Trade-off Analysis in Web Development

An experiment on the use of QFD

Sven Ziemer
Department of Computer and Information Science
Sem Sælandsvei 7–9
NO-7491 Trondheim, Norway
sven.ziemer@idi.ntnu.no

Tor Stålhane
Department of Computer and Information Science
Sem Sælandsvei 7–9
NO-7491 Trondheim, Norway
stalhane@idi.ntnu.no

Magnar Sveen
Department of Computer and Information Science
Sem Sælandsvei 7–9
NO-7491 Trondheim, Norway
magnar.sveen@idi.ntnu.no

ABSTRACT

Many Web Applications are depending on keeping a loyal user group. This is partly archived by frequent updates, including new content, new presentation and new functionality. Time-to-market (TTM) is an important requirement for these updates. This pose a challenge for the quality of Web Applications, which have to be balanced properly. When prioritising new requirements, the benefits and the consequences have to be assessed and conflicts between requirements have to be resolved. This can be done by using trade-off analysis. A tool that can be used to facilitate a trade-off analysis is the Quality Function Deployment (QFD) method. This paper reports on an experiment that we have conducted to observe how QFD contributes to resolve conflicts between requirements, to make a prioritisation and to enhance the communication in a development team.

Categories and Subject Descriptors

D.2 [**Software Engineering**]: Miscellaneous; D.2.8 [**Software Engineering**]: Metrics–*Process metrics, Product metrics*

General Terms

Experimentation, Measurement, Management

1. INTRODUCTION

Two of the factors that determines the success ofWeb Applications is the number of returning end-users and the number of end-users that are using the services offered by a Web Application. To run a Web Application successfully depends therefore partly on how good the development team behind it adapts to the competitive environment this application is running in. A typical Web Application will be updated frequently to react to changes in the particular business, to improve the quality of the functionality it offers and to keep it an attractive place on the Web. A Web Application will also be compared with its competitor(s) and will therefore be updated to strengthen its position vis a vis its competitor(s). The time to develop new functionality, to perform all changes and to test them is limited. New functionality is often tested and validated by the end-users. A development team must be flexible to react to these requests immediately [10].

To be flexible, development teams will apply development practises that focus on TTM and flexibility. Teams are small and work in parallel [10]. Conflicting requirements are detected late and have to be resolved quickly. There is only a restricted amount of information available to solve such conflicts and to find a proper balance between TTM and quality. A simple method to help projects to balance TTM and quality and to resolve conflicts between requirements would be beneficially. In the WebSys project we have modified the QFD method to reflect the opinion of individual stakeholders and to guide a project team to find a balance that most stakeholders can support [18].

In this paper we present an experiment to test how this modified version of QFD supports the decisionmaking of small projects. The rest of this paper is organized as follow: In section 2 we present the modified QFD method and how to use it in projects. In section 3 and 4 the design of the experiment and the results are described. A discussion of the experiment and future work follows in section 5. Conclusion are given in section 6.

2. TRADEOFFANALYSIS IN WEBDEVELOPMENT

As mentioned above many Web Applications are developed in a competitive environment. In order not to loose potential users to competing Web Applications, new functionality is introduced to increase the attractiveness of the Web Application. The quality of new functionality is often not as important as its presence. If the quality is not good enough it can be improved later. This focus on TTM is sometimes called Rush-to-market. New functionality is sometimes introduced for marketing reasons only [10]. The introduction of new and cutting-edge technologies adds to the attractiveness of a Web Application. Also, the appearance of a Web Application will be changed (makeover) to give the impression of ”newness” [3].

The development activities have to adopt to this competitive environment. This is done by following a number of development practises, including [9]:

– Evolutionary Development – Applications are delivered in successive releases with short time intervals. This makes it possible to correct bugs and improve the qualitative when

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
3-WoSQ ‘05, May 17, 2005, St Louis, Missouri, USA.
Copyright 2005 ACM 1-59593-122-8/05/0005…$5.00.

www.manaraa.com

this becomes a problem. This is sometimes called Web Gardening [9].

