formal languages while documenting the requirements has unclear effect on the software requirement specifications (SRS) quality, because ambiguities arise during transformation of SRs from NL to structured or semi-structured languages (Kamsties, Berry, & Paech, 2001). Therefore, the majority of companies continue writing SRS in NL and a good practice in the requirements engineering remains to be performing the validation of the SRS after the requirements are elicited, analysed and documented (Wiegers & Beatty, 2013a). The validation of the SRS is often done by performing reviews and inspections. However, reviewing large specifications is a challenging tasks, which neither gives the answer on the quality of SRS nor guarantees that all flaws of the SRS are revealed (Popescu, Daniel, Spencer, Nenad, & Berry, n.d.); (Saito et al., 2013). As a consequence, there is a need for more reliable techniques allowing to validate SRS.

In accordance with the IEEE Standard 1028, the a high quality SRS does not contain any of the following anomalies: missing, superfluous, ambiguous, inconsistent, not conforming to standards, risk-prone (unstable or requirements with high interdependence), incorrect, not-implementable, editorial. The current research is focused primarily on the identification of ambiguities, because they represent the most undesirable consequences for the project, namely wrongly implemented functionalities, which result in the costs overrun and delay of the product release (Denger, Berry, & Kamsties, n.d.).

We defined a novel method of validation large requirements specification written in NL by the means of surveying the stakeholders and a subsequent calculation of the requirements ambiguity metric. Considering the best practices on metric design by Mund (2013), we proposed the metric, which is easy to understand and apply, and is accompanied with the precise instructions how to apply in practice the method of revealing ambiguities before calculating the metric.

To assess the accuracy of the method description and the efficiency of the proposed method, we conducted 3 experiments on real industrial SRS involving the employees of the company where the experiments were conducted. The results of the experiment showed, that the method is more efficient in revealing ambiguities than the experts assessment and especially efficient in disclosing ambiguities in short and laconic requirements.

2.2 Related work

The requirements validation is an inalienable part of the requirements engineering process (Sharma, Richa, & Biswas, 2012), thus, much research is dedicated to the requirements specifications quality. The majority of the scholars focused on identification of several quality criteria, while the current research concentrated on detecting ambiguities and quantitatively assessing the quality of the SRS. The following researchers are considered as the basis for our work.

Fabbrini et al. (2014) proposed NL SRS quality model, which allows to quantitatively measure the quality of SRS sentences. A prototype of an automated tool that parses NL requirements specifications and applies the quality model was developed. The model represents the analysis of the document to ascertain whether it contains any defects of ambiguity, completeness or understandability. The following ambiguity indicators were considered: implicit subjects sentences, optional phrases, subjective sentences, vague sentences, weak sentences. The rules that allow to recognize the above mentioned indicators are predefined in the lexical analyser component of the tool. The prototype was tested on more than 800 real industrial SRS and disclosed between 49 and 73 per cent of faults.

Although our research also concentrates on revealing ambiguities, we propose to identify defects surveying the potential stakeholders. It gives the following advantages over application of automated tools: our method might be applied to evaluate SRS written in any NL, not only in the language(s), for which the tool have a lexical analyser; all ambiguities revealed by our method are actual ambiguities, while ambiguities identified by automated tool still require reviewing.

Ron S. Kenett et al.(n.d.) suggested several formulas to calculate SRS metrics to measure the document quality quantitatively. The goal of both our research and the research of Ron S. Kenett is to assess the quality of the SRS and identify the parts of the specification that don't meet the quality standards and require reworking. The quality criteria defined in IEEE standards were taken by the authors as the bases for definition what is a good requirement. Kenett et al. identify whether the requirement is ambiguous based on the presence of sentence attributes that define a syntactically correct sentence written in English. Among other metrics, the authors propose the Accuracy metric, which is based on the fact how frequent the ambiguous information and “to be determined” (TBD) notes are disclosed in the document. To asses the overall quality of the document, the calculated metrics can be compared against the historical data (Kenett, n.d.)

A quality assessment method based on the document characteristics, namely document content and structure, was proposed by Thitisathienkul et al.(2014). The method is applicable for SRS written in NL. The authors introduced the measurement process model, which aims at guiding users in applying the method in practice. The measurement process model is a simple and powerful means of evaluating the quality of SRS. In the first step, the users should identify the scope of requirements for measurement and assign responsibilities. Secondly, they should decide what characteristics of SRS should be considered as good or bad characteristics of SRS and assess whether they can be measured. Further, the metrics for quality measurement should be defined and applied. After that the results of metrics application should be interpreted, the possible improvements of the reviewed SRS should be proposed.

In our research, we followed the fundamental principles of metric definition, proposed by Thitisathienkul et al. In particular, we established the rules of choosing the scope of requirements for the metric creation (in our case - requirements sampling strategy). Than, we identified the way of data collection (by surveying the stakeholders) and people responsible (in our method the survey design is performed by the experts). Thereafter, we described the measurement process and evaluation (the unambiguity metric). Consequently, the results of an independent employment of our method on various SRS can be compared (Thitisathienkul, Patra, & Nakornthip, 2014).

2.3 Research Question

The following non-trivial aspect of the method definition were in focus:

- What rules and principles of creating alternative interpretations of requirements are the most beneficial for revealing ambiguities?
- Who must be included into the group of stakeholders, whose task is to answer surveys?
- What participants must be included into the group of experts, whose task is to design survey questions?

Additionally, we investigated how efficient is the method of validating large requirements using surveys in comparison with other methods.

2.4 Research Approach

2.4.1 Method description

The development of the method description was the first stage of our research. A draft was made based on the the artifacts of the pilot method application, which took place at Siemens AG in 2013. The description consisted of three focus areas: how to sample requirements from the requirement specification, how to create a survey based on the sampled requirements and how to choose the stakeholders who will answer the survey. To clarify the details about how the pilot survey was organized, we arranged a workshop at Siemens AG. Among the workshop participants were a Project Manager who was the method inventor and a Requirements Engineer who participated in the initial method application. Further development of the method description was done iteratively based on the analysis of the tree runs of method application and on the observations of the participants. Refer to Appendix A to see our research timeline.

2.4.1.1. Requirements sampling

During the research, the question about a right sampling strategy was raised several times. The initial sampling strategy was based on the statistic approaches to population sampling, namely on the random sampling. During the workshop, it was decided that if a SRS has both functional and nonfunctional requirements, both types must be sampled for the survey. Additionally, from the experience of the industry partner, different chapters of the SRS could be written by different requirements engineers under different working conditions (e.g. close to the development phase under the time pressure or during the normal workflow) and be of a different quality. Thus, to validate quality of the SRS, the requirements from different parts of the SRS must be included into the survey. The initial sampling strategy transferred to the one comparable to the stratified random sampling strategy, where the stratas are SRS chapters, functional and nonfunctional requirements.

Besides, the decision of how many questions should be included into the questionnaire was changed. An initial strategy to use the Acceptable Quality Limit or AQL-Method that is suppose to minimize the survey measurement error (Charles, 1977) proved to be inefficient to apply in practice for sampling requirements. If the SRS consisted of e.g. 20 or 30 requirements, the questionnaire should have been consisted of 3 to 5 questions, what could be not enough even for sampling requirements from all parts of SRS. On the other hand, to validate SRS consisting of around a hundred requirements approx 20-30 requirements must be sampled. Due to the fact that the sample size has direct effect on the cost of a survey. (Arnab, Raghunath, Zewotir, & North, 2014), including such number of requirements was undesirable. Additionally, a long questionnaire represents a core concern because the studies revealed its negative effect on the collected data quality (Adigüzel, Feray, & Michel, 2008). Therefore, the number of the requirements on the questionnaire was fixed based on the practical experience of the industry partner.

2.4.1.2. Stakeholders sampling

The principles of stakeholders selection were embraced from the pilot method application. Sampling survey respondents, we primarily focused on including the stakeholders with various positions, namely software engineer, software architect, usability designer, quality assurance engineer, tester. project manager, technical writer, product manager. The various stakeholders have different tasks and responsibilities, thus, examine the requirements from different angles. Several positions were decided to be optional, because not every project has a representative of such position. Due to a relatively small number of participants, every stakeholder role was included only once or twice. The stakeholders' experience was not taken into consideration because previous studies showed a low interdependence between the employee qualification level and their ability of interpreting requirements (Saito et al., 2013). Nevertheless, such criteria as relation to the project was introduced while sampling the stakeholders. We included both, stakeholders of the native project, for which the SRS was studied, and the stakeholders of external projects to observe how this criteria influence the correct requirements interpretation.

2.4.1.3. Questionnaire design

The questionnaire creation strategy was adopted from the pilot method application. According to it, the experts should brainstorm all possible interpretations of the requirement, the correct interpretation is provided by the author of requirements. However, to guide the experts during the questionnaire design and to impart the structure to the process, we additionally specified five types of interpretations that can be generated around a requirement.

Although we used close-end questions, which are considered more advantageous over open-end questions, because they provide more uniform responses that are easy to process (Banda, 2008), we additionally introduced a "None of the above answers" option. The answer aimed at avoiding a non-response error that could arise when a stakeholder has his/her own interpretation that is not expressed by the experts, which created the questionnaire, and avoids choosing any answer from the proposed list (Phung et al., 2015). From the practical experience of the industry partner, it was decided that every requirement should have three to five interpretations (including "None of the above answers").

2.4.1.4. Metric calculation

Developing the metric, we relied on the fact that a requirement can be considered as unambiguous only if all stakeholders share the same understanding of its meaning (Berry, n.d.). However, we concede that even a high quality requirement can be misinterpreted by some stakeholder(s) because of objective reasons (e.g. external stakeholder might not have some key for the understanding knowledge). Thus, we regard that 90 to 100 per cent of common interpretations is acceptable and indicates an excellent quality requirement. The coefficient 1,0 was assigned to such requirements.

