

An exploratory study on the effects of perceived value and stakeholder satisfaction on software projects

Huijgens, Hennie; Van Deursen, Arie; Van Solingen, Rini

DOI

[10.1145/2915970.2915987](https://doi.org/10.1145/2915970.2915987)

Publication date

2016

Document Version

Accepted author manuscript

Published in

Proceedings of the 20th International Conference on Evaluation and Assessment in Software Engineering, EASE 2016

Citation (APA)

Huijgens, H., Van Deursen, A., & Van Solingen, R. (2016). An exploratory study on the effects of perceived value and stakeholder satisfaction on software projects. In *Proceedings of the 20th International Conference on Evaluation and Assessment in Software Engineering, EASE 2016* Article a31 Association for Computing Machinery (ACM). <https://doi.org/10.1145/2915970.2915987>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

An Exploratory Study on the Effects of Perceived Value and Stakeholder Satisfaction on Software Projects

Hennie Huijgens, Arie van Deursen and Rini van Solingen

Report TUD-SERG-2016-015

TUD-SERG-2016-015

Published, produced and distributed by:

Software Engineering Research Group
Department of Software Technology
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology
Mekelweg 4
2628 CD Delft
The Netherlands

ISSN 1872-5392

Software Engineering Research Group Technical Reports:

<http://www.se.ewi.tudelft.nl/techreports/>

For more information about the Software Engineering Research Group:

<http://www.se.ewi.tudelft.nl/>

Note: Accepted for publication in the Proceedings of EASE 2016, the 20th International Conference on Evaluation and Assessment in Software Engineering, published by ACM.

© 2016 ACM. Personal use of this material is permitted. Permission from ACM must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

An Exploratory Study on the Effects of Perceived Value and Stakeholder Satisfaction on Software Projects

Hennie Huijgens
Delft University of Technology and
Goverdson, The Netherlands
h.k.m.huijgens@tudelft.nl

Arie van Deursen
Delft University of Technology
The Netherlands
arie.vandeursen@tudelft.nl

Rini van Solingen
Delft University of Technology and
Prowareness, The Netherlands
d.m.vansolingen@tudelft.nl

ABSTRACT

Context: In this paper we present an exploratory study on the insights of organizations into the *perceived value* of their software projects. Our study is based on the notion that quantifying and qualifying project size, cost, duration and defects needs to be done in relation with stakeholder satisfaction and perceived value. **Objectives:** We expect that bringing *perceived value* into the equation will help in increasing the impact such organizations deliver. **Method:** In order to find out whether our approach is practically feasible in an industrial setting, we performed an exploratory study in a Belgian telecom company. **Results:** In this study we evaluate 22 software projects that were delivered during one release. Fifty-three (53) key stakeholders provide *stakeholder satisfaction* and *perceived value* measurements in 103 completed surveys. **Conclusions:** We conclude that a focus on shortening overall *project duration*, and improving communication on intermediate progress improved *stakeholder satisfaction* and *perceived value*. Our study does not provide any evidence that steering on costs helped to improve these.

CCS Concepts

• **General and reference** → **Cross-computing tools and techniques** → **Metrics**.

Keywords

Software Economics, Evidence-Based Software Engineering, Perceived Value, Stakeholder Satisfaction, Cost Duration Index.

1. INTRODUCTION

An often cited result of the 1994 Standish CHAOS research [1] is that 70% of all software projects are problematic. Standish defines these as so-called ‘challenged projects’, meaning they were not delivered on time, within cost, and with all specified functionality [2].

This is in a certain way along the lines of what we found when studying a series of 22 finalized software projects in a Belgian telecom company. We found that the average cost overrun was 28% (ranging from -41% to 248%), and that the average duration overrun was 70% (ranging from 9% to 168%). There was only one single project that performed within a 10% cost and duration overrun boundary. As such, these projects were challenged if we adopt

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. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
EASE '16, June 01-03, 2016, Limerick, Ireland
© 2016 ACM. ISBN 978-1-4503-3691-8/16/06...\$15.00
DOI: <http://dx.doi.org/10.1145/2915970.2915987>

the way Standish defines success and failure; being the extent in which a project conforms to its original plan.

However, did all the other 21 projects fail? Is it fair to say that a project with cost overrun is a failure? Is it reasonable to say that a project that performed completely according to plan, but delivered software that no one uses, is a success?

Supported by many critical reviews of the Standish criteria [3] [4] [2], we define success and failure in this paper from a different angle, trying to include the balance between *perceived value* and *cost* into the equation. We consider that a project that is late and over budget yet returns high *perceived value* according to its stakeholders, may still be called successful. By analyzing project metrics such as *cost*, *duration*, *defects*, and *size* of the projects in connection with *stakeholder satisfaction*, *perceived value* and *quality of estimations*, we show that stakeholders define success and failure of a project different from solely measuring cost and duration overrun. Especially in domains where value is more important than predictability, e.g. agile ways of working, a limited view on conformance to planning, seems illogical. Due to the fact, that measuring the real – delivered – value of software deliveries is difficult, if not impossible, we focus in this paper specifically on *perceived value*.

In this paper, we seek to better understand the relation between *cost* and *duration* on the one hand, and *perceived value* on the other. To that end, we analyze a set of projects conducted at a Belgian telecom company (referred to in this paper as BELTEL). We analyze key attributes such as *cost*, *duration*, and *defects*, and contrast these with the opinion of a series of stakeholders using indicators such as *stakeholder satisfaction*, *perceived value*, and *Estimation Quality Factor*. In order to understand any relations between success and failure of software projects and *stakeholder satisfaction* and *perceived value* we propose the following research question:

RQ1 How do stakeholder satisfaction and perceived value relate to cost, duration, defects, size and estimation quality factor of software projects?

In answering this question, we make the following contributions:

1. We propose a light-weight value measurement technique based on post-release interviews.
2. We provide data on 22 industrial projects for which 53 key stakeholders provide *stakeholder satisfaction* and *perceived value* measurements in 103 completed surveys.
3. We contrast the resulting *perceived value* and *stakeholder satisfaction* statements with collected data on *costs*, *duration*, *defects*, *size* and *quality of estimations* and look for links between them.

The remainder of this paper is structured as follows. In Section 2 the background of the model that we use for analysis purposes is



Figure 1. A Cost Duration Matrix showing the BELTEL projects that are subject of the study.

described. Section 3 outlines the industrial context. Our research approach is elaborated in Section 4 and results are described in Section 5. We discuss the results in Section 6 and in Section 7 we compare the outcomes with related work. Finally, in Section 8 we make conclusions and outline future work.