- Parallel development – In order to to reduce the time for each development cycle, the application is developed in parallel development. This means that a developer can work on different releases in one iteration.
- Ad hoc development process – The development of Web Applications lacks the rigour and systematic approach, quality control and assurance that is found in traditional software development.
- Requirements – Requirements are evolving over time. Prototypes are used to communicate with the customer, and end-users are sometimes used to test, validate and refine the requirements.

These practises are helping the developers to deploy new versions of a Web Application at short time intervals. Updates every week or every second week are not uncommon [10]. But what is good for marketing purposes of a business is not without problems for the quality of the applications [9]. Quality is itselfs an important competitive advantage. If the quality of a Web Application is not good enough, users will go to other Web Sites. After all, the nearest competitor is only one mouseclick away. When a Rush-to-market attitude dominates the development activities the focus is mostly on the functionality and not so much on quality (non-functional requirements). TTM, newness, and quality should therefor be balanced carefully. This is not an easy task. Given the short TTM, it is hard to have full knowledge of all the consequences of a number of choices and decisions that a development team have to take.

To find the right balance between TTM, newness and quality one has to consider the pros and cons from the available options. This is mainly based on qualitative information – such as experience and beliefs – and not so much on exact knowledge or quantifiable information. The information is supplied by the involved stakeholders who express their beliefs on the available options. A decision can be made by the team by looking for a combination of options that can satisfy as many stakeholders as possible. This goalsetting has to be realistic and consider all aspects of the software product. Based on the present information, the solution that is considered the best possible by the team is chosen. This process is inspired from utility maximisation found in decision theory [15].

By using Trade-off analysis we want to support the effective decision making of small software projects. Trade-offs will be helpful for projects in several ways, by improving [18]:

- Awareness – Awareness about conflicting requirements and the possibility to balance different options to get a successful product.
- Negotiation – A way to consider the available options and their consequences. This may result in changing one or more goals.
- Communication – Bringing together all stakeholders and sharing all available information to reach a decision.

2.1 Quality Function Deployment

When choosing a tool for trade-off analysis we decided that our tool must fulfil the following requirements:

- Be easy to understand for all parties involved.
- Include both qualitative and quantitative information.
- Be applicable to all activities in a web project.
- Be applicable with all development methods and process models.

We chose the Quality Function Deployment method – QFD for short – because we find it easy to learn and use, and it is well suited for the trade-off analysis described above, because of its flexibility. It is described in [1] and also in [6].

QFD is a method for answering important questions during the requirement analysis and other phases of a software project. In the requirement phase, which is the only phase we are looking at here, QFD looks at three questions:

- What – what shall we implement?
- How – what method, technique, etc. shall we use to realise each requirement?
- How much – how much resources shall we put into meeting the requirements?

Traditionally, the QFD matrix uses three symbols to indicate the relationships between the ”whats” and the ”hows” (strong correlation, weak correlation and conflicting relation). In order to use QFD in trade-off analysis, we use numerical values in the matrix as follows:

- 9 points, strong positive relation.
- 3 points, weak positive relation.
- 0 points, no relation.
- Corresponding negative values for negative relations.

We find that the following factors make this method well suited for trade-off analysis (An example of a trade-off analysis is given in [18]):

- The requirement are prioritised by the stakeholders. In addition, the level of todays product can be compared with the most important competitor(s).
- The method allows us to include importance and dif- ficulty scores, which give us a good idea of the work needed when using each tool or method and for implementing each function
- By relating the prioritising of the stakeholders to todays level and the level of the most important competitor(s) the trade-off analysis supports the making of a realistic decision.

3. THE EXPERIMENT

As stated earlier, one of the main concerns in a trade-off situation is communication. We hope that QFD would make the communication more efficient, since it would help the stakeholders to see the consequences of their decisions and how their decisions influenced the other stakeholders success criteria. Although a trade-off can be considered as a purely technical

www.manaraa.com

problem [2], [11], we have chosen to take a somewhat wider perspective and also consider the people side of this problem.