Other coefficients were assigned based on the requirements classification taken from the pilot project and the idea that the value of a requirement drops exponentially with the reduction of the number of common interpretations. Requirements which have from 80 to 89,9 per cent of common interpretations are considered as relatively good quality requirements, to which the coefficient 0,8 was assigned. If 60 to 79,9 per cent of stakeholders interpreted a requirement in the same way, such requirement has only partially common interpretation, coefficient 0,6 is assigned to the requirement. Less than 60 per cent of identical interpretations indicate a requirement with various interpretations, to which a coefficient 0,3 is assigned

As a SRS is a collection of the requirements, in order to calculate the metric, we sum the number of requirements with the adjustment on the above mentioned coefficients. Metric formula looks as following:

$$Unambiguity\ Metric = \frac{K1 * N1 + K2 * N2 + K3 * N3 + K4 * N4}{N1 + N2 + N3 + N4} * 100\%$$

where coefficients K1..K4 are equal to 1, 0.8, 0.6 and 0.3, variables N1..N4 indicate the number of requirements, assigned to the coefficients K1..K4 respectively.

2.4.2 First survey

Having an initial draft of the method, the first run of the method application was conducted at one of the projects by Siemens AG. A requirement specification and a feature specification served as the basis for the survey questions, which were created during the workshop. The experts who participated in the workshop and whose task was to devise the alternative interpretations for the selected requirements were a project Software Architect and a Requirements Engineer. At the beginning of the workshop, the goal and all main principles of the method were recapped for the participants. Thereafter, during a brainstorming session, three to four alternative interpretations for each of the sampled requirements were conceived. After the workshop, the questionnaires were composed and sent out to ten survey participants, namely to two external and eight internal ones.

When the survey results were obtained, we analysed the percentage of correct answers and wrong answers separately among internal and external stakeholders, as well as the percentage of correctly interpreted requirements for every participant. Additionally, we calculated a standard deviation and a standard variation to determine whether the number of correct interpretations provided by a stakeholder standed out against other ones.

To evaluate the outcome of method application, the first and subsequent survey findings were compared with the quantitative linguistic analysis of the requirements. We quantitatively estimated the ratio of ambiguity indicators in the requirements and analysed how the survey findings correlate with the data. To this effect, we developed a Java console application to parse and perform a lexical analysis of the sampled requirements. The methodological base of the quantitative linguistic analysis was the literature on the assessment of the requirements specifications, written in natural languages (Fabbrini, Fusani, Gnesi, & Lami, n.d.). See figure 1 with the list of lexical ambiguity indicators, provided by (Fabbrini et al., n.d.) et al. (n.d.). In addition to the linguistic analysis of the requirements sampled for the first questionnaire, we analysed the detailed feature specifications written based on the requirements.

Ambiguous sentence type	Ambiguity indicators
Implicit subject sentences	this, these, that, those, it, they, above, below, previous, next, following, last, first
Optional sentence	possibly, eventually, if appropriate, if needed
Subjective sentences	having in mind, taking into account, taking into consideration, similar, better, similarly, worse, as much as possible, as good as possible , as quick as possible
Vague sentences	clear, well, easy, strong, weak, good, bad, efficient, efficiently, low, useful, significant, adequate, fast, slow, old, new, future, recent, past, today's, near, far, close, back, in front
Weak sentence	can, could, may, might

Table 1. Ambiguity indicators of (Fabbrini et al., n.d.) et al. (n.d.).

Along with the answers to the questionnaire, we received the feedback regarding the comprehensibility of the requirement interpretations from several stakeholders. The observation of the workshop along with the analysis of the first survey results and the participants' feedback allowed us to improve the way of designing the alternative interpretations and apply it during the second and the third runs more efficiently as well as to amplify the method description.

In particular, it was anew realised that the alternative interpretations must be formulated clear and understandable, should not mislead the stakeholders. Additionally, we noticed that the experts experience difficulties while creating the alternative interpretations of the requirements first time. Thus, it was proposed to find a weak point of the requirement, namely an equivocal collocation or sentence, which become be a source of several interpretations.

2.4.3 Second survey

Initially, we had planned that the second and the third runs of the surveys would be conducted sequentially. However, due to the time restrictions the second and third survey were conducted simultaneously. For the second survey, we used the epics of user stories of a currently running project. The experts group embraced a Project Lead and a Requirements Engineer. Within the scope of the workshop, the alternative interpretations for the requirements were created following the method guidance.

Additionally, it was decided to ask the experts their opinion on how unambiguous was every requirement on the scale of 5 points, where 1 meant "not clear at all, ambiguous", and 5 meant "completely clear". The purpose of the collection of opinions was to use this information during the analysis of the survey results, in order to comprehend whether the expert assessment on the requirement unambiguity correlates with the actual ambiguity revealed by the method.

2.4.4 Third survey

The third project was based on the obsolete or pending requirements for a currently running project, the participants included a Requirements Engineer and a Principal Key Expert. During the workshop, there was a misunderstanding of the goals and approach of the survey creation among the participants, since one of the experts misinterpreted the goal of the method. As a result, the workshop was interrupted several times to clarify the objectives and rules of how to create the alternative interpretations.

The novelty of the third survey was not to use the "None of the above answers" as a correct answer to analyse whether it has any effect on how many participants will chose this answer.

2.5 Used Data Sources

For creating the method description, both primary and secondary empirical data sources were used (Hox & Boeije, 2005), which represent mostly the qualitative data sources consisting of our observations during three runs of surveys and artifacts of the pilot project (Taylor-Powell; (Hox & Boeije, 2005); Renner, 2003). A literature review was done to adopt the established best practices practices of designing surveys and sampling the stakeholders. Additionally, we used the feedback of the stakeholders, who responded to the survey questions, to improve the survey designing strategy.

To design the surveys, requirement specifications of various types were used. The first project was based on the excerpt of the software requirements specification, composed of twenty requirements, and a corresponding user story specification. Ten requirements were sampled out to create the questionnaire. For the second survey, user stories specification of sixteen user stories was used, ten of them became a basis for the questionnaire. To create the third questionnaire, we employed the excerpt of the software requirements specification that was formed of twenty three requirements, among which the selection of ten was chosen.

2.6 Research Results

The outcome of this research consists of two parts. The first one is the final description of the method. The second is the method validation through the analysis of the data gathered in the applications of the method in three different projects.

2.6.1. Method description

2.6.1.1. Requirements sampling

After trying to apply various sophisticated methods of deciding the sample size during the first survey run, the questionnaire length was fixed to ten questions following the Siemens AG experience. The arrangement has both practical relevance and the scientific background. According to the industry partner experience, the employees are unwillingly engaged in any studies or activities, that require more than 30-40 min of participation. Thus, the decision was made to reduce the probability of the non-response errors as well as to ensure the high-quality of collected data (Adigüzel et al., 2008).

For the requirements sampling, the application of a simple random sampling after dividing requirements on subgroups was considered to be the best option. The criteria for the division are functional and nonfunctional requirements, SRS chapters, authors of the requirements.

2.6.1.2. Stakeholders sampling

The stakeholders of four mandatory occupations (software engineer, architect, usability designer, product manager) and four optional occupations (project manager, quality assurance engineer, technical writer, tester) were selected as respondents.

No dependency was spotted between the stakeholders with the number of correct replies and their occupations. In the first survey, the integration tester achieved the highest number of correctly interpreted requirements, whereas the lowest number was also gained by the testers. (see Table 2). In the second survey, the highest ratio of the correctly interpreted requirements belongs to the project manager (see Table 3), at the same time, technical write and quality assurance engineer have interpreted the requirements in the most correct way in the third survey (see Table 4). Furthermore, the standard deviation on the average number of the correctly interpreted requirements was small in the first and third surveys and quite small in the second survey, which indicated low difference between the representatives of different occupations

#	Participant	Occupation	Relation to the project	# of correct answers	Average	Standard deviation	Standard variation
1	TK	Software engineer	internal	4	4,57	1,49	0,32
2	WP	Usability designer	internal	6			
3	KR	Project Manager	internal	5			
4	BD	Technical writer	internal	5			
5	HS	Integration tester	internal	7			
6	RT	Tester	internal	3			
7	PA	Tester	internal	2			
8	WS	Architect	external	6	5,00		
9	SL	Quality assurance engineer	external	5			
10	TM	Product Manager	external	4			
11	HO	Product Manager	internal	n/a	n/a	n/a	n/a
12	HT	Tester	internal	n/a			
13	WS	Architect	external	n/a			

Table 2. Stakeholders of the first survey.

The fact that no correlation was revealed between the stakeholders occupation and the number of correctly interpreted requirements yet again confirms the necessity to engage the representatives of as many occupations as possible for obtaining the most objective analysis of ambiguities, hidden in the SRS.

.

#	Participant	Occupation	Relation to the project	# of correct answers	Average	Standard deviation	Standard variation
1	KS	Software engineer	internal	5	5,50	2,32	0,56
2	SE	Product Manager	internal	7			
3	BG	Tester	internal	6			
4	LM	Integration tester	internal	4			
5	FM	Usability designer	external	2	1,50		
6	LS	Quality assurance engineer	external	1			
7	MG	Project Manager	external	n/a	n/a	n/a	n/a
8	GD	Product Manager	internal	n/a			
9	BD	Technical writer	internal	n/a			
10	KP	Architect	internal	n/a			
11	HO	Product Manager	external	n/a			

Table 3. Stakeholders of the second survey.

According to our research, in two of three surveys, there was almost no difference in the rates of how internal and external stakeholders recognize the correct requirements. However, the second survey resulted in a significantly lower ability of correctly interpreted requirements by the external stakeholders (see Table 3). Therefore, we would recommend further investigation on the objective ability of internal and external stakeholders to reveal ambiguous in requirements.