2. BACKGROUND

In former research we built a model based on the consideration that *project size*, *project cost*, *project duration* and the *number of defects* detected during a software project are interrelated with each other [5] [6]. The model takes a project's size, measured in function points, as starting point. It then compares the actual costs (in Euros) and duration (in months) for a project of this size to benchmarked data, taken from a set of 446 finalized software projects in the financial and telecom application domains. This is done using two power regressions conducted on the 446 projects, permitting the computation of the 'expected' cost and duration of a project of a given size (measured in function points) [5] [6].

The model can be used to compare a portfolio of projects to the benchmark, by means of a *Cost Duration Matrix* [5] [6], as shown in Figure 1 for the 22 projects under study in this paper. Each project is shown as a circle. The larger the circle, the larger the project is (in function points), and the 'redder' the project is, the more defects per function point it contains. The position of each project in the matrix represents the cost and duration deviation of the project relative to the benchmark, expressed as percentages. The horizontal and vertical 0%-lines represent zero deviation, i.e. projects that are exactly consistent with the benchmark. A project at (0%, 0%) would be one that behaves exactly in accordance with the benchmark; a project at (-100%, -100%) would cost nothing and be ready immediately; and a project at (+100%, +100%) would be twice as expensive and takes twice as long as expected from the benchmark. As can be seen from the figure, most of the 22 projects in the portfolio are cheaper than what the benchmark would predict

(right of the 0%-cost bar), yet take longer than expected (below the 0%-duration bar). The 0%-lines divide the *Cost Duration Matrix* into four quadrants:

1. *Time over Cost* (top left); projects that score better than average for duration, yet worse than average for cost. In Figure 1, there are no projects in this quadrant.
2. *Good Practice* (top right); projects that score better than average for both cost and duration. In Figure 1, there are just two projects in this quadrant (projects 4.2 and 5.3).
3. *Cost over Time* (bottom right); projects that score better than average for cost, yet worse than average for duration. This is where the majority of projects are in Figure 1.
4. *Bad Practice* (bottom left); projects that score worse than average for both cost and duration. In Figure 1, there are four projects in this quadrant.

The overall performance of the portfolio is furthermore summarized through the two red 'median' lines: On average, projects in the portfolio take 48% more time than expected from the benchmark, yet are 42% cheaper. The *Cost Duration Matrix* provides a tool to compare two project portfolios in terms of project cost and duration. Our comparisons are based on the benchmark of 446 projects from the finance and telecom industries, described in more detail in [5] [6]. The benchmark of 446 projects includes 115 projects that are originally from BELTEL as well, making it a suitable benchmark to compare the new 22 projects against.

3. INDUSTRIAL CONTEXT

BELTEL is a Belgian telecom company that can be characterized as a typical large information-intensive company with a mature software delivery organization that offers a mix of delivery approaches, ranging from plan-driven to agile (Scrum). For the majority of its software development activities BELTEL has a strategic, long-term contract with one large Indian supplier, referred to in this paper as

INDSUP. Projects relate to different business domains (e.g. Internet, Mobile Apps, Data warehouse, Billing, Customer Relationship Management). During the past three years, BELTEL has adopted a metrics program to collect data on size, cost, duration, the number of defects, and the quality of estimations of finalized software projects. This data is used to analyze project performance at BELTEL, to benchmark project performance, and to continuously improve the software delivery process within BELTEL. Approximately six months ago, BELTEL changed its strategic focus from cost-based (steering on efficiency and operational excellence) to value maximization and shortening time-to-market. To facilitate this, BELTEL has collected additional data, addressing business value and customer satisfaction.

In the present paper, we will compare these with the data on costs and duration that were also collected, in order to better understand the relationships between various project success indicators. Development projects at BELTEL are conducted independently, yet grouped for deployment into so-called releases. Once a project passes its system test it is promoted to a release, which typically contains multiple projects. Releases are further tested and deployed as a whole. Within BELTEL eight subsequent releases are performed each year. In this paper, we will study data from 22 projects coming from four different releases.

4. RESEARCH APPROACH

The goal of this paper is to contrast project metrics such as *cost*, *duration*, *defects* and *quality of estimations* with *stakeholder satisfaction* and *perceived value*. We argue this will help to better understand the backgrounds of software projects as a guide for building future software portfolios.

As described before the Standish criterion [1] states that success and failure are related to the quality of project estimates. In order to explore alternatives, we test for association between paired samples, using Pearson's product moment correlation coefficient and resulting p-values in case our data is normally distributed or Spearman Rank Correlation when the data is not normally distributed. To mitigate the risk that we find coincidental correlations we perform an exploratory study that confronts correlated metrics with findings from qualitative results from analysis of the free format text from the surveys.

4.1 Definitions

In this paragraph we describe and explain the major metrics that are collected and analyzed for the subject projects.

4.1.1 Project Metrics

Four project metrics are collected on each project that is subject of the case study. For all projects, we collect *project size* (measured in function points), *project cost* (in Euros), *project duration* (in months), and the *number of defects* found during the project. Based on this, we determine the cost per function point, days per function point, and defects per function points, using in each case size in function points as weighting factor.

4.1.2 Estimation Quality Factor

Estimation Quality Factor (EQF) is a measure of the deviation of a forecast to the actual. EQF is a forecasting metric that depicts the quality of forecasts made during a project. The measure was defined by Tom DeMarco [7]. DeMarco defines EQF by:

$$EQF = \frac{\text{Area under actual value}}{\text{Area between forecast and actual value}}$$

We use the formulization proposed by Eveleens and Verhoef [8]. It allows us to quantify the quality of forecasts. A low EQF value means that the deviation of the forecasts to the actual is large. EQF is measured for both cost and duration.

4.1.3 Cost Duration Index

The *Cost Duration Index* is a measure of the relative position of a project within the *Cost Duration Matrix* (see Figure 1). The index is represented as a number between zero and one hundred. In practice most projects score between 80 and 99. A high index corresponds to a good position in the *Cost Duration Matrix* (best is top-right in the *Good Practice* quadrant). The index is based on the geometric mean of two proportions comparing the actual value to the benchmark value:

$$p = \sqrt{\frac{\text{Actual Duration}}{\text{Benchmark Duration}} * \frac{\text{Actual Cost}}{\text{Benchmark Cost}}}$$

We subsequently normalize this p to a value ranging from 0-100 with 100 being best via:

$$\text{Cost Duration Index} = \frac{(p_{\max} - p)}{(p_{\max})} * 100$$

4.1.4 Stakeholder Satisfaction

Stakeholder satisfaction is a measure of the satisfaction of stakeholders of a specific project with the way a project was performed and with the results as delivered by that project. *Stakeholder satisfaction* is measured by asking stakeholders of a specific project to rate their satisfaction on two aspects; the way a project was performed (the project's process), and with the results as delivered by a project (the project's result), for which we use questions with a 1 to 5 rating scale.