3.1 Hypothesis and instrumentation

A good hypothesis formulation is important to any experiment since it will influence all further work such as organisation, instrumentation and analysis. Our hypothesis was that:

H0: "The use of QFD will not improve the quality of the communication in a trade-off meeting".

H1: "The use of QFD will improve the quality of the communication in a trade-off meeting".

If we want to test the hypothesis, we need to find a way to measure communication quality – the instrumentation of the project. Since communication quality to a large extent is a subjective matter, we decided to solve the problem by using a post-experiment questionnaire. The questionnaire used a Likert scale from 1 to 5 for all questions. The questions pertaining to communication quality, and thus to our hypothesis, were:

- Did you find it a difficult decision to make?
- Did the stakeholders different requirements create any conflicts in the group?
- Did you find the communication in your group to be constructive?
- Did you all agree on the final decision?

The answers to these questions will, in our opinion, be a good measure of communication quality. In addition to the scores of the individual questions we also looked at how much the persons in each group agreed on the answers to the questions. Had they, so to speak, been at the same meeting?

3.2 Experimental design

The participants were volunteers from the students taking software engineering courses – second and third year students. They were randomly organized into two groups – the treatment group – 14 persons – and the control group – 17 persons. The treatment group received a 15 minutes introduction to the QFD method. The participants were organised into small groups – three or four persons each – that should perform the trade-offs. We had

QFD-matrix for JavaScript:

	Requirements	Weights			Tools and Methods					
	S = System developer, P = Programmer, M = Marked manager	S	P	M	JavaScript	Load test	Browser test	Total S	Total P	Total M
1	Short time to market				3	-3	-3			
2	Fancy effects				3	-3	-3			
3	High performance				0	9	0			
4	Easily updated				3	0	0			
5	Easy of use				9	0	3			
6	Independent of browser				-3	0	9			
	Total System developer									
	Total Programmer									
	Total Marked manager									

Figure 1: QFD matrix

five groups that used QFD and five groups that just sat round a table and discussed the problem.

Each group consisted of persons playing roles and each role had its own agenda for the project that they were about to start. The roles and their agendas were as follows:

- Marketing manager – wants a fancy system and a short time to market.
- Systems developer – wants an easily updated browser and a user friendly system.
- Programmer – wants a challenge and the opportunity to use new technology and methods.

For the four persons groups, we used two programmers instead of one.

The technology involved was:

- Pure HTML – static information and to a large degree independent of the browser.
- Flash – a new technology as far as our firm is concerned. It gives highly dynamic systems but is hard to change once it has been put into operation.
- JavaScript – dynamic HTML that is user friendly but can cause problems with some browsers.

Each of these technologies had a QFD table that shows the impact on the stakeholders agendas. The influence of each technology and the importance of each project goal for each stakeholder were fixed. The QFD table for JavaScript is shown in figure 1.

Each stakeholder inserted his own priorities – weights – into the table and were thus able to see the effects on "his" total score if the project decided for instance to chose JavaScript.

3.3 Experiment operation

The experiments were run in meeting rooms on campus. The participants who had received QFD training were randomly assigned to one of the treatment groups while the others were randomly assigned to one of the control groups. We first run the experiments with the control groups and then, on the next day, the experiments with the treatment groups. The rooms used for each trade-off meeting were arranged so as to prevent discussions between the groups. It also made it difficult for one group to hear what was going on in any other group. In addition, one of the authors – the experiment supervisor – was present at all times in order to prevent information sharing between groups.

The participants were given their role cards and a problem description – the goal of the trade-off meeting. The role card gave a short description of their hopes – what they want to get out of the discussion – and fears – what they want to prevent. For instance, the role card for the marketing manager contained the following information:

- Want: fancy effects, short time-to-market.
- Fears: a boring website.

When the group had reached a decision, they gave a signal to the experiment supervisor, who handled them their post-experiment questionnaire. The participants were told to fill in their questionnaire without talking to each others. The experiment supervisor controlled this process. The data from the

www.manaraa.com

questionnaire were arranged into tables for later analysis. The tables pertaining to the communication quality are shown in table 1 in the analysis section.