#	Participant	Occupation	Relation to the project	# of correct answers	Average	Standard deviation	Standard variation
1	AC	Software engineer	internal	5	4,88	1,54	0,30
2	WP	Usability designer	internal	4			
3	MA2	Project Manager	internal	5			
4	HD	Product Manager	internal	6			
5	BV	Technical writer	internal	7			
6	BT	Tester	internal	3			
7	FW	Architect	internal	6			
8	SA	Tester	external	3	5		
9	ZJ	Quality assurance engineer	external	7			
10	MA	Software engineer	internal	n/a	n/a	n/a	n/a

Table 4. Stakeholders of the third survey.

2.6.1.3. Survey design

Our research showed that the design of the survey is the most complicated part of the method application and the success of the requirements validation strongly depends on the quality of the questionnaire, namely of the created alternative requirement interpretations.

During the workshops on questionnaire creation, we observed experts having difficulties in sharing common understanding on how the interpretations should look. Although the strategy of finding equivocal collocations proved its strength and the proposed types of created interpretations eased the questionnaire design to some extent, further research on systematisation of alternative requirements interpretations is needed.

Moreover, after the first survey, we received the feedback from two stakeholders, who stated that they revealed ambiguities and deceptions in the proposed alternative interpretations, which hampered answering the survey. After the second and the third surveys, we did not receive any unfavorable feedback, which can be viewed as a positive sign in increasing the questionnaire quality. However, further research on facilitating the design of alternative requirement interpretations and improving characteristics of the questionnaire is recommended.

2.6.1.4. Experts for the questionnaire design

Although the choice of the experts who design the questionnaire was never the main focus of our research, observing three runs of the method application we could conclude, that the experts play a key role in the survey success or failure, because their main responsibility is to create the questionnaire for the survey. Therefore, the experts should have not only proven qualification and experience in the software engineering, but also be acknowledged of the method objectives, rules of the design of alternative requirement interpretations.

2.6.2. Method efficiency

2.6.2.1. First survey

According to the first survey results, two of ten requirements were interpreted by 80 per cent of stakeholders commonly (requirements #2 and #8). For the requirement #2, around 2,5 per cent of lexical ambiguity indicators were revealed, which is lower than the average rate, the requirement #8 did not contain any lexical ambiguity indicators. Seven of ten requirements had less than 60 per cent of common interpretations. For most of them, the number of lexical ambiguity indicators was around an average value or greater than average (see Table 5). In some cases, when no lexical ambiguity indicators were spotted but the percent of correct interpretations was low (e.g. requirement #9), the requirement consisted only of one short sentence, which explained both the absence of ambiguity indicators in the requirement and its high ambiguity.

Thus, we can assume that the method of validation of the large requirements specifications is more efficient in revealing ambiguities in short and laconic requirements than the lexical analysis of the ambiguity indicators.

Question #	Stakeholders										% of correct interpretations	% of lexical ambiguity indicators
	Internal							External				
	TK	WP	KR	BD	HS	RT	PA	WS	SL	TM		
1	1	3	1	1	3	1	1	1	3	3	40,0	0,0
2	1	1	1	1	1	3	1	1	2	1	80,0	2,5
3	1	4	3	1	1	1	1	1	1	4	70,0	0,0
4	1	2	2	3	2	1	1	2	2	3	50,0	6,3
5	2	1	1	3	4	3	2	1	1	1	20,0	2,0
6	1	4	2	2	2	3	3	4	4	1	30,0	7,9
7	3	3	1	3	4	3	1	4	4	3	30,0	2,9
8	1	3	3	3	3	3	2	3	3	3	80,0	0,0
9	3	3	1	4	2	4	1	1	4	3	30,0	0,0
10	2	2	2	4	3	4	1	2	3	3	40,0	5,0
# of correct interpret.	4	6	5	5	7	3	2	6	5	4	Average	2,66
Average	4,57							5,00			Correlation coefficient	-0,298
Standard deviation	1,494434118											
Standard variation	0,32%											

2 - ordinal number of the correct interpretation

Table 5. First survey results

The correlation coefficient between the percentage of correctly interpreted requirements and detected lexical ambiguity indicators is equal to -0,3, which indicates a weak negative relationship between two variables (see Figure 1).

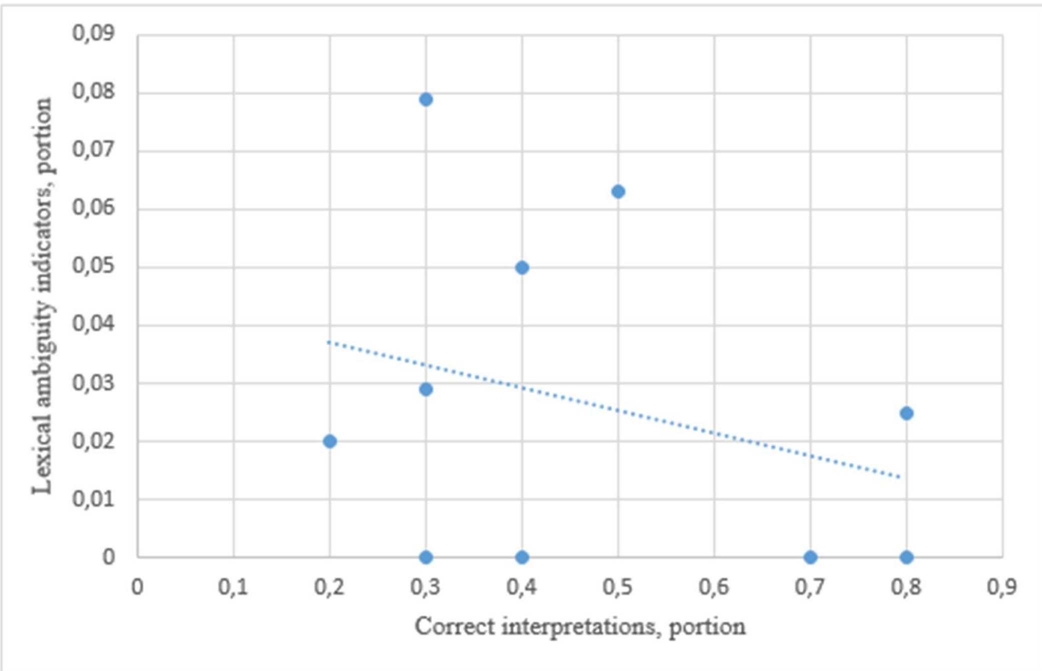

Figure 1. First survey: correlation between correct interpretations and lexical ambiguity indicators,

The ambiguity metric for the first SRS is equal to 43 per cent and calculated as for the first and subsequent SRSs as follows:

$$Unambiguity\ Metric = \frac{K1 * N1 + K2 * N2 + K3 * N3 + K4 * N4}{N1 + N2 + N3 + N4} * 100\%$$

$$= \frac{1 * 0 + 0{,}8 * 2 + 0{,}6 * 1 + 0{,}3 * 7}{2 + 1 + 7} = 43\ \%$$

2.6.2.2. Second survey

The second survey is characterized by a high non-response rate, namely 45 per cent of respondents did not submit their responses (see Table 3), One of possible explanations of a low response rate could be that the survey was conducted under the time pressure during the holiday season.

Question #	Stakeholder						% of correct interpretations	Non-ambiguity (expert assessment)	% of lexical ambiguity indicators
	Internal				External				
	KS	SE	BG	LM	FM	LS			
1	3	3	3	2	4	1	50,0	2,5	6,7
2	4	2	2	2	4	2	33,3	3,0	3,8
3	4	2	2	3	4	3	33,3	3,5	0,0
4	4	4	3	3	4	3	50,0	4,0	3,4
5	1	3	1	1	4	3	50,0	5,0	4,7
6	2	3	3	3	4	4	50,0	5,0	2,0
7	4	2	3	2	4	2	50,0	5,0	0,0
8	3	2	2	2	4	1	50,0	5,0	2,2
9	4	3	3	4	4	4	33,3	4,0	0,0
10	3	1	1	1	1	1	16,7	5,0	5,4
# of correct interpret.	5	7	6	4	2	1	Average	4,200	2,821
Average	5,5				1,5		Correlation coefficient	0,083	-0,031
Standard deviation	2,316606714								
Standard variation	0,56%								

2 - ordinal number of the correct interpretation

Table 6. Second survey results

The number of correctly interpreted requirements by external stakeholders was quite low, namely 1,5 answers on average. Furthermore, “None of the above answers” was chosen nine of ten times by an external stakeholder (see Table 6). The answers to the requirements .#3, #6, #7 and #9 have also contained an above average number of “None of the above answers”, which may signify either low quality response data or misleading questionnaire.

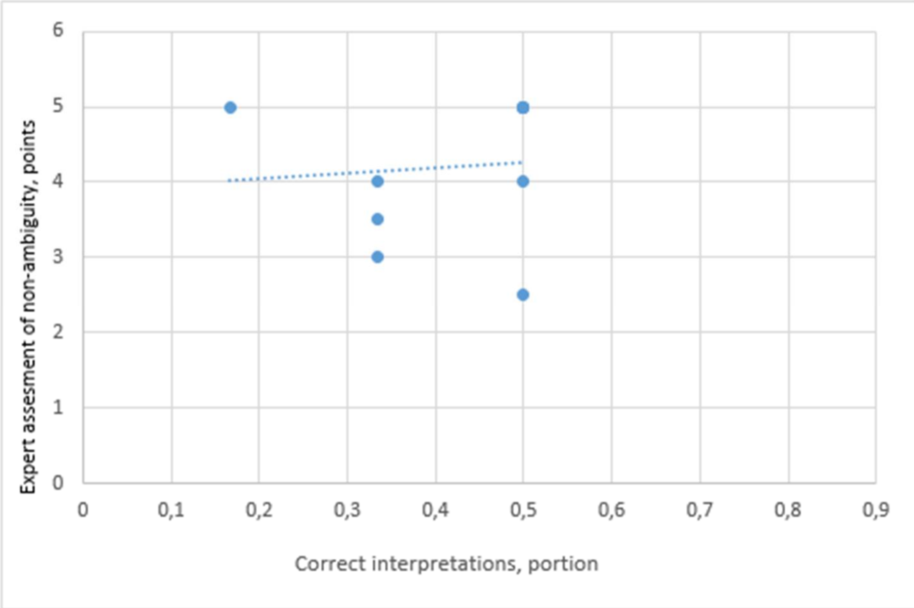

Figure 2. Second survey: correlation between correct interpretations and experts' assessment of requirements ambiguity

The correlation coefficients between the number of correct interpretations and the expert assessment of requirements' unambiguity is equal to 0,08 and between the number of correct interpretations and the number of detected lexical ambiguity indicators is 0,03, which basically indicate no relationship between all values (see Figure 2, Figure 3).

