Exploritory Study in Features Used to Determine the Quality Level of Software Requirements

Mary E. Biddle
University of Kentucky
mary.biddle@uky.edu

Abstract

Software requirements documents are the foundation on which a project is built. The software quality and project success are dependent on the software requirements. Currently there are no standards for measuring the quality level of a requirements document. Without a standard, empirical studies cannot be compared or reproduced. The results are purely subjective and limited to internal comparisons with similar subjects. Typically the requirement quality is determined through internal reviews. This study is the first step in developing quality metrics for requirements. Subjects were chosen that would typically attend a review. They were asked to provide a list of features they were looking for in the review. The experiment successfully extracted a list of common attributes that can be used in further studies to create some metrics of quality for requirements documents.

1. Introduction

Software requirements documents are the basis for all development. High quality requirements documents affect the quality of the software and success of the project(1). Without quantifiable quality metrics it is difficult to compare the quality of requirements documents. Without the ability to compare, it is difficult to empirically prove the affect the documents have on the software or project. This paper describes the first phase of a study to determine a standard that could be used to assign quality metrics to a requirements document for future empirical studies. Since reviews are the typical mechanism for determining quality, a survey was used to formulate what reviewers look for in the review.

This paper is organized into 9 sections. Section 2 discusses prior work. Sections 3 and 4 explain the design of the experiment. Section 5 examines the threats to validity. In section 6 the statistical results will be analyzed and the results presented. Conclusions and future work can be found in section 7. Acknowledgements are presented in section 8. References are in section 9. The paper ends with an appendix that provides the raw data from the survey.

2. Background

The quality of software requirements has been reported as affecting the project results in several studies(2)(3). These experiments cannot be reproduced, or compared to similar experiments conducted by different organizations. One of these experiments claimed the results could not be reproduced in this statement, "A more severe threat concerns repeatability. How do different assessors affect the result?"(2). Another study reported concerns that expressed similar issues, "The first probable threat to validity might be reliability of SRS quality evaluation data."(3). Even with the number of papers addressing quality attributes of requirements documents, there are no clear metrics that could be used to compare one project to another(4)(5)(6)(7).

3. Hypotheses

The purpose of this study is to help define a list of attributes that can be used in determining hypotheses for future studies. This is considered a pre-study or exploratory study. According to the article by Barbara Ketchenham and all (8), "exploratory studies are an important mechanism for generating hypotheses and guiding further research".

Worst Case Response. Since six people responded with potentially ten responses, the worst case would be all six people would respond with ten different answers. This would generate sixty different responses.

Best Case Response. The best case would contain ten responses. All six people would generate the same ten responses.

The difference between the worst case and the best case is 60-10=50. The medium value is (50/2)+10 = 35.

H0 (Null Hypothesis): Closer to the worst case response. The Null Hypothesis would indicate that the reviewers look at completely different features. There are no common values among the reviewers. The null hypothesis will be true if it approaches the worst case of 60 different responses are generated. Values closer to the worst case would be true, therefore N0>35 features.

H1 (Alternate Hypothesis): Closer to the best case response. The alternate hypothesis indicates there are common features that reviewers look for in a document. The alternate hypothesis, will be true if the number of features approaches the best case or H1<35 features.

4. Experiment

4.1. Independent Variables.

The independent variables are the ones controlled in the experiment(9). The independent variable in this experiment is the survey question. Only one question was asked of the subjects, called the Big Question. The Big Question was "Please provide in priority order, with 1 being the most important, the top ten items you would use to measure the quality of a requirements document."

4.2. Dependent Variables.

The dependent variables are the results that are measured to test the hypotheses(9). The dependent variable in this experiment is the responses to the question.

4.3. Subjects.

The subjects in this experiment are the people that received the survey. Quota sampling was used to select our subjects. The subjects consisted of people that had experience working with software requirements. People were chosen from several different software development areas, developer, tester, requirements writer, project manager, and marketing. Someone from each area responded except marketing.

Part of the survey consisted of a background questionnaire which can be found in appendix A. Table 1 shows the subjects that responded to the survey. Sally has worked in industry 5 years as a developer. She has been a PHD student for 2 years. Sally has worked with 1 requirements document.