3.4 Validity

One of the main concerns when making an experimental design is validity. The results from the validity discussions tell us what kinds of validity we can claim for the results from our experiment. We have used the validity groups used by Claes Wohlin (see [17]). The in-dept discussion on this topic is too long for a short paper. The readers should instead consult Magnar Sveens report, [13]. Here we will just identify what we consider the most important threats to validity for this experiment.

- Internal validity. Our main concern is the instrumentation for the experiment. If one or more of the questions in the post-experiment questionnaire has been misunderstood or interpreted differently by one or more persons, the results will not be valid.
- Conclusion validity is, among other things, concerned with sample size and is discussed under analysis.
- External validity. Our main concern here is the generalizability of the setting, especially the fact that none of the participants in the experiment were real stake holders.
- Construct validity. Our main concern here is the monomethod bias of the experiment. We have, however, taken any reasonable precaution to reduce this effect – for instance by analysing differences in both communication quality and group agreement.

3.5 Analysis

All results from the experiment – the data from the postexperiment questionnaire – are ordinal values. According to statistical orthodoxy we should use non-parametric statistical methods to analyse the data. Practitioners have, however, left this stance a long time ago – see for instance [12] and [7]. We have thus chosen to use the t-test to check for differences between scores on the Likert scale and the Ftest to check differences in variances. Just to put things into perspective, we quote one the grand old men of statistics – John Tukey – who states that:

> *"An oversimplified and over-purified view of what measurements are like cannot be allowed to dictate how data is to be analyzed" [14].*

The other problem we are facing is the question of sample size. We have used the formulas and ideas of Will Hopkins [8]. The important concept in sample size calculation is effect *size* – ES. If the control group has the mean μ1 and the treatment group has the mean μ2, then ES = (μ1 – μ2)/σ. If C is a function of the probabilities of type I and type II errors, then for many situations – including ours – we have that the sample size N can be found from $N = C/ES^2$. For a probability of 5% for type I errors and a probability of 20% for a type II errors, we get C = 32 and thus $N = 32/ES^2$. This assumes, however, that we have the same number of persons in both the treatment group and the control group. If the two numbers are different, we can claim statistical power equivalent to N = 4 * n1 * n2/(n1 + n2). In our case, however, the difference – n1 = 14 and n2 = 17 – is too small to matter.

A total of 31 persons participated. Thus, we can observe an effect size of approximately 1.0. This is a large to moderate effect.

The observations pertaining to the questions used to measure the communication quality are summed up in table 1. Using the t-test we find that all differences are significant at the 5 % level.

4. RESULTS

The t-tests showed that the observed differences for the communication quality factors were all significant at the 5% level. The effect sizes are between 0.8 and 1.0 and are, however, on the limit of what we can expect to observe with just 31 persons.

The results can be summed up as follows:

Did you find it a difficult decision to make? Those using QFD experienced fewer difficulties.

Did the stakeholders different requirements create any conflicts in the group? Those using QFD experienced fewer conflicts.

Did you find the communication in your group to be constructive? Those using QFD found the communication to be less constructive.

www.manaraa.com

Did you all agree on the final decision? Those using QFD found that they agreed more on the final decisions.

If it had not been for the assessment of communication constructiveness, the group the conclusion would be straight forward. Another question pertaining to communication was: ”Did it become clear to you what the different stakeholders found important?” Here those who used QFD scored slightly lower than those who did not use this method. This difference is, however, not significant at the 5% level.

As stated earlier, we also wanted to check the variation for each of the groups engaged in the trade-offs. A small variance for a group on a question from the exit questionnaire will be used as an indicator for better communication. To check this, we used a standard F-test with a 5 % significance level. Only one question showed a significant difference between trade-off meetings with and without using the QFD. ”Which stakeholder would you say got the most requirements fulfilled?” Here the groups using QFD all agreed on the outcome, while the groups not using this method had a quite mixed opinion of who got what. In all other cases, the variance for the answers was the same.

5. DISCUSSION AND FUTUREWORK

The result of the experiment encourage us to continue the work on trade-off analysis in Web Development projects. There are, however, a number of issues that we need to discuss.