4.1.5 Perceived Value

Value is essentially delivered to customers of BELTEL. However, in this specific context it is difficult, if not impossible, to measure real value as delivered to customers. As an alternative, we measure *perceived value* as a qualitative measure of the perception of stakeholders of each project.

Perceived value is measured for each stakeholder in a specific project, on four aspects: BELTEL's customers, BELTEL financial, BELTEL internal process effectiveness, and BELTEL innovation. We base the use of the four perspectives *Customer*, *Financial*, *Internal Process*, and *Innovation* on the Balanced Scorecard [9]. Based on the results per project of the four perceived value measures a *perceived value (overall)* is calculated, with the number of measures (excluded the choice "Don't know") as weighting factor.

4.2 Project Selection

Because we are particularly interested in data of finalized projects, all metrics are measured once a release is finalized, since only then we know the real cost and duration of projects. Because we want to measure the effects of *stakeholder satisfaction* and *perceived value* on a software portfolio as a whole, we did not make any selection in the subset of projects within each release, except for the fact that we only selected projects that delivered software functionality (the projects could be counted in Function Points). Projects that do not include any software component (e.g. infrastructure projects or configuration projects) are excluded from our study.

4.3 Data Collection procedure

A major part of the data collection for our case study was performed within the measurement capability that is operational within BELTEL. Data collection on *cost*, *duration*, *number of defects*, *size*, and calculation of both *Estimation Quality Factor* metrics was performed by members of a measurement team that was supported (for performing Function Point counts) by measurement staff of BELTEL's main Indian supplier. All project data was stored in a measurement repository that is provided to use for our study. The lead author of the study was working as lead of BELTEL's measurement team.

Besides the project data that was collected as an operational practice, we collected data on *stakeholder satisfaction* and *perceived value*. To do so we conducted a questionnaire survey with stakeholders from BELTEL. The list of stakeholders was prepared in cooperation with the Project Managers of the applicable software projects, and consists of a mix of business and IT representatives that were involved in the subject projects. We asked the participants, who are stakeholders of a specific software project within a release, to rate their satisfaction with the way the project was performed and to rate their perception of the value that was added by the project. Besides ratings on a 1-5 rating scale we asked the participants to add free format text as an explanation of their perceptions. The questionnaire consists of five questions:

1. *What was your role in project PROJECT_NAME?*
2. *How satisfied are you with the way project PROJECT_NAME was performed (the project's process)?* (1-5 rating scale);
3. *How satisfied are you with the results of project PROJECT_NAME (the results as delivered by the project)?* (1-5 rating scale);
 - a. *How would you rate the delivered value of project PROJECT_NAME to the following aspects* (1-5 rating scale, with 'Don't know' as an option; this choice was excluded from further analysis)? a) *BELTEL's Customers (Value in terms of delivered to customers of BELTEL)*; b) *BELTEL Financial (Value in terms of financial revenue for BELTEL)*; c) *BELTEL Internal Processes (Value in terms of improvement and/or proper performance of BELTEL's internal processes)*; d) *BELTEL Innovation (Value in terms of innovation of BELTEL's products or services delivered to its customers)?*
4. *Are there any additional comments or suggestions you'd like us to know about this project?* (Free format text).

With regard to question 4: the additional information (between brackets) was shown to the participants when hovering over a question mark next to the text of each of the four aspects.

4.4 Analysis Procedure

In order to explore potential relationships between the collected metrics, we tested for association between paired samples. Because all sample data is normally distributed (see Table 2 for details on skewness and kurtosis), we used a Pearson's product moment correlation coefficient test for this purpose. In order to understand the underlying principles that can explain the outcomes of the quantitative analysis, we studied the free format text from the surveys.

5. RESULTS

5.1 Description of the Projects

Within the scope of our study we evaluated four software releases within BELTEL, covering a total of 22 software projects. Table 1 gives a brief description of each project, where the numbering of

Table 1. An overview of the projects in scope of the case study.

| Project ID | Project Description |
|--------------|--|
| Project 3.1 | Rules- and regulations driven small Billing project |
| Project 3.2 | Implementation of a control on a Billing application |
| Project 3.3 | Release-based enhancements on CRM-application (Scrum) |
| Project 3.4 | New campaign management tool (3 rd part of a program) |
| Project 3.5 | Release-based enhancements on a mobile App (Scrum) |
| Project 4.1 | Enhancements on a Billing application |
| Project 4.2 | Release-based enhancements on CRM-application (Scrum) |
| Project 4.3 | Frontend project: Connect Google Play |
| Project 4.4 | Rules & Regulations enhancement: fee for customers |
| Project 5.1 | Release-based enhancements on CRM-application (Scrum) |
| Project 5.2 | New campaign management tool (4 th part of a program) |
| Project 5.3 | Data warehouse 4 sprints of an iteration (Scrum) |
| Project 6.1 | Enhancement to integrate payment by credit-card-aliases |
| Project 6.2 | Enhancement to implement Apple Store code |
| Project 6.3 | Release-based enhancements on CRM-application (Scrum) |
| Project 6.4 | Adapt a procedure on an online platform |
| Project 6.5 | E-invoice for a subset of customers in a Billing system |
| Project 6.6 | Easy Script for cleanup of master MSISDN |
| Project 6.7 | Rules & Regulations project on a Billing application |
| Project 6.8 | Frontend enhancement: Shopper user interface e-services |
| Project 6.9 | Once-only migration project |
| Project 6.10 | New Order Management System (part of program, Scrum) |

the projects indicates in which release each project was finalized (e.g. Project 6.4 finalized in Release 6). The software projects in scope represent a varied outline of BELTEL's software project portfolio. It includes projects of different business domains, sizes, cost patterns, durations, and delivery approaches. Some projects are typically once-only, with teams that were put together for the purpose of one project only. Others are part of subsequent iterations within a release structure with a steady heartbeat and a fixed, experienced team. Sixteen projects are characterized as plan-driven, while six followed a more agile (Scrum) delivery approach.

All projects were performed separately. Yet from the User Acceptance Testing onwards they were combined as a release deployed into BELTEL's production environment. Looking at the total cost of a release, on average 60% was spent on software projects. The remaining cost were spent on infrastructure projects, small innovations, and configuration projects, and as such do not fit into the *Cost Duration Matrix* approach. These projects are out-of-scope for this case study.