Overall, six of ten requirements achieved 50 per cent of common interpretations, the result of the rest four requirements is lower. The ambiguity metric for the second SRS is equal to 30 per cent.

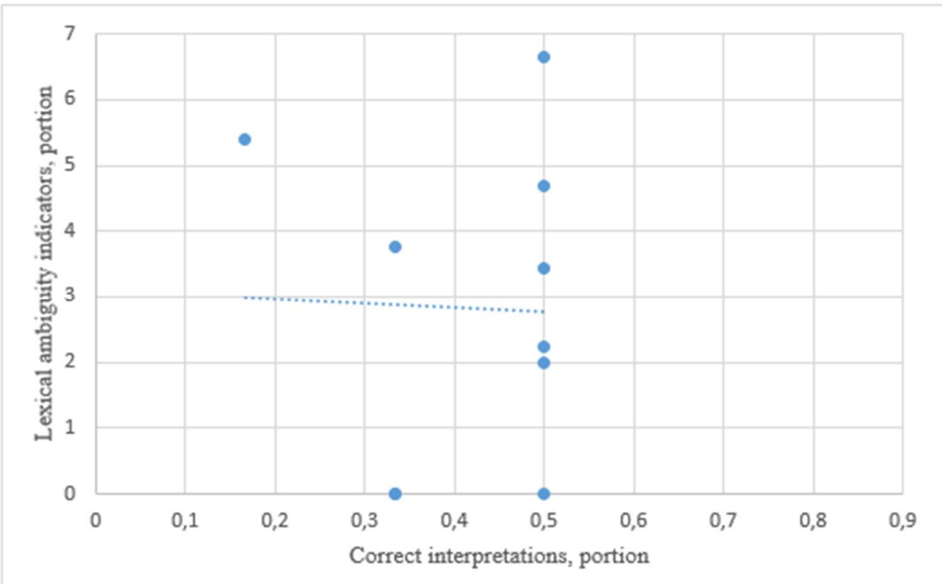

Figure 3. Second survey: correlation between correct interpretations and lexical ambiguity indicators

2.6.2.3. Third survey

The results of the third survey showed that one of ten requirements were identically interpreted by all stakeholders (requirement #8), four requirements were commonly interpreted by 60 to 80 per cent of respondents (requirements #2, #4, #6 and #10), five requirements collected less that 60 per cent of correct answers. On average, the requirements were correctly interpreted by a comparable number of internal and external stakeholders (see Table 7).

Question #	Stakeholders									% of correct interpretations	Non-ambiguity (expert assessment)	% of lexical ambiguity indicators
	Internal							External				
	AC	WP	MA2	HD	BV	BT	FW	SA	ZJ			
1	4	1	3	1	2	3	1	1	4	11,11	2	7,14
2	1	1	3	1	2	1	1	2	1	66,67	4,5	2,43
3	4	3	2	2	3	3	3	1	3	22,22	5	5,17
4	1	3	1	1	1	3	1	1	1	77,78	5	1,96
5	3	3	3	3	2	1	3	1	2	22,22	4	0,00
6	2	2	2	1	2	3	2	3	2	66,67	5	2,66
7	3	3	3	3	3	2	3	1	2	11,11	5	0,00
8	1	1	1	1	1	1	1	1	1	100,00	5	0,00
9	3	2	1	1	1	2	1	2	1	55,56	5	2,50
10	1	1	2	1	1	1	1	2	1	77,78	5	5,17
# of correct interpret	5	4	5	6	7	3	6	3	7	Average	4,550	2,703
Average	5,14							5,00		Correlation coefficient	0,513	-0,111
Standard deviation	1,536590743											
Standard variation	0,30%											

2 - ordinal number of the correct interpretation

Table 7. Third survey results

The correlation coefficient between the percentage of correct interpretations and experts assessment on unambiguity is equal to 0.51, which indicates a moderate positive relationship between two evaluations. One requirement that stands out of the correlation is requirement #7, which was correctly interpreted only by two stakeholders, and which though was assessed by the experts with 5 points as a clear requirement (see Figure 4). The ambiguity metric for the third SRS amounted to 49 per cent.

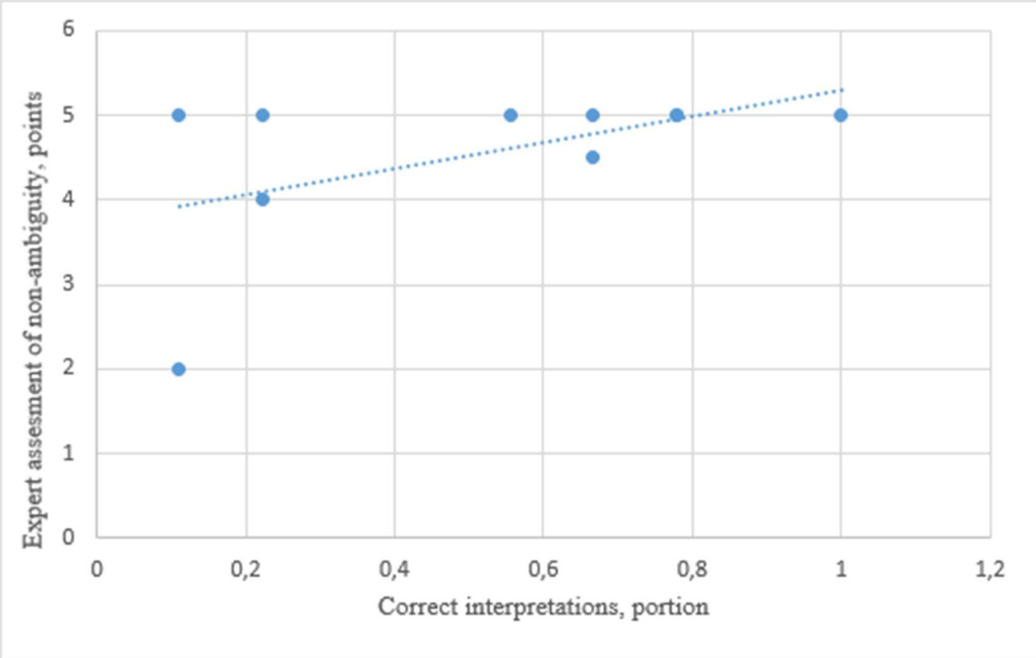

Figure 4. Third survey: correlation between correct interpretations and experts' assessment of requirements ambiguity

The correlation coefficient between the number of correct interpretations and the number of detected lexical ambiguity indicators is equal to -0.11, which indicates a very weak negative relationship (see Figure 5). There were less than average or zero lexical ambiguity indicators detected in the most of the requirements, which were commonly interpreted by the stakeholders; in contrast, more than average lexical ambiguity indicators were found in the requirements, which the stakeholders interpreted differently. However, several exceptions occurred, for example requirements #5 and #7, where the ambiguity was only disclosed by the method of validating large requirements specifications.

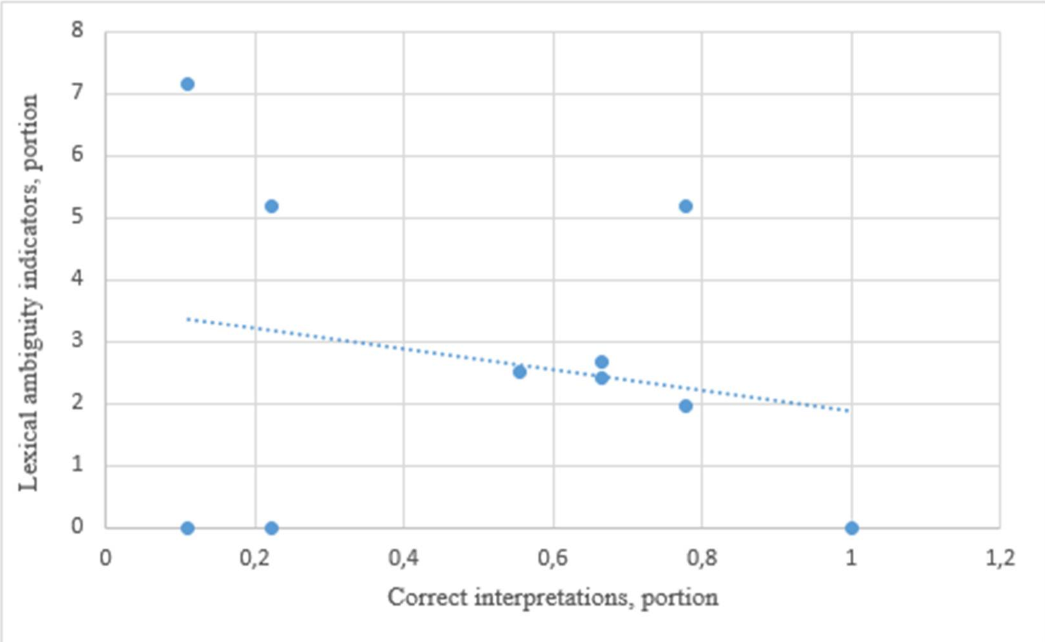

Figure 5. Third survey: correlation between correct interpretations and lexical ambiguity indicators

In accordance with the questionnaire design, the "None of the above answers" option was never a correct answer in the third survey. Confirming to the survey results, the answer was chosen 22 per cent of times on average, which is though slightly lower than the average probability of choosing any answer in the survey, but still comparable.