The first issue is the one about the less constructive communication from the groups using QFD in the experiment. We did not expect this to happen as we believe that performing a trade-off analysis with QFD will improve the communication. We found some explanations and also learnt also a lesson. In our opinion, the explanation is that the use of a new and unknown method – in this case QFD – resulted in a simpler communication pattern. It seems that the focus change from communication with other people to communication with or through a tool. Those who used QFD spent most of their time doing calculations to see what happened to their own priorities while those who did not use any special method communicated with the other participants in the meeting. This is consistent with the observations of the experiment supervisor. The two situations are as shown in figure 2. The situation to the right hand side opens up for more communication and thus for more con- flicts but also more understanding of the other stakeholders and their needs.

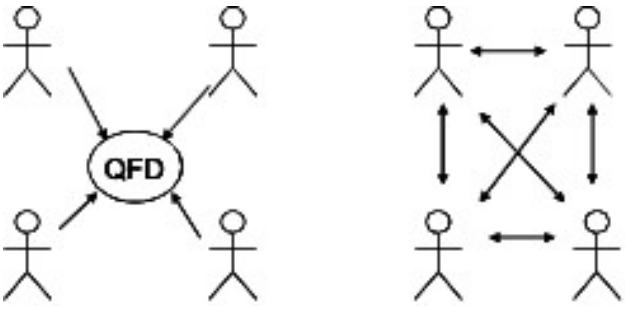

Figure 2: Difference in the communication pattern between QFD and non-QFD groups

Another explanation is that there were a time limit of 40 minutes for every group. Given the previous explanation and the time limit, there was not enough time to discuss the options. What we can learn from this is that everything that can be done to direct the focus on the trade-off and not the method should be done, for instance tool support and better training with QFD. Also, the guidelines for using such a method can include a rule that every stakeholder has to explain his weighting of the requirements, or that the group will discuss the weight for the influence from technology A on requirement M. The goal of the QFD method is

Did you find it a difficult decision to make?		
Category	Number of answers in each category	
	Using QFD	Not using QFD
1 (easy)	5	1
2	7	8
3	2	6
4	0	0
5 (hard)	0	0

Did the stakeholders different requirements create any conflicts in the group?		
Category	Number of answers in each category	
	Using QFD	Not using QFD
1 (none)	8	2
2	2	6
3	3	4
4	1	5
5 (many)	0	0

Did you find the communication in your group to be constructive?		
Category	Number of answers in each category	
	Using QFD	Not using QFD
1 (no)	0	0
2	1	0
3	4	0
4	6	8
5 (yes)	3	9

Did you all agree on the final decision?		
Category	Number of answers in each category	
	Using QFD	Not using QFD
1 (did not)	0	0
2	0	0
3	0	2
4	2	5
5 (totally)	12	10

Table 1: The results from the post-experiment questionnaire

www.manaraa.com

to support an easy and fast decision making process. Any new discussion will slow down the process. These rules should therefore only be used when the focus is too much on the tool and not on the problem at hand.

An interesting point is that the groups using QFD – even if they had a less constructive communication – found the decision easier to make. This troubles us because this can also imply that a group makes a decision without noting that there is an obvious better decision. All agreed on the early decision and are blinded by it. This again emphasises how important it is to strengthen a constructive communication inside the team.

The use of subjective evaluation rises the question of how reliable the collected data are [16], and how this evaluation is done. This issue is related to the discussion about validity. We will discuss this in some detail. The method described above is dependent on expert opinions and will be no better than the experiences available from the persons participating in the trade-off process. An open question is how to represent the collected information. We chose to use a Likert scale with 5 values (strong positive relationship, weak positive relation, no relation, weak negative relation and strong negative relation). People with a mathematical or statistical background often have uneasy feelings about doing math on values from a Likert scale. Whether this is permissible or not has been debated for more than 50 years and the debate is still raging between the persons pure at hearth and the persons with a more pragmatic view. A good summary and discussion of the positions of the parties involved can be found in [5], pg. 150 – 154, which also contains a large number of references. In this short paper, however, we will only note two points – 1) It has been working for QFD in an industrial setting for more than 40 years and 2) several active researchers have found it permissible to use statistical methods on values from a Likert scale, e.g. [4].