Person	PHD (years)	Industry (years)	Position	Project using req
Sally	2	5	Developer	1
Will	1.5	11	tester	10+
Bob	11	12	Architect	30+
Warren	0	15	Designer	1
Walt	3	0	Tester	4
Doug	0	23	Project Manager	10

Table 1. Subjects

The subjects have either provided input, used, developed, or reviewed requirements documents. The information acquired from the survey shows how the subjects have worked with requirements. In table 2, Sally has not provided input into or actually written requirements. She has reviewed and used requirements.

Person	Provide input into Req	Written Req	Reviewed Req	Used Req
Sally	no	no	yes	yes
Will	yes	no	yes	yes
Bob	yes	yes	yes	yes
Warren	yes	yes	yes	yes
Walt	no	yes	yes	yes
Doug	yes	yes	yes	yes

Table 2. Subject Requirements Experience

5. Experiment Design

This is an exploratory experiment using an open ended survey. Each subject was asked one question that requested their top ten answers. The subject's identity was kept anonymous during the request and in the results. The information this experiment is trying to extract is the criteria software engineers use to evaluate a software requirements document.

6. Threats to Validity

This section describes the threats to validity for the experiment(9).

6.1. Conclusion Validity

Conclusion validity is concerned with the ability to draw the correct conclusion from the treatment and the outcome.(10)

One of the threats to validity for this study is the sample size. Only ten people were asked to participate. Of those ten people only six people responded. There was a fairly good representation

of the different types of software engineers, but only one or two of each type responded. The results may not be a good representation of the general population.

6.2. Internal Validity

Internal validity is concerned with the treatment causing the outcome.(10)

The way the group was selected may be a threat to internal validity. The individuals were selected because of their availability and their extensive background with requirements. They may not be a good representative of the general population of requirements reviewers.

6.3. Construct Validity

The relation between theory and observation is the concern of construct validity.(10)
This is an exploratory study that only has one method and one dependent variable. Although this is the nature of the study, it is a threat to the construct validity. The results may under-represent the construct.

6.4. External Validity

The conditions that limit the ability to generalize the results to industrial practice is considered external validity.(10) There does not appear to be any threat to validity in this experiment.

7. Analysis and Results

Of the six people that responded, two listed 10 answers, two provided 7 answers, one gave 8 answers, and one listed 6 answers. There was a total of (2*10) + (2*7) + 8 + 6 = 48 features. From the responses to the questionnaire (see appendix A for the raw results), table 3 was generated. Three people claimed Readable was an attribute needed in determining quality. Two people said the Readable attribute was the most important and one person said it ranked 7.

Table 3. Survey Results.

Quality Attributes of a Requirements Document	Cites	Priority									
		1	2	3	4	5	6	7	8	9	10
1. Readable	3	2						1			
2. Organized Format	4					1	1			1	1
3. Unambiguous	5			2	1	1	1				
4. Consistent	3		1	1		1					
5. Complete	4		2		1		1				
6. Testable	4				1		2	1			
7. Concise	4			2	1						1
8. Correctly Captures Customer's needs	4	3							1		
9. Enough Detail	3			1	1			1			
10. Traceable	3		1					1	1		

Quality Attributes of a Requirements Document	Cites	Priority									
		1	2	3	4	5	6	7	8	9	10
11. Source/Rationale/Value	2		1			1					
12. Version Control (library)	3	1							1	1	
13. Change Control	1		1								
14. Reviewed	1					1					
15. List of Stakeholders	1					1					
16. Part of defined process	1							1			
17. Accurate	1				1						

Table 3: Survey Results

Another view of the total values of each feature is found in Chart 4. Three people responded with 1. Readable as a feature. Looking at the data, it appears that 5 features were chosen by 3 people. 5 features were chosen by 4 people. 5 features were chosen by 1 person. One feature was chosen by 5 people. One feature was chosen by 2 people.