2.7 Results Discussion

2.7.1. Limitations

In our three examples, the requirements specifications under study were medium to small in size, consisting only of functional requirements. Thus, there was no possibility to fully apply the sampling strategy prescribed by the method, and sampling strategy was close to the simple random sampling.

Additionally, due to the limited accessibility of the experts, the expert groups of the second and the third runs of the surveys included only two persons, one of them was the authors of the requirement specifications. For this reason, the expert assessment on the quality of the requirements might have a portion of subjectivity, the created alternative interpretations might be not as divers as if

For evaluation of the results, we used a simple Java parser, which was developed based on revealed information about the quality model of F. Fabbrini et al (n.d.). Since all quality indicators could be not uncovered by the authors in the article, the linguistic analysis performed by the prototype of software tool developed by F. Fabbrini et al. could vary from our analysis.

The design of the survey is one of the most difficult and important parts of the method. Although we initially aimed at describing the method in a way allowing to exclude "human factor" so that it can be applied in any organisation, the three runs of the survey revealed that the outcome of the method application depends heavily on the quality of the questionnaire. Consequently, the expertise, preparation to the workshop and qualification of experts, who create the survey play crucial role. Thus, further research on the designing survey questions is recommended.

2.7.2. Results validation

The validation of the method description was through the three runs of the surveys. We observed how clear for the participants were the objectives of the method, the rules of its application and the results received. If the experts, who were responsible for the surveys creation, or the stakeholders, who answered the surveys questions, were confused during the method application, the description of the method was enhanced and a special attention was paid to this aspect during the next run.

The validation of the method results was by the means of two techniques: linguistic analysis and the experts' assessment. During the linguistic analysis, we compared the percentage of the ambiguity indicators found by a Java application developed based on the NL SRS Quality Model proposed by F. Fabbrini et al (2014) with the percentage of the correctly interpreted by the stakeholders requirements. Additionally, we asked the experts to assess how clear from their point of view each requirement was. The experts' opinion was also compared with the surveys results and analysed.

2.8 Conclusion

In the current thesis we developed the description of a novel method of validating large requirements specifications using surveys. The method was initially described based on the pilot application of the method and the best practices derived from the literature. Further, we applied the method on the running projects by Siemens AG three times, improving the method description after each iteration based on our observations of the participants and on the feedback of the stakeholders. We showed that the stakeholders of different occupations are equally good in revealing ambiguities in the SRS, thus, the representatives as many positions as possible respond the survey questions. We also ascertained that the quality of the questionnaire and the clarity of alternative interpretations are at high importance for revealing actual ambiguities in the requirements. Thus, the qualification and the engagement of the experts who create the questionnaire play key role while applying the method.

In addition, we compared the results obtained from the method applications with the linguistic analysis on ambiguity indicators and with the expert assessment of the requirements quality. We suppose that the method is more efficient in revealing ambiguities in short and laconic requirements, however, additional research on the question is advised.

Although further research on facilitation the design of the questionnaire is required, we are convinced that the method can be successfully applied for revealing abilities in large requirements specification at the industry

3. Elaboration Chapter

3.1. Statistical survey

Statistical survey is a powerful means used to collect statistical information about economic, political, social and cultural aspect of the life (Bethlehem, 2009; Calinescu, Melania, Sandjai, & Barry, 2013) Commonly, the information is collected by distributing a homogeneous questionnaire, which contains the questions on the theme under the study, to the respondents. (Bethlehem, 2009; Banda, 2008).

Conducting a survey is a sophisticated process, that involves complex decision making and requires careful preparation (Bethlehem, 2009; Banda, 2008). Different researchers distinguish various stages of the survey lifecycle, however, the majority agree on the following ones:

Survey design-> Data collection -> Data editing -> Nonresponse correction -> Analysis -> Publication

Subsequent chapters don't contain a detailed description of the above mentioned stages. Instead, they reveal the pitfalls and describe the most challenging aspects of organizing and running the surveys.

3.1.1. Questionnaire design

The success of the survey and the accuracy of the collected data is heavily dependent on the quality of the questionnaire. The well-designed questionnaire ensures that the collected data is actual and relevant, and that the research goals are met on time and budget (Doutriaux & Crener, 1982). Therefore, the survey preparation phase requires special attention and carefulness (Banda, 2008). However, before starting to create the questionnaire, it is recommended to clarify the survey objectives. They must include the clear goals of what and how many respondents are to be surveyed, how the collected data is to be analysed and used (Banda, 2008; Doutriaux & Crener, 1982). Also the questionnaire length must be under careful consideration, because the dependence between the quality of the collected data and the questionnaire size has been proven (Adigüzel et al., 2008).

Moreover, the questionnaire format and the type of questions, which will be used in the questionnaire, must be agreed on, because it has an impact on the way the collected data is processed and analyzed, as well as the quality of the findings (Moy & Murphy, 2016). Two the most common used question types are open-ended questions and close-end questions (Banda, 2008; Moy & Murphy, 2016). The advantages of open-ended questions are that they allow respondents express their ideas and thoughts to the topic on their own words, on the contrary, allow respondents to answer in their own words which makes the response creditable. However, the processing of the responses requires complex coding, which entails high costs and time expenses (Moy & Murphy, 2016). In contrast, the uniform close-ended questions gether unified responses that are easy to process and analyses. Nevertheless, a significant disadvantage of close-ended questions are that the person who designs the survey has to foresee all possible responses to the questions. In case he/she fails, valuable information may be lost (Banda, 2008).

Additionally, the way the questions are phrased influences how actively and willingly the respondents answer the questions, therefore, the wording must be clear and easy to follow (Shaughnessy, Zechmeister & Jeanne, 2011). The questionnaire layout must be clean, so that the respondents can easily read the questions (Banda, 2008).

To ensure the quality of the composed questionnaire, it is recommended either to run the pilot survey or to redirect the questionnaire to the experts, who can determine whether indeed it is clear, does not contain any sources of misinterpretations and provides all possible answers in case the close-ended questions are used (Moy & Murphy, 2016).

3.1.2. Quality of surveys and survey errors

There are several types of widely accepted survey errors: sampling errors and non-sampling errors, which in turn are divided into coverage errors, non-response errors, and measurement errors.(Phung, Hardeweg, Praneetvatakul, & Waibel, 2015) All of them can distort the collected data and jeopardize the legitimacy of the survey (Moy & Murphy, 2016). Thus, a careful study of the errors and reasons why they appear is recommended in order to avoid their occurrence.

3.1.2.1. Sampling errors

Sampling error represents the difference between the survey statistics (conducted on the sampled respondents) and the actual statistics that could have been collected in case the large population would have been surveyed. The sampling error is occurs in the research because the sample chosen for the survey does not display the population under study accurately (Assael, Henry, & John, 1982). The decision on the sample size and on the respondents included is a crucial task for many researchers (Barlett et al, 2001). The sampling error can normally be administered while opting for an appropriate methodology, sampling design and sample size as well as assuring the presence of different groups including small subgroups in the sample (Phung et al., 2015).

3.1.2.2. Non-sampling errors

Although non-sampling are more often perceived as an inalienable issue of conducting surveys in the developing countries, they also appear in the developed countries and there was a few studies were conducted about their causes and consequences (Phung et al., 2015). Nonsampling errors, which are more sophisticated than sampling errors and more difficult to control, consist of two factors. Firstly, non-response error, take place when sampled respondents do not provide answers to the survey, which entails that the responses do not represent objectively the selected sample. Secondly, the response error happens when sampled respondents misreport the answers or respond inaccurately (Assael et al., 1982). It may be done in purpose, e.g. while answering sensitive questions (Biemer, 2010) or because the respondents are fatigued, are influenced by interviewers or other external factors (Assael et al., 1982). Understanding of non-sampling errors and their causes is vital because they may lead to erroneous conclusions made after analysing the data. (Phung et al., 2015)

The following types of non-sampling errors are known:

1. Measurement error is one of the most commonly met and damaging error types, and represents the difference between the collected statistical data and the actual data. In

particular, it reflects how exactly coincide the respondents' actual opinion or behavior with the one expressed in the survey (Thomas, 2014). The error arises due to the fact that the respondents may intentionally or accidentally provide inaccurate answer to the question or an interviewer can mislead the respondent and cause the error (Biemer, 2010). However, one of the most common reasons of the measurement errors is the poor quality of the questionnaire (Biemer, 2010).

2. Coverage error takes place when the sampled respondents are not representative and don't display the population under study accurately. A coverage error occurs while designing the survey and deciding on the respondents and is represented by two types of errors: over-coverage, which arise when the not-target respondents are embraced in the selected sample, and under-coverage, when the sample does not contain crucial for the research respondents (Phung et al., 2015).

3. Non-response errors arise when the researchers fail to gain the expected information from respondents, may happen because of the lack of access to the respondent or because of his/her unwillingness to respond (Calinescu et al., 2013; Phung et al., 2015). The non-response errors result in inability to calculate and derive the correct conclusion about the subject under study (Calinescu et al., 2013). One distinguishes two types of non-response errors: unit non-response and item non-response errors. The unit non-response error occurs when the data from a respondent is missing completely, whereas the item non-response error takes place in case when the data about sample unit (respondent) is collected partially (Phung et al., 2015). Questions of sensitive topics are often subjected to non-response errors. Sensitive topic questions are the ones to which most people usually feel "uneasy" to provide the response if they are asked (Albaum, Gerald, Roster, & Smith, 2014).