A related issue is the possible misuse of the weights. They are subjective measures, and by choosing higher weights a stakeholder might try to increase his influence. If this becomes a problem it will be an easy task to ask every stakeholder to explain his weighting. Any obvious misuse will then be detected.

In the future we would like to replicate this experiment with some changes to improve the communication pattern (as discussed above), and we are planning to perform case studies with companies using QFD to perform trade-off analysis. We will also work on an approach for multidimensional trade-off analysis, based on QFD.

6. CONCLUSIONS

In this paper we have reported on an experiment on using QFD for a trade-off analysis. QFD is easy to use and can be used on quantitative information. The results from the experiment shows that decision become easier to make and that there are less conflicts between the stakeholders. There is, however, the problem that groups using QFD had a less constructive communication. Thus this method has to be used as a supplement to a discussion and not as a replacement for it. We would advise small Web Development teams to use this method for trade-off analysis and to be observant about the discussed weaknesses.

7. REFERENCES

[1] Y. Akao. Quality Function Deployment. Productivity Press, Cambridge, MA, 1990.

[2] B. Boehm. Safe and simple software cost analysis. IEEE Software, 17:14–17, September/October 2000.

[3] F. Brassington and S. Pettit. Principles of Marketing, Third edition. Prentice Hall, 2003.

[4] D. Davis. Business Research for Decision Making. Duxburry Press, fourth edition edition, 1996.

[5] T. Dybå. Enabling Software Process Improvement: An Investigation of the Importance of Organizational Issues. PhD thesis, NTNU, 2001. IDI Report 7/01.

[6] R. B. Grady. Practial Software Metrics for Project Management and Process Improvement. Prentice Hall PTR, 1992.

[7] D. Hand. Statistics and the theory of measurement. Journal of the Royal Statistics Society, Series A, 159, 1996.

[8] W. G. Hopkins. A new view of statistic. Technical report, University of Queensland, 2001.

[9] S. Murugesan, Y. Deshpande, S. Hansen, and A. Ginige. A new discipline for web-based system development. In Proceedings of the Workshop on Web Engineering, ICSE99, 1999.

[10] B. Ramesh, J. Pries-Heje, and R. Baskerville. Internet software engineering: A different class of processes. Annals of Software Engineering, 14(1-4):169–195, 2002.

[11] G. Ruhe and D. Greer. Quantitative studies in software release planning under risk and resource constraints. In International Symposium on Empirical Software Engineering, 2003.

[12] S. Stevens. Psychological Scaling: Theory and Applications; Eds: H. Gulliksen and S. Messick, chapter Ratio Scales, Partition Scales, and Confusion Scales. Wiley, 1951.

[13] M. Sveen. An experiment with the use of tradeoffs and quality function deployment in web development (master thesis). Technical report, Departmend of Computer and Information Science, NTNU, 2005.

[14] J. W. Tukey. The Collected Works of John W. Tukey, Voll III: Philosophy and Principles of Data Analysis: 1949-1964; Ed: L. V. Jones, chapter Data Analysis and Behavioral Science or Learning to Bear the Quantitative Mans burden by Shunning Badmandments. Wadsworth & Brooks/Cole Advanced Books & Software, 1986, 1961.

[15] P. Weirich. Realistic Decision Theory. Oxford University Press, 2004.

[16] C. Wohlin and A. A. Andrews. Assessing project success using subjective evaluation factors. Software Quality Journal, 9:43–70, 2001.

[17] C. Wohlin, P. Runeson, M. H¨ost, M. Ohlsson, B. Regnell, and A. Wessln. Esperimentation in Software engineering – An Introduction. Kluwer Academic Publishers, 2000.

[18] S. Ziemer and T. Stålhane. The use of trade-offs in the development of web applications. In Workshop Web Web Quality, 2004.

www.manaraa.com