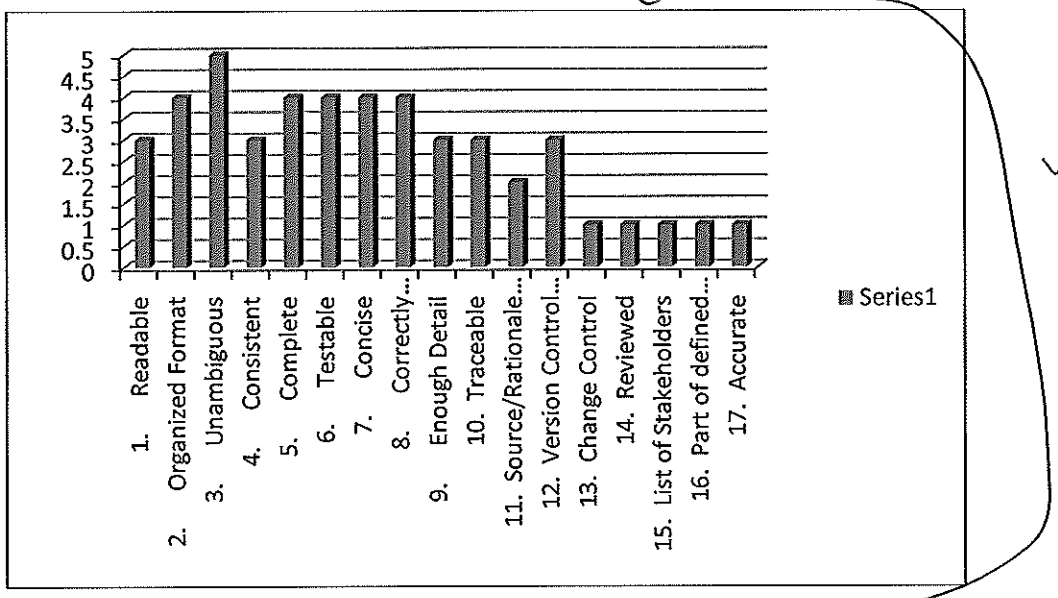

Table 4: Total Subjects Choosing Features

The dispersion of the responses is shown in Table 5. One person chose a feature (Num Chosen) was found 5 times (Frequency) which gave a relative frequency of 29.4%. Since there is a total of 6 people responding, when only one person chose one feature, the percentage of people making that choice is 16%. It is significant that 83% of the people chose one feature (unambiguous). It is also significant that 50% of the people and 66% of the people chose 5 unique features. When there is a possibility of choosing 47 different values, this dispersion is relatively small.

where was Related Work section?

Num Chosen	Frequency	Relative Frequency	Percentage of People
1	5	29.40%	16%
2	1	5.80%	33%
3	5	29.40%	50%
4	5	29.40%	66%
5	1	5.80%	83%

Table 5: Dispersion

8. Conclusions and Future Work

The results are very exciting. The survey produced a relatively small list of items that can be explored in future studies. In the future, I plan to offer the same survey to a larger group of people. Then after some analysis, future experiments will be formulated to test possible metrics.

need more here, not so informal... look @ Weikert and Billy's papers - you can get thru OpenConf, right?

9. Acknowledgements

Many thanks to all the help Dr. Jane Hayes gave by reviewing and encouraging me in this work.

10. References

~~Works Cited~~

[1] 1. **The Standish Group International Inc.** The CHAOS Report. [Online] 2006. www.standishgroup.com.

[2] 2. *Investigating the Impact of Software Requirements Specification Quality on Project Success.* **Eric Knauss, Christian E. Boustani, and Thomas Flohr.** Hannover : PROFES, 2009. LNBIP 32, pp 28-42, 2009.

3. *How does Requirements Quality Relate to Project Success or Failure?* **Mayumi Itakura Kamata, Tetsuo Tamai.** s.l. : 15th IEEE International Requirements Engineering Conference, 2007.

4. *Assessing the Quality of Software Requirements Specifications.* **Eric Knauss, Christian El Boustani.** s.l. : 16th IEEE International Requirements Engineering Conference, 2008.

5. **Gibson, J. Chris.** Quality Software Requirements. *http://www.stickyminds.com.* [Online]

6. *Experiences on Analysis of Requirements Quality.* **Petra Heck, Paivi Parviainen.** s.l. : The Third International Conference on Software Engineering Advances, 2008.

7. *Specifying Good Requirements.* **Firesmith, Donald g.** 3 July-August, s.l. : Journal of Object Technology, 2003, Vol. 2.

8. *Preliminary Guidelines for Empirical Research in Software Engineering.* **Barbara A. Kitchenham, Shari Lawrence Pfleeger, Lesley M. Pickard, Peter W. Jones, David C Hoaglin, Khaled El Emam, and Jarrett Rosenberg.** 8, August 2002, IEEE Transactions on Software Engineering, Vol. 28.