4. Data-processing error refers to the mistakes and inaccuracies which occur during data entry, editing, coding, formatting questionnaires, assignment of survey weights (Biemer, 2010).

3.1.3 Data analysis

3.1.3.1.Factorial analysis

Factor analysis is widely recognized means in the the social sciences for determining the structure of initial data and revealing the factors that influence it (Doutriaux & Crener, 1982; Norman & Streiner, 2003). Examples of factor variables are income level of two cities, or age of the population, occupation of the respondents. The factorial analysis allows to define various hypotheses (e.g. the influence of the respondents occupation on their income level) and confirm or refute the correlation of factors defined in the hypothesis (Norman & Streiner, 2003). Factors consistently regard to observable phenomena, however, they can represent hypothetical constructs, such as intelligence, depression, or coping ability (Norman & Streiner, 2003). Usually, the factorial analysis compares and analyses two or more factors to identify determine statistical relevance and importance of the factors and their cross-influence (Taherdoost, Sahibuddin & Jalaliyoon, 2014; Norman & Streiner, 2003). A commonly used technique of the factorial analysis is the correlation matrix, which demonstrates the correlation between variables, and which is recommended to be applied for the correlation coefficients over 0,3 (Taherdoost, Sahibuddin & Jalaliyoon, 2014).

3.1.3.2. Regression analysis

The regression analysis is used to determine the relationships between dependent and independent variable, in other words, it assesses how will evolve the other variable in case the one is changed (Chromy & Abeyasekera, 2005; Seltman, 2015). The variables under analysis are called explanatory and dependent variable, normally, it is not possible to determine which variable is explanatory (Norman & Streiner, 2003). Regression analysis belongs to the predictive modelling techniques and is widely used for forecasting (Norman & Streiner, 2003).

3.2. Software requirements and validation of their quality

According to Wiegers and Beathy (2013), the requirement is a customer demand for the ability of the software to perform particular functionalities or to possess certain characteristics, which bring the value to the stakeholders. Requirements engineering is the first phase of the SDL, when the customer needs are revealed, documented and passed further to the development of the product. Therefore, the issues in SRS, which were not detected in proper time, will affect all later development phases and result in higher project costs and delayed release time (Firesmith & Donald, 2007; Mund, Jakob, Fernandez, Henning, & Jonas, 2015). Moreover, according to the studies, around 40 to 50 percent of defects discovered in a software product have the origin in the requirements engineering phase (Wiegers & Beatty, 2013b). In the worst case, the defects of the SRS cause around 32,65 per cent of the projects failure (Lopes Margarido 2011).

As a result, the quality of the software requirements has become the keystone for the project success or failure and many methodologies aimed at decreasing the ambiguities in the NL SRS have recently been developed. While some of them focus on the validation the quality of the SRS, others try to prevent the appearance of the issues on the stage of writing SRS (Denger et al., n.d.). The following approaches of dealing with the requirements ambiguities can be distinguished (Tjong, Sri, Nasreddine, & Michael, 2006a):

1. The definition of linguistic rules, including undesired, inherently ambiguous words and word expressions.

2. The definition of special writing norms and patterns, which should be applied while writing requirements.

3. The definition of guidances and checking techniques.

3.2.1. Approaches focused on linguistic rules

According to numerous studies, particular words and word combinations might introduce the ambiguity to software requirements because of their inherently ambiguous nature (Tjong, Sri, Nasreddine, & Michael, 2006b; Fabbrini et al., n.d. ; Rupp, Chris, & Andreas, 2014). They can be classified as vague words, optional, subjective, weak ones (Denger et al., 2013; Fabbrini et al., n.d.). Therefore, the use of such words during writing a high quality software requirements must be eliminated.

3.2.2. Approaches focused on the definition of writing pattern

A clear sentence structure contributes significantly to avoiding ambiguities in the SRS (Denger et al., n.d.; Thongglin, Kanjana, Sylviane, & Peter, 2013). The introduction of various types of the requirements writing patterns helps to systematise the information in the sentence, make the structure clear and understandable. While some researchers introduce a general pattern for all requirements, other develop various patterns depending on the sentence semantics (Denger et al., n.d.).

The drawback of the approach is the fact that in some cases it might be problematic to write requirements using the patterns, because the information necessary in the pattern structure might be missing (Denger et al., n.d.).

Along with the writing patterns, it is recommended to avoid writing software requirements in the passive voice, and use several subjects in one sentence connecting them with conjunctions "and", "but", "or" etc. (Fabbrini et al., n.d.; Rupp et al., 2014; Tjong, Sri, Nasreddine, & Michael, 2006c). Additionally, an inevitable characteristic of a good quality requirement is the presence of the full and clear information about what action, on what object, under which conditions and by whom is performed (Rupp et al., 2014).

3.2.3. Reading techniques

The requirements validation in form of various reading techniques (inspection, review, walkthrough) focuses on checking the quality of the requirements to verify that the documented requirements demonstrate the characteristics and behaviour of the real world system to be developed (Sharma, Richa, & Biswas, 2012), as well as correspond to the recognised norms of the requirements quality, such as being correct, clear, traceable, feasible, unambiguous, complete, verifiable, traceable (Wiegers & Beatty, 2013b). Although different types of the reading techniques require different level of preparation, participants involved and the disposable budget, they are considered to be the most basic and accessible way of validating the quality of the software requirements (Chen, Pak-Lok, Sau-Fun, Tse, & Yu, n.d.; Kamsties et al., 2001).

4. Appendices

A: Research timeline.

	Activity	Duration weeks	Feb. 2016 weeks				Mar. 2016 weeks				Apr. 2016 weeks				May 2016 weeks				June 2016 weeks				July 2016 weeks				Aug. 2016 weeks				Sept. 2016 weeks			
			1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Method description	Literature study	21	■	■	■	■	■	■	■	■	■	■	■	■	■	■						■	■	■			■		■	■	■			
	Agreement on the research goals	2	■	■																														
	Studying pilot project artifacts	4			■	■	■	■																										
	Initial method draft	4			■	■	■	■																										
	Workshop to explain pilot project details	1						■																										
	Draft was added with clarification of the PM	2						■	■																									
	Literature best prctices added to method draft	5							■	■												■	■	■										
	Method improved by the survey learnings	6																		■	■	■								■	■	■		
1 survey	Choice of SRS for the survey	2						■	■																									
	Studying SRS, arranging workshop	3							■	■	■																							
	Survey planning	3							■	■	■																							
	Workshop to create questionnaire	1									■																							
	Qustionnarie finalization	1									■																							
	Sending out questionnaire	1									■																							
	Results collection	6												■	■	■	■	■	■															
	Results analysis	3																		■	■	■												
2 survey	Choice of SRS for the survey	3																■	■	■														
	Studying SRS, arranging workshop	2																						■	■									
	Survey planning	2																						■	■									
	Workshop to create questionnaire	1																								■								
	Qustionnarie finalization	1																								■								
	Sending out questionnaire	1																								■								
	Results collection	4																										■	■	■	■			
	Results analysis	2																												■	■			
3 survey	Choice of SRS for the survey	2																	■	■														
	Studying SRS, arranging workshop	2																						■	■									
	Survey planning	2																						■	■									
	Workshop to create questionnaire	1																								■								
	Qustionnarie finalization	1																								■								
	Sending out questionnaire	1																								■								
	Results collection	4																										■	■	■	■			
	Results analysis	2																												■	■			

B: Final Method Description

1. Purpose

The purpose of this method is quantifying the degree to which the participating stakeholders of a software development project share the same interpretation of the given requirements.
One of the possibilities to check the quality of requirements specifications is to apply the method of validating large requirements using surveys. The method is employed as follows. From the software requirements specification (SRS), a subset of requirements is sampled (see chapter 2) and alternative interpretations of these requirements are created using the guidelines in chapter 3. Thereafter, the survey is conducted on the stakeholders (see chapter 4), and the results of the survey are analyzed. In case of satisfactory results, the development team may continue with the further stages of the software development lifecycle (SDL). If the quality of the SRS is lower than the acceptable level, the document should be improved (see Figure 1).