Table 2 gives an overview of the descriptive statistics of the projects involved in the case study. Project 6.3, a small (16 FPs) release-based enhancement on a CRM-application that was performed in a Scrum way, represents the minimum cost and a short duration of 6.51 months. This project scores in the *Good Practice* quadrant in Figure 1 and shows the highest score of all for *stakeholder satisfaction* for both process and result. To put things in perspective, the maximum cost is linked to Project 6.10, an implementation of a part of a new order management system. This project took 10.06 months to finalize. Yet, also this project scores in the *Good Practice* quadrant, mainly due to the high number of function points that are delivered; 324 FPs. This project also scores well for both *stakeholder satisfaction* and *perceived value*. We note that

Table 2. Descriptive statistics of the Project Data.

| Survey Data (n = 22) | | | | | |
|----------------------|---------------------|--------------------|---------------------------|--------------------|-------------------|
| | Cost Duration Index | Project Cost (EUR) | Project Duration (Months) | Project Size (FPs) | Number of Defects |
| Minimum | 80.80 | 8,000 | 4.96 | 12 | 1 |
| First Quartile | 88.85 | 44,001 | 8.37 | 25 | 3 |
| Median | 92.07 | 66,209 | 10.18 | 39 | 9 |
| Third Quartile | 95.06 | 118,876 | 11.73 | 126 | 23 |
| Maximum | 96.71 | 296,000 | 19.03 | 324 | 223 |
| Mean | 91.41 | 99,615 | 10.20 | 79 | 29 |
| Skewness | -0.94 | 1.27 | 0.78 | 1.71 | 3.19 |
| Kurtosis | 1.16 | 0.77 | 1.43 | 2.71 | 10.89 |
| St. Deviation | 3.84 | 78209.91 | 3.22 | 82.00 | 55.20 |

both referred at projects are performed in a Scrum way. See the Technical Report [10] for detailed project data.

Besides project metrics we collected data on *stakeholder satisfaction* and *perceived value* by sending an online survey questionnaire to applicable stakeholders of each software project once the technical go live was performed. The overall completion rate of all surveys was 69%. Over a period of four releases 103 surveys were completed by 53 individual respondents. One respondent could answer surveys for different projects in one release, or repeated surveys for a series of iterative projects over different releases. An extended overview of all metrics that are collected for each project that is subject in this case study is to be found in the Technical Report [10].

5.2 Results of the tests for association

To identify potential relationships between the different metrics that we collected we performed a series of tests on paired samples of each metric, by using Pearson's Product Moment Correlation Coefficient (see the Technical Report [10] for details). The results of these tests are shown in Table 3. The table is setup in the form of a matrix that pairs sets of two metrics. For each pair the correlation coefficient is shown, including (between brackets) the associated p-value. A color indicates correlation: dark grey indicates a strong (positive or negative) linear relationship, bright grey indicates a moderate linear relationship, light grey indicates a weak linear relationship. Analysis of the results of the pairwise correlation test resulted in the following four observations.

Observation 1: Project duration relates to stakeholder satisfaction, perceived value, estimation quality and the cost-duration index, but not with project size.

Where in many other organizations a clear correlation is found between *project size* and *project duration*, BELTEL shows an atypical pattern. *Project size* and *project duration* are not related in any way, indicating that regardless the size of a project the duration is always on average nine months. A strong downhill (negative) linear relationship that occurs between *project size* and *Days per FP* confirms this finding, indicating that delivery of one function point goes faster when projects are bigger in size.

Duration shows a moderate downhill (negative) linear relationship with *Cost Duration Index*, *stakeholder satisfaction* (both *process* and *result*), *perceived value* (*process*), and *Estimation Quality Factor* (*Duration*). In other words; shorter project durations relate

with a better position on the *Cost Duration Index* (e.g. more towards *Good Practice*), more satisfied stakeholders for both *process* and *result*, higher *perceived value* with relation to the internal process, and higher values for the *Estimation Quality Factor* with regard to duration, indicating duration forecasts of better quality.

Observation 2: Perceived value relates positively (weak) with project size, costs per function point, and days per function point.

Two observations are related to *perceived value*. A weak uphill (positive) linear relationship occurs between *project size* and *perceived value* (*overall*). A moderate relationship is found between *project size* and *perceived value* (*innovation*). Both findings are an (however weak) indication that *perceived value* is higher for bigger projects in size.

Another finding with regard to *perceived value* is that moderate downhill (negative) linear relations occur between on the one hand *perceived value* (*Overall*) and on the other *Cost per FP* and *Days per FP*. This indicates that lower *Cost per FP* and lower *Days per FP* relate to higher scores for *perceived value* (*overall*). This matches the finding above that *perceived value* is higher for bigger projects in size. We assume that the (however weak) relationship between *project size* and *perceived value* as described in the paragraph above helps to enhance this effect.

A third finding on *perceived value* is about a strong downhill (negative) linear relationship that is observed between *perceived value* (*financial*) and *Estimation Quality Factor* (*Cost*). This indicates that a high *perceived value* with regard to financial aspects corresponds with cost estimates of low quality. We have no explanation for this phenomenon, and will explore its broader occurrence in further research.

Observation 3: Cost duration index moderately relates to stakeholder satisfaction for both process and result and to the quality of estimations for duration.

With regard to the *Cost Duration Index* we observe a moderate uphill (positive) linear relationship between on the one hand *Cost Duration Index* and on the other *stakeholder satisfaction* (both *process* and *result*), and *perceived value* (both *overall* and *innovation*). A moderate uphill (positive) linear relationship, however not strongly significant, is found between *Cost Duration Index* and *Estimation Quality Factor* (*Duration*). In other words, a high position on the *Cost Duration Index* (towards *Good Practice*) relates to high scores for *stakeholder satisfaction* for both *process* and *result*, high scores on *perceived value* for both *overall* and *innovation*, and cost forecasts of better quality.

Observation 4: Stakeholder satisfaction for both process and result are interrelated to each other and with the quality of estimations for duration

Three findings are combined in the observation with regard to *stakeholder satisfaction*. First, *stakeholder satisfaction* (*process*) and *stakeholder satisfaction* (*result*) are moderate uphill (positive) related to each other, indication that a high score for satisfaction with regard to the process corresponds with a high score for satisfaction with the result that is delivered. We see a moderate uphill linear relationship between *stakeholder satisfaction* (*result*) and *perceived value* (*innovation*), indicating that a high satisfaction with the delivered result relates with a high score of value with

Table 3. Matrix with test results of association between paired samples, using Pearson’s product moment correlation coefficient.

| | Project Size | Project Cost | Project Duration | Number of Defects | Cost per FP | Days per FP | Defects per FP | Cost Duration Index | Stakeholder Satisfaction (Process) | Stakeholder Satisfaction (Result) | Perceived Value (Overall) | Perceived Value (Customer) | Perceived Value (Process) | Perceived Value (Financial) | Perceived Value (Innovation) | Estimation Quality Factor (Cost) |
|--------------------------------------|-----------------|-----------------|------------------|-------------------|-----------------|-----------------|-----------------|---------------------|------------------------------------|-----------------------------------|---------------------------|----------------------------|---------------------------|-----------------------------|------------------------------|----------------------------------|
| Project Cost | 0.76 (0.00) | | | | | | | | | | | | | | | |
| Project Duration | 0.05 (0.82) | 0.21 (0.34) | | | | | | | | | | | | | | |
| Number of Defects | 0.77 (0.00) | 0.64 (0.01) | 0.01 (0.97) | | | | | | | | | | | | | |
| Cost per FP | -0.53 (0.01) | -0.14 (0.54) | 0.11 (0.63) | -0.26 (0.31) | | | | | | | | | | | | |
| Days per FP | -0.70 (0.00) | -0.56 (0.01) | 0.17 (0.45) | -0.37 (0.14) | 0.68 (0.00) | | | | | | | | | | | |
| Defects per FP | 0.07 (0.80) | 0.24 (0.35) | 0.14 (0.60) | 0.55 (0.02) | 0.38 (0.11) | 0.05 (0.86) | | | | | | | | | | |
| Cost Duration Index | 0.50 (0.02) | 0.04 (0.84) | -0.50 (0.02) | 0.23 (0.38) | -0.89 (0.00) | -0.68 (0.00) | -0.39 (0.12) | | | | | | | | | |
| Stakeholder Satisfaction (Process) | 0.14 (0.54) | 0.05 (0.81) | -0.46 (0.03) | 0.16 (0.55) | -0.34 (0.12) | -0.26 (0.25) | -0.61 (0.01) | 0.46 (0.03) | | | | | | | | |
| Stakeholder Satisfaction (Result) | 0.15 (0.51) | -0.06 (0.79) | -0.53 (0.01) | 0.11 (0.68) | -0.26 (0.25) | -0.16 (0.48) | -0.05 (0.85) | 0.45 (0.03) | 0.54 (0.01) | | | | | | | |
| Perceived Value (Overall) | 0.45 (0.04) | 0.28 (0.22) | -0.04 (0.86) | 0.26 (0.33) | -0.56 (0.01) | -0.67 (0.00) | -0.03 (0.91) | 0.43 (0.05) | 0.18 (0.42) | 0.40 (0.07) | | | | | | |
| Perceived Value (Customer) | 0.20 (0.39) | 0.11 (0.64) | 0.26 (0.26) | 0.09 (0.74) | -0.23 (0.31) | -0.31 (0.17) | 0.13 (0.64) | 0.07 (0.76) | -0.08 (0.72) | 0.21 (0.36) | 0.52 (0.02) | | | | | |
| Perceived Value (Process) | 0.12 (0.60) | 0.06 (0.80) | -0.52 (0.02) | 0.29 (0.27) | -0.41 (0.07) | -0.39 (0.08) | 0.03 (0.92) | 0.44 (0.04) | 0.17 (0.45) | 0.33 (0.14) | 0.68 (0.00) | 0.05 (0.82) | | | | |
| Perceived Value (Financial) | 0.23 (0.34) | 0.35 (0.13) | 0.12 (0.61) | 0.00 (0.99) | -0.03 (0.91) | -0.23 (0.32) | -0.03 (0.91) | 0.04 (0.87) | 0.14 (0.55) | 0.08 (0.72) | 0.71 (0.00) | 0.05 (0.82) | 0.39 (0.08) | | | |
| Perceived Value (Innovation) | 0.56 (0.01) | 0.25 (0.28) | -0.14 (0.55) | 0.30 (0.27) | -0.60 (0.00) | -0.61 (0.00) | -0.13 (0.62) | 0.62 (0.00) | 0.23 (0.32) | 0.58 (0.01) | 0.70 (0.00) | 0.25 (0.27) | 0.32 (0.16) | 0.36 (0.11) | | |
| Estimation Quality Factor (Cost) | 0.04 (0.91) | -0.18 (0.61) | -0.21 (0.56) | -0.38 (0.40) | -0.18 (0.61) | -0.27 (0.44) | -0.47 (0.29) | 0.51 (0.13) | -0.17 (0.64) | 0.16 (0.66) | -0.21 (0.59) | -0.10 (0.81) | 0.23 (0.56) | -0.76 (0.02) | 0.41 (0.27) | |
| Estimation Quality Factor (Duration) | -0.20 (0.40) | -0.18 (0.45) | -0.60 (0.01) | -0.11 (0.70) | 0.06 (0.82) | -0.10 (0.67) | -0.18 (0.50) | 0.24 (0.30) | 0.43 (0.06) | 0.47 (0.04) | 0.00 (0.99) | -0.22 (0.38) | 0.30 (0.22) | -0.16 (0.54) | 0.12 (0.63) | 0.26 (0.47) |

The table above shows results from a test of association between paired samples of the 22 software projects from the case study, using Pearson’s product moment correlation coefficient. The overview shows for each test the correlation coefficient and between brackets the p-value. A color indicates samples that are correlated: dark grey indicates a strong (positive or negative) linear relationship, bright grey indicates a moderate linear relationship, light grey indicates a weak linear relationship.

regard to innovation aspects. Third, weak uphill linear relationships are observed between *stakeholder satisfaction* (both *process* and *result*) and *Estimation Quality Factor (Duration)*. This might indicate that high satisfaction corresponds with estimations of good quality. Finally, we mention two strong uphill (positive) linear relationships that are found with regard to *project size*, *project cost* and *number of defects*. *Project size* and *project cost* are strongly related, showing a strong uphill linear relationship, indicating that bigger projects (in FPs) have on average higher costs. An equal effect is seen between *project size* and *number of defects*. A strong uphill linear relationship indicates that bigger projects (in FPs) show more defects during the project itself. This effect is known from many related studies [11] [5] and as such not a surprise in our research.

Summarized we conclude that strongly significant relationships are found between the metrics that are in scope of our case study, however these are moderate. In the next paragraph we challenge our observations by linking them to the free format text that resulted from the surveys that are performed at closure of each release.