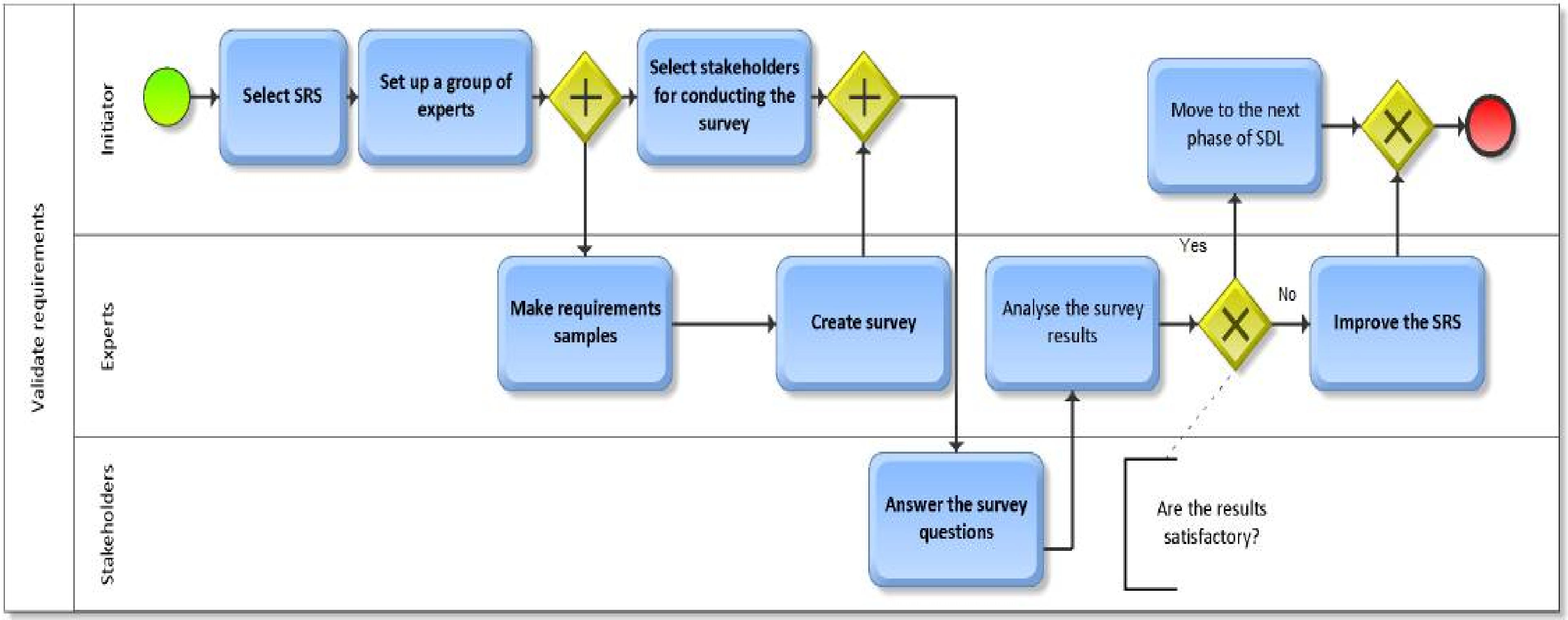

Figure 1. The process of validating large requirements specifications

2. Requirements Sampling Strategy

To ensure the most accurate metric results, the proportional stratified random sampling must be applied while choosing the pieces of requirements for the survey. The following should be taken into consideration:

1. The *functional* and *nonfunctional* requirements should be sampled.
2. Requirements should be chosen from different chapters of SRS.
3. Requirements should be chosen from parts of SRS, which are written by different authors.

The sample size may vary depending on the desired level of accuracy, however, *the recommended sample size is 10 requirements.*

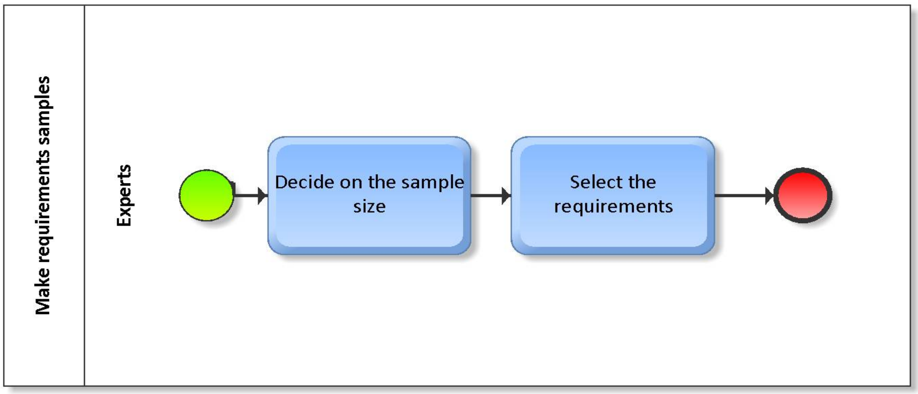

Figure 2. Requirements sampling

3. Survey Creation

The survey on the interpretation of the requirements specification represents the core component of the method. The questionnaire for the survey is created by the devised group of experts, which includes at least two of the following: a Project Manager, Product Manager, Software Architect, and Requirements Engineer of the external project. The correct interpretation is provided by the author of the requirement, who is not included in the group of experts.

To create the questionnaire, the experts read each requirement carefully and find equivocal words and expressions in the requirement. In case those are found, the experts create alternative faulty interpretations of the requirement considering revealed ambiguity. A correct interpretation provides the author of the requirement.

While formulating faulty requirement interpretations, the experts apply the listed below answer rules to provide the diversity of interpretations for a given requirement. In the questionnaire, each requirement should have from 3 to 5 of the following possible interpretations:

1. The correct interpretation
2. A negation of the correct interpretation
3. Interpretation, which is close to the correct one, but lacking some details
4. Interpretation, which is close to the correct one, but providing some excessive faulty information
5. The correct interpretation, which is formulated on a wrong abstraction level, and for this reason is considered as a faulty one
6. "None of the above answers" option. It is recommended to include this option into the list of answers, for around 10-15 percent of the questions this answer should be correct.

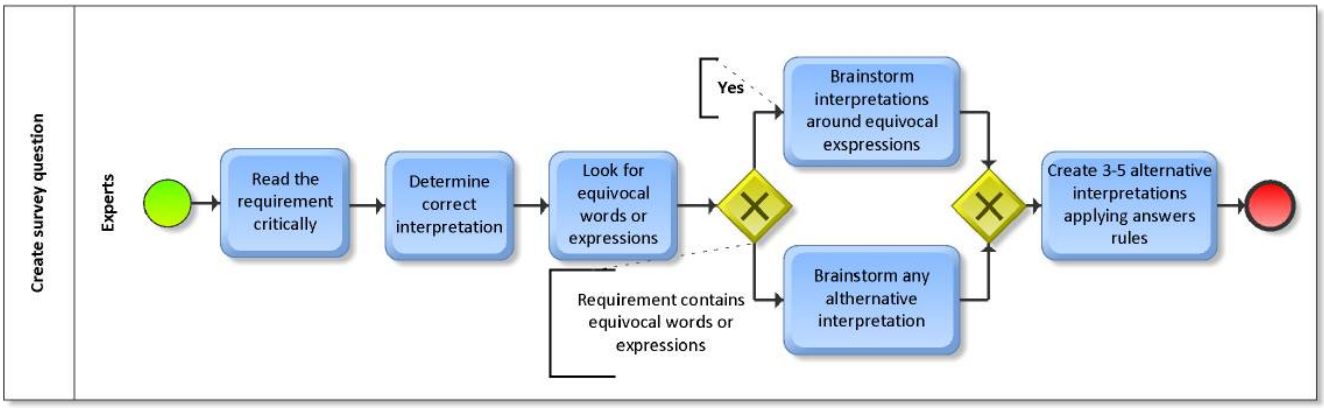

Figure 3. Creating survey questions

While formulating both correct and faulty interpretations of the requirements, the experts should not necessarily cover all the aspects of a given requirement.

The questionnaire should be design in the way that answering the survey questions does not take the stakeholders longer than 40 minutes.

4. Stakeholder Sampling Strategy

The choice of the stakeholders participating in the survey may affect its outcome. To receive the most objective metric results, the stakeholders of various roles must be engaged in the survey. When the participants of some roles are mandatory, the engagement of others might be optional. The following stakeholders must be among participants:

a. Software engineer (mandatory)

b. Architect (mandatory)

c. Usability designer (mandatory)

d. Project Manager (optionally)

e. Product Manager (of not related projects, mandatory)

f. Quality assurance engineer (optionally)

g. Technical writer (optionally)

h. Tester (optionally)

The survey should be conducted on the stakeholders of different relation to the project. The following stakeholders must be among participants:

a. Employees who are involved into the project (80 per cent of the participants)

b. Employees who are not involved into the project (20 per cent of the participants)

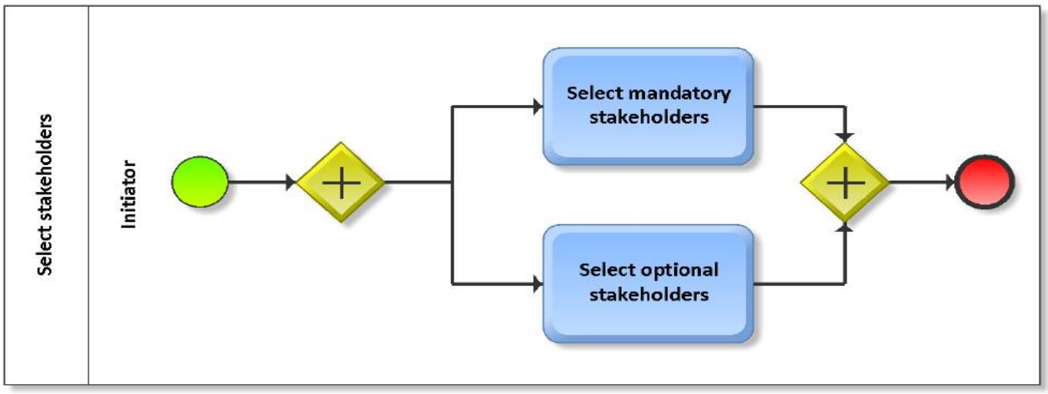

Figure 4. Stakeholders selection process.

5. Unambiguity metric

After the survey is conducted, the examined requirements are divided into three categories based on the number of correct interpretations of the given requirement.

In case 90 to 100 per cent of stakeholders interpreted a requirement alike, it is the requirement with common interpretations. If 80 per cent and more of the stakeholders responded identically, the requirement is regarded as the requirement with mainly common interpretation. If from 60 to 80 per cent of the stakeholders share the same understanding of requirement meaning, the requirement is considered to have partially common interpretation. Requirements with lower that 60 per cent identical responses are treated as requirements with different interpretations.

To calculate the metric, each category is assigned to the coefficient according to Table 1.

Coefficient	Coefficient value	Correct interpretations
K1	1	90 % and more
K2	0,8	from 80% to 89,9%
K3	0,6	from 60% to 79,9%
K3	0,3	less than 60%

Table 1. Values of ambiguity coefficients

$$Unambiguity\ Metric = \frac{K1 * N1 + K2 * N2 + K3 * N3 + K4 * N4}{N1 + N2 + N3 + N4} * 100\%$$

where N1..N4 is the number of requirements, assigned to the coefficient K1..K4 respectively.