5.3 Results of the free format text analysis

In order to compare the outcomes of the quantitative analysis of the project metrics with the survey we coded the free format text that resulted from the surveys. We used the tool Qualyzer¹ for this purpose. See Table 4 for the outcomes of the coding process. In this paper we only include a subset of comments given by participants from the survey. All free format text from the survey can be found in the Technical Report [10].

¹ <http://qualyzer.bitbucket.org>

Table 4. Results of the analysis of free format text.

| | Points of attention for satisfaction and value | Count |
|-----|---|-------|
| A1 | Quality (good quality 27, bad quality 12) | 39 |
| A2 | Communication (good communication 21, bad communication 17) | 38 |
| A3 | Deployment (issues with implementation 19, issues in production 9, bad or delayed implementation 9) | 37 |
| A4 | Requirements (requirements not clear 15, good requirements 9, requirements creep 5, bad documentation or design problems 4) | 33 |
| A5 | Stakeholders (satisfied stakeholders 29, low stakeholder involvement 3, unsatisfied stakeholders 1) | 33 |
| A6 | Duration (good estimation of duration; in-time delivery 16, bad estimation of duration 7) | 23 |
| A7 | Testing (good testing or good test environment 8, problems with testing 9, delayed testing 3) | 20 |
| A8 | Process (smooth, lean, or mature process 11, (agile) process needs improvement 3, bad process 2, process not according to standards 2) | 18 |
| A9 | Project Management (scope problems 10, good project management 3, scope delivered 3) | 16 |
| A10 | Agile Development (good product owner 4, good backlog management tool 2, use of tools unclear 1, agile process needs improvement 1, traditional release in agile process 1) | 14 |
| A11 | Supplier Management (issues with supplier 11, good relation with supplier 1, bad alignment between parties 1) | 13 |
| A12 | Team Aspects (good team spirit 7, team not fixed 1) | 8 |
| A13 | Release Management (bad alignment project and release 6, release delayed 1) | 7 |
| A14 | Value Aspects (good value 2, issues with value 4) | 6 |
| A15 | Cost Aspects (within time and budget) | 1 |

Table is sorted on Count.

5.3.1 Quality, Deployment and Testing (A1, A3, A7)

The first thing that strikes us when looking at the results of the coding process is that aspects with regard to quality are high on the list of items that apply to the stakeholders. Most remarks (27) were about good quality:

'No bad surprises after implementation' (P41).

'No incidents occurred and positive feedback from the users' (P25).

However, a number had to do with low quality issues:

'Bad quality of deliverables' (P13).

'Escalation because of bad quality took way to long, the last couple of weeks everything need to be done at once' (P18).

A large number (37) of negative comments given in the survey was related to the deployment of projects within a release into BELTEL's production environment. Most had to do with issues that occurred during this process (e.g. problems with environments or incidents in production that needed to be fixed):

'It took more than a year to get this project live. It even had to be rolled back from production a few times' (P13).

'The processes ran well but still some improvement can be done in deployment of fixes or in selecting the correct team to solve a ticket' (P19).

'Overall satisfied but one-third of the functionalities does not work' (P28).

'Project is still not fully delivered. Some outstanding issues are still being resolved' (P48).

An explanation for the fact that many issues occur after going technically live is that BELTEL uses the first week (or sometimes a longer period) to test deployments in the production environment. Usually projects are not commercially live during that period.

Comments with regard to testing are related to these deployment issues. Also here we find a majority of comments that are related to issues with test environments and the test process itself.

'We encountered issues (no test environment available) that must not be there and be handled by INDSUP' (P14).

'The internal processes of INDSUP and data warehouse team are not always very smooth which is causing tests to be blocked for days/weeks' (P19).

'A lot of discussion on how we need to test...' (P39).

Observation 5: Most comments of satisfied stakeholders are about good quality, however dissatisfied stakeholders say test and deployment need improvements.

5.3.2 Communication (A2)

The second most mentioned point on the stakeholder's list is about communication. Many remarks are related to good communication:

'Upfront discussion of all test results, hence no surprises at the go/no go meeting' (P48).

'Overall for Project 3.3, I am positive about the continuous flow in communication and delivery' (P01)

'Regular status meetings are good for clarity' (P02).

A number of remarks have to do with good communication between parties. A remarkable finding here was that these four remarks all were related to external suppliers in the frontend development of website and app development, and not with the main strategic supplier INDSUP:

'Great collaboration between BELTEL's departments and the development partner' (P21).

'Very good result, very good collaboration' (P31).

However, not all is well with communication. Besides the 21 positive remarks, there are 17 suggestions for improvement.

'Communication and involvement for agile items is limited to the bare minimum, so the added value of release management is not really big here. The whole agile process is still pretty blurry to most of its stakeholders, so this definitely needs to be improved' (P48).

'The <name of a supplier> team even delivers items in the application without anyone being aware' (P13).

Observation 6: Many satisfied stakeholders comment about good communication. A similar number of dissatisfied stakeholders says communication needs to be improved.

5.3.3 Requirements (A4)

Most of the comments (15) related to requirements were about unclear requirements that hinder a project's progress:

'Interpretation from requirements can be different and cause issues at testing phase' (P40).

'What is not clear to me is the actual content of a requirement. A requirement starts off in the tool and you can have several comments. But it is not clear to me what end requirements are agreed with necessary stakeholders' (P45).

A limited number of comments (9) were made on bad documentation, design problems and requirements creep:

'Requirement changes even during testing phase' (P13).

'There should be a standard for the documentation, since it is an entry criterion for user acceptance testing and too much time is lost with discussions on the content' (P47).

But also some comments were made on the availability of good requirements:

'Clear requirements, clear testing process' (P25).

Observation 7: Many comments of dissatisfied stakeholders are about unclear requirements, bad documentation, and requirements creep.

5.3.4 Stakeholder Satisfaction and Duration (A5, A6)

Many of the comments related to stakeholder aspects (29) were about satisfied stakeholders. Most comments had to do with the quality of delivery and the time-to-market of delivery:

'Business was happy with the results so I'm happy too' (P48).

'No issues were found. High business value' (P44).

'In a short time frame the most important functionalities were delivered' (P42).

Project duration and time-to-market is mentioned by many (23) participants, were most comments (15) are about on-time delivery:

'Went relatively quick and was delivered on time' (P33).

'This project is the first one that was partially delivered in the new <agile> way-of-working so some growing pains were observed. Nevertheless, the speed and quality of delivery wasn't bad' (P48).

Observation 8: Many comments of satisfied stakeholders are related to good quality of duration estimates. Dissatisfied stakeholders comment about long duration and schedule overrun.