Referencies

Adigüzel, F., Feray, A., & Michel, W. (2008). Split Questionnaire Design for Massive Surveys. *JMR, Journal of Marketing Research*, *45*(5), 608–617.

Albaum, G., Gerald, A., Roster, C. A., & Smith, S. M. (2014). Topic sensitivity and research design: effects on internet survey respondents' motives. *Asia Pacific Journal of Marketing and Logistics*, *26*(1), 147–161.

Arnab, R., Raghunath, A., Zewotir, T., & North, D. (2014). Variance Estimation from Complex Survey Designs: A Case Study of Household Income and Expenditure Survey Design 2002/03, Botswana. *Communications in Statistics - Theory and Methods*, *44*(1), 63–79.

Assael, H., Henry, A., & John, K. (1982). Nonsampling vs. Sampling Errors in Survey Research. *Journal of Marketing*, *46*(2), 114.

Banda, J. P. (2008). Designing of household sample survey: Practical guidelines. *New York, NY: United Nations, Department of Economic and Social Affairs Statistics Division*.

Barlett, J.E., Kotrlik, J. W., Higgins,C. (2001). Organizational research: Determining appropriate sample size in survey research. *Information Technology, Learning, and Performance Journal; Spring 2001; 19, 1; ABI/INFORM Collection pg. 43*

Berry, D. M. (n.d.). Ambiguity in Natural Language Requirements Documents. In *Lecture Notes in Computer Science* (pp. 1–7).

Berry, D. M., & Erik, K. (2004). Ambiguity in Requirements Specification. In *Perspectives on Software Requirements* (pp. 7–44).

Bethlehem, J. (2009). *Applied Survey Methods: A Statistical Perspective*. John Wiley & Sons.

Biemer, P. P. (2010). Total Survey Error: Design, Implementation, and Evaluation. *Public Opinion Quarterly*, *74*(5), 817–848.

Calinescu, M., Melania, C., Sandjai, B., & Barry, S. (2013). Optimal resource allocation in survey designs. *European Journal of Operational Research*, *226*(1), 115–121.

Charles, D. D. (1977). Carcass beef yield as a basis for carcass evaluation and marketing.

Australian Journal of Agricultural Research, *28*(6), 1133.

Chen, T. Y., Pak-Lok, P., Sau-Fun, T., Tse, T. H., & Yu, Y. T. (n.d.). Towards a problem-driven approach to perspective-based reading. In *7th IEEE International Symposium on High Assurance Systems Engineering, 2002. Proceedings.* http://doi.org/10.1109/hase.2002.1173126

Chromy, JR., Abeyasekera, S. (2005). Statistical analysis of survey data. In: Department of Economic and Social Affairs Statistics Division, United Nations. *Household Sample Surveys in Developing and Transition Countries. New York: United Nations*, (pp. 389-417)

Denger, C., Berry, D. M., & Kamsties, E. (n.d.). Higher quality requirements specifications through natural language patterns. In *Proceedings 2003 Symposium on Security and Privacy*. http://doi.org/10.1109/swste.2003.1245428

Doutriaux, J., & Crener, M. A. (1982). Which statistical technique should i use? a survey and marketing case study. *Managerial and Decision Economics*, *3*(2), 99–111.

Fabbrini, F., Fusani, M., Gnesi, S., Lami, G. (2014). Quality Evaluation of Software Requirements Specifications. *I.E.I. - C.N.R. Pisa, Italy*

Fabbrini, F., Fusani, M., Gnesi, S., & Lami, G. (n.d.). The linguistic approach to the natural language requirements quality: benefit of the use of an automatic tool. In *Proceedings 26th Annual NASA Goddard Software Engineering Workshop*. http://doi.org/10.1109/sew.2001.992662

Firesmith, D., & Donald, F. (2007). Common Requirements Problems, Their Negative Consequences, and the Industry Best Practices to Help Solve Them. *The Journal of Object Technology*, *6*(1), 17.

Hox, J. J., & Boeije, H. R. (2005). Data Collection, Primary vs. Secondary. In *Encyclopedia of Social Measurement* (pp. 593–599).

Kamsties E., Berry D., Paech B. Detecting Ambiguities in Requirements Documents Using Inspections

Seltman, H.J. (2015). Experimental Design and Analysis. Chapter 9. *Carnegie Mellon*

University

Kenett, R. S. (n.d.). Software specifications metrics: a quantitative approach to assess the quality of documents. In *Proceedings of 19th Convention of Electrical and Electronics Engineers in Israel*. http://doi.org/10.1109/eeis.1996.566920

Lopes Margarido, I., Faria J.P., Vidal, R. M, Vieira, M. (n.d.) Classification of defect types in requirements specifications: Literature review, proposal and assessment. *Department of Informatics Engineering, Faculty of Engineering of the University of Porto, Portugal*

Moy, P., & Murphy, J. (2016). Problems and Prospects in Survey Research. *Journalism & Mass Communication Quarterly*, *93*(1), 16–37.

Mund J. (2013). Quality Assessment of Requirement Specifications using Metrics

Mund, J., Jakob, M., Fernandez, D. M., Henning, F., & Jonas, E. (2015). Does Quality of Requirements Specifications Matter? Combined Results of Two Empirical Studies. In *2015 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*. http://doi.org/10.1109/esem.2015.7321195

Norman, G. R., & Streiner, D. L. (2003). *PDQ Statistics*. PMPH-USA.

Phung, T. D., Hardeweg, B., Praneetvatakul, S., & Waibel, H. (2015). Non-Sampling Error and Data Quality: What Can We Learn from Surveys to Collect Data for Vulnerability Measurements? *World Development*, *71*, 25–35.

Popescu, D., Daniel, P., Spencer, R., Nenad, M., & Berry, D. M. (n.d.). Reducing Ambiguities in Requirements Specifications Via Automatically Created Object-Oriented Models. In *Lecture Notes in Computer Science* (pp. 103–124).

Rupp, C., Chris, R., & Andreas, G. (2014). Das SOPHIST- RE gelwerk – Psychotherapie für Anforderungen. In *Aus der Praxis von klassisch bis agil* (pp. 123–164).

Saito, S., Shinobu, S., Mutsuki, T., Masatoshi, H., Tsuyoshi, K., & Mikio, A. (2013). Requirements clinic: Third party inspection methodology and practice for improving the quality of software requirements specifications. In *2013 21st IEEE International*

Requirements Engineering Conference (RE). http://doi.org/10.1109/re.2013.6636732

Shaughnessy, J.; Zechmeister, E.; Jeanne, Z. (2011). Research methods in psychology (9th ed.). *New York, NY: McGraw Hill.* pp. 161–175.

Sharma, R., Richa, S., & Biswas, K. K. (2012). Using norm analysis patterns for automated requirements validation. In *2012 Second IEEE International Workshop on Requirements Patterns (RePa)*. http://doi.org/10.1109/repa.2012.6359965

Taherdoost, H., Sahibuddin, S., Jalaliyoon, N. (2014). Exploratory Factor Analysis: Concepts and Theory. *2nd International Conference on Mathematical, Computational and Statistical Sciences (MCSS '14) and the 7th International Conference on Finite Differences, Finite Elements, Finite Volumes, Boundary Elements (F-and-B '14).* http://wseas.us/e-library/conferences/2014/Gdansk/MATH/MATH-49.pdf

Taylor-Powell, E., Renner, M (2003) Analyzing Qualitative Data. *University of Wisconsin-Extension Cooperative Extension, Madison, Wisconsin*

Thitisathienkul, P., Patra, T., & Nakornthip, P. (2014). Quality assessment method for software development process document based on software document characteristics metric. In *Ninth International Conference on Digital Information Management (ICDIM 2014)*. http://doi.org/10.1109/icdim.2014.6991412

Thomas, R. K. (2014). Fast and Furious … … or Much Ado About Nothing? *Journal of Advertising Research*, *54*(1), 17–31.

Thongglin, K., Kanjana, T., Sylviane, C., & Peter, G. (2013). Thai software requirements specification pattern. In *2013 IEEE 12th International Conference on Intelligent Software Methodologies, Tools and Techniques (SoMeT)*. http://doi.org/10.1109/somet.2013.6645650

Tjong, S., Sri, T., Nasreddine, H., & Michael, H. (2006a). Improving the Quality of Natural Language Requirements Specifications through Natural Language Requirements Patterns. In *The Sixth IEEE International Conference on Computer and Information Technology (CIT'06)*. http://doi.org/10.1109/cit.2006.103

Tjong, S., Sri, T., Nasreddine, H., & Michael, H. (2006b). Improving the Quality of Natural

Language Requirements Specifications through Natural Language Requirements Patterns. In *The Sixth IEEE International Conference on Computer and Information Technology (CIT'06)*. http://doi.org/10.1109/cit.2006.103

Tjong, S., Sri, T., Nasreddine, H., & Michael, H. (2006c). Improving the Quality of Natural Language Requirements Specifications through Natural Language Requirements Patterns. In *The Sixth IEEE International Conference on Computer and Information Technology (CIT'06)*. http://doi.org/10.1109/cit.2006.103

Umber, A., Ashfa, U., & Bajwa, I. S. (2011). Minimizing ambiguity in natural language software requirements specification. In *2011 Sixth International Conference on Digital Information Management*. http://doi.org/10.1109/icdim.2011.6093363

Wiegers, K., & Beatty, J. (2013a). *Software Requirements*. Pearson Education.

Wiegers, K., & Beatty, J. (2013b). *Software Requirements*. Pearson Education.

1028-2008 - IEEE Standard for Software Reviews and Audits. (n.d.). Retrieved August 07, 2016, from https://standards.ieee.org/findstds/standard/1028-2008.html