5.3.5 Agile, Value, and Process (A10, A14, A8)

A more agile delivery process is one of the key innovations that are implemented within the software delivery organization of BELTEL. Knowing this we argue that the low number of comments related to this aspect (14) does not reflect the strategic choice of BELTEL for a new delivery approach, including the investments made in coaching and implementing tools that support an agile way of working. Eight (8) comments were positive about the quality of the product owner and the backlog management tool in use:

'Good team spirit, dedicated and committed product owner, smooth testing, implementation as per time, budget, and quality' (P38).

'Possibility to have the requirements with the highest priority delivered first. Business is satisfied' (P44).

However, some comments were related to the agile process itself that needed improvement:

'The whole agile process is still pretty blurry to most of its stakeholders so this definitely needs to be improved' (P48).

For an organization that made delivery of value a strategic innovation remarkably few comments were made on value aspects. Two were about good value being delivered, while most had to do with the lack of value:

'Don't have full view on commercial impact' (P21).

'No real feeling on the benefit of this project' (P45).

With regard to process aspects 7 comments were about needs for improvement:

'The process went too slow and it was not always clear who had which role and how that role would fit into the project' (P18).

'Could have been done in a more structured way' (P17).

However, about as many comments were related to a good process:

'Very lean way of working, high flexibility' (P44).

Observation 9: The low number of comments related to agile processes does not reflect the strategic choice of BELTEL for a new delivery approach.

5.3.6 Supplier Management (A11)

A number of comments (13) were about issues with suppliers, where also BELTEL's main supplier INDSUP was mentioned several times:

'Very long delays and complete lack of knowledge and initiatives from <name of supplier>' (P14).

5.3.7 Cost

Finally, a remarkable observation is that only once a comment is made related to cost of projects:

'Implementation as per time, budget, and quality' (P38).

No comments were made about the quality of estimations with regard to project cost.

Observation 10: Cost does not seem an important issue for stakeholders within BELTEL's project organization.

6. DISCUSSION

Four important learnings arise from summarizing the observations from the quantitative and qualitative analysis. First, quality of the deliverables (both good quality and to be improved quality) (observation 5), in combination with testing aspects and deployment into the production environment, is commonly mentioned in comments by all participants. Unclear requirements, bad documentation, requirements creep, and bad quality of test and deployment resources are perceived as causes for bad quality of deliverables (observation 7). Good teams, release-based working (repeated delivery by the same team), on-time delivery, and a smooth process are perceived as causes for good quality of deliverables (observation 8). However, a link with too long project durations for

smaller projects in size (observation 1) cannot be confirmed in the qualitative analysis.

Quantitative analysis shows a moderate relation between *stakeholder satisfaction* for both *process* and *result* on the one hand and *Estimation Quality Factor (Duration)* and *Cost Duration Index* on the other (observation 4 and 3). The comments given in the surveys confirm these observations with regard to project duration. Stakeholders of projects within BELTEL are satisfied when delivery of results is in-time, where we assume this relates to good quality of duration estimates (observation 8). However, it needs to be said that the word estimate or estimation are never used in the comments. On the other hand, dissatisfaction of stakeholders is often linked with too late delivery and long project durations (long waiting time). This finding does not hold for project cost. *Project cost* seems not an important issue for stakeholders (observation 10). Only one comment is made related to this.

A third finding is about communication (observation 6). Many of participants mentioned this to be of importance. Half of the comments were about good communication, such as good alignment between parties, good collaboration, and short feedback loops. The other half mention communication to be improved, such as provide information on processes and innovations (e.g. agile delivery), ongoing discussions, and miscommunication with suppliers. We assume that the relative absence of comments that are related to the ongoing innovation of implementing a more agile delivery process in combination with the very limited focus on value might be of importance here (observation 9). This low interest in agile innovation among BELTEL's stakeholders in a way reflects our findings in the quantitative analysis too. We found a relation between *perceived value* and *Functional Size* of software project deliverables, however this was only a weak correlation. Besides that, a moderate correlation was found between *perceived value* and *Cost per FP* and *Defects per FP* (observation 2). This indicates that *stakeholder satisfaction* is more related to interaction and being informed, than with conformance to planning and estimation. A strategy of 'no last minute surprises' is as such much better to increase *stakeholder satisfaction* than attempting to improve estimation and planning practices.

However, a warning is in place here: we notice that many positive comments on communication also are linked to two specific Product Owners within BELTEL. We did not focus our research on roles within the subject projects, but this suggests that the fulfillment of a role by a specific person may be of greater influence on *stakeholder satisfaction* and *perceived value* than the subject delivery model. Note that this resonates with the first line of the Agile Manifesto: "Individuals and Interactions over Processes and Tools" [12].

6.1 Implications

What can we do with these results? Our findings imply that companies such as BELTEL can improve their *stakeholder satisfaction* and *perceived value* by paying more attention to communication on the delivery approach (which is Scrum for BELTEL) and by ensuring good overall communication between all parties involved.

A second point of attention for BELTEL and others alike is to improve project documentation and overall requirements, and testing and deployment resources (e.g. environments, tools, process) since this might be of influence to satisfaction of project stakeholders.

6.2 Threats to Validity

With regard to construct validity constraints we emphasize that we asked stakeholders for perceptions on satisfaction and value.

Perceptions are not the same as actual measurements, this is especially the case for our value measurements. We prefer to measure the real business value as delivered by each software project. However, two problems occur with regard to this. Holistic measurements on value are often difficult to make for a single project (e.g. *Return on Investment* and *Net Present Value*). Besides that, such measures (e.g. *Net Promoter Score*) cannot easily be related to software projects, mainly because too many different factors are of influence for such measurements.

A threat to internal validity that we acknowledge is the fact that 'fishing for p-values' might hold a risk that some of the correlations we find are a coincidence. However, the number of parameters in our model is too low to perform a reliable generalized linear model test with multiple data points. To prevent from systematic error we perform an exploratory test in which we do test for p-values, yet we confront these with findings of the qualitative analysis.

In order to minimize systematic error with regard to subjectiveness of stakeholders in their survey answers, we included representatives from both IT and business that were involved in any way in a subject project. We considered to also include participants that did not know the subject projects in the assessment of perceived value. However, the study was performed in an operational context within BELTEL. Answering surveys, subsequent a release, was implemented as an operational capability. When designing the study we considered that it was undesirable to interfere stakeholders more than necessary in their operational activities, and not to engage them in surveys related to projects in which they did not participate. Another attempt we made to prevent from bias, was to perform anonymous surveys, although one can argue that based on specific roles a lack of anonymity could introduce potential bias.

In order to reduce bias due to ambiguity of survey answers with regard to the four aspects of value (customer, internal process, financial, and innovation) we applied additional text on the survey that was shown when participants hovered over a question mark linked to each question.

One other threat for our study is in the fact that the lead author of this paper is also a member of the measurement team within BELTEL. However, we prevent from bias because the BELTEL measurement team is independent and objective in its collection of data.

The extent to which the results of our study can be generalized to other companies than BELTEL is difficult to answer because we performed an exploratory study in one specific company. Especially because our findings relate to specific situations, maturity, and development approaches we argue that a one-on-one generalization to other companies is not valid. Instead we argue that evidence-based software engineering [13] in a way we perform for this study within BELTEL is a precondition for mature improvement within other companies too. In a way we argue that not our findings itself, but the method we use to collect and analyze project metrics might be of use for other software companies. As such, we recommend additional research on relations between project metrics and satisfaction and value within BELTEL but also within other software companies as we expect this to be of great interest for the software engineering industry.

7. RELATED WORK

Many studies include critical reviews of the Standish Chaos Report [3] [2] [14] [4] [15] [13] [16] [17]. The Standish Group reported in their 1994 CHAOS report that the average cost overrun of software projects was as high as 189%. Jørgensen and Moløkken-Østvold [2] conclude that this figure is probably much too high to represent

typical software projects in the 1990s and that a continued use of that figure as a reference point for estimation accuracy may lead to poor decision making and hinder progress in estimation practices [2]. Glass [3] states that objective research study findings do not, in general, support those Standish conclusions [3].

Although quite some research has been performed in the area of value estimation [18] [19] [20], and success criteria for software project [21] [22], most of these approaches seem poorly adopted in industrial software project management settings.

8. CONCLUSIONS

The outcomes of our exploratory study indicate that “within time and cost” does not automatically lead to satisfied stakeholders. A focus on shortening overall project duration, and good communication (e.g. no last minute surprises) has a positive effect on stakeholder satisfaction, while too late delivery and long project durations dissatisfies them. Our study does not provide any evidence that steering on costs helped to improve these. A novelty in the results of our study is that we linked perceived value to functional size of projects. However, we only found a weak positive relation between both metrics, where we expected them to be interconnected in a stronger way.

ACKNOWLEDGMENTS

We thank BELTEL for its generosity to allow us to use company data in our research, and all survey respondents for their help on sharing their ideas on improvement of software projects with us. Furthermore, we thank Georgios Gousios for his great help.

REFERENCES

- [1] Standish Group International, "The Chaos Report," 1994. www.standishgroup.com/sample_research/PDFpages/Chaos1994.pdf.
- [2] M. Jørgensen and K. Moløkken-Østfold, "How large are software cost overruns? A review of the 1994 CHAOS report," *Information and Software Technology*, vol. 48, no. 4, pp. 297-301, 2006.
- [3] R. L. Glass, "The Standish Report: Does It Really Describe a Software Crisis?," *Communications of the ACM*, vol. 49, no. 8, pp. 15-16, 2006.
- [4] J. L. Eveleens and C. Verhoef, "The rise and fall of the chaos report figures," *IEEE software*, vol. 27, no. 1, pp. 30-36, 2010.
- [5] H. Huijgens, R. v. Solingen and A. v. Deursen, "How to build a good practice software project portfolio?," in *ACM Companion Proceedings of the 36th International Conference on Software Engineering (ICSE)*, 2014.
- [6] H. Huijgens, G. Gousios and A. v. Deursen, "Pricing via Functional Size - A Case Study of a Company's Portfolio of 77 Outsourced Projects," in *ACM/IEEE 9th International Symposium on Empirical Software Engineering and Measurement (ESEM)*, Beijing, China, 2015.
- [7] T. DeMarco, "An algorithm for sizing software products," *ACM SIGMETRICS Performance Evaluation Review*, vol. 12, no. 2, pp. 13-22, 1984.
- [8] L. J. Eveleens and C. Verhoef, "Quantifying IT forecast quality," *Science of computer programming*, vol. 74, no. 11, pp. 934-988, 2009.
- [9] R. S. Kaplan and D. P. Norton, *Putting the balanced scorecard to work*, 1995.
- [10] H. Huijgens, A. v. Deursen and R. v. Solingen, "An Exploratory Study on the Effects of Perceived Value and Stakeholder Satisfaction on Software Projects - Technical Report - TUD-SERG-2016-012," Delft University of Technology, 2016.
- [11] B. Boehm, "Software Engineering Economics," *IEEE Transactions on Software Engineering*, vol. 10, no. 1, pp. 7-19, 1984.
- [12] Beck et al., "Manifesto for Agile Software Development," 2012. [Online]. Available: www.agilemanifesto.org.
- [13] T. Dyba, B. A. Kitchenham and M. Jørgensen, "Evidence-based software engineering for practitioners," *Software, IEEE*, vol. 22, no. 1, pp. 58-65, 2005.
- [14] K. Moløkken and M. Jørgensen, "A Review of Surveys on Software Effort Estimation," *IEEE - Proceedings of ISESE - International Symposium on Empirical Software Engineering*, pp. 223-230, 2003.
- [15] K. El Emam and A. Gunes Koru, "A replicated survey of IT software project failures," *IEEE Software*, vol. 25, no. 5, pp. 84-90, 2008.
- [16] R. L. Glass, "IT Failure Rates-70% or 10-15%?," *IEEE Software*, vol. 22, no. 3, pp. 110-112, 2005.
- [17] J. Sutherland, A. Viktorov, J. Blount and N. Puntikov, "Distributed Scrum: Agile Project Management with Outsourced Development Teams," in *40th International Conference on System Sciences*, Hawaii, 2007.
- [18] B. Boehm, "Value-Based Software Engineering," *ACM SIGSOFT Software Engineering Notes*, vol. 28, no. 2, pp. 1-12, 2003.
- [19] S. Biffl, A. Aurum, B. Boehm, H. Erdogmus and P. Grünbacher, *Value-Based Software Engineering*, Berlin Heidelberg: Springer, 2006.
- [20] S. Faulk, D. Harmon and D. Raffo, "Value-Based Software Engineering (VBSE): A Value-Driven Approach to Product-Line Engineering," in *First International Conference on Software Product-Line Engineering*, 2000.
- [21] N. Agarwal and U. Rathod, "Defining 'success' for software projects: An exploratory revelation," *International journal of project management*, vol. 24, no. 4, pp. 358-370, 2006.
- [22] D. J. Bryde, "Methods for managing different perspectives of project success," *British Journal of Management*, vol. 16, no. 2, pp. 119-131, 2005.

TUD-SERG-2016-015
ISSN 1872-5392

