The Effects of Perceived Value and Stakeholder Satisfaction on Software Project Impact

Hennie Huijgens, Arie van Deursen and Rini van Solingen

Report TUD-SERG-2017-001
The Effects of Perceived Value and Stakeholder Satisfaction on Software Project Impact

Hennie Huijgens\textsuperscript{a, 1}, Arie van Deursen\textsuperscript{b}, Rini van Solingen\textsuperscript{c}

\textsuperscript{a}Delft University of Technology and Goverdson, Mekelweg 4, 2628 CD Delft, The Netherlands
\textsuperscript{b}Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands
\textsuperscript{c}Delft University of Technology and Prowareness, Mekelweg 4, 2628 CD Delft, The Netherlands

ARTICLE INFO
Keywords:
Software Economics
Evidence-Based Software Engineering
Perceived Value
Stakeholder Satisfaction
Cost Duration Index

ABSTRACT
Context: In this paper we present a multiple case study on the insights of software organizations into stakeholder satisfaction and (perceived) value of their software projects. Our study is based on the notion that quantifying and qualifying project size, cost, duration, defects, and estimation accuracy needs to be done in relation with stakeholder satisfaction and perceived value.

Objectives: We contrast project metrics such as cost, duration, number of defects and estimation accuracy with stakeholder satisfaction and perceived value.

Method: In order to find out whether our approach is practically feasible in an industrial setting, we performed two case studies; one in a Belgian telecom company and the other in a Dutch software company.

Results: In this study we evaluate 22 software projects that were delivered during one release in the Belgian telecom company, and 4 additional large software releases (representing an extension of 174% in project size) that were delivered in a Dutch software company. Eighty-three (83) key stakeholders of two companies provide stakeholder satisfaction and perceived value measurements in 133 completed surveys.

Conclusions: We conclude that a focus on shortening overall project duration, and improving communication and team collaboration on intermediate progress is likely to have a positive impact on stakeholder satisfaction and perceived value. Our study does not provide any evidence that steering on costs helped to improve these. As an answer to our research question - how do stakeholder satisfaction and perceived value relate to cost, duration, defects, size and estimation accuracy of software projects? – we found five take-away-messages.

1. INTRODUCTION

An often cited result of the 1994 Standish CHAOS research \cite{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 \cite{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 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?

1.1 Problem Statement

Supported by many critical reviews of the Standish criteria \cite{2} \cite{3} \cite{4}, we define success and failure in this paper from a different angle, trying to include the balance between value and cost into the equation. In previous research we defined success and failure in terms of cost, duration and number of defects of
a software project [5] [6] [7]. Looking at the outcomes of this we consider that a project that is late and over budget – and thus in terms of our study bad practice, or in other words unsuccessful – yet returns high value according to its stakeholders, may still be called successful, because of the fact that it delivers high value.

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, we focus in this paper specifically on perceived value. The underlying idea is that, since finding evidence in the bottom-line financial administration is hard, if not impossible, the best we can do is involve stakeholders for a qualitative indication of value. However, as this is strongly dependent on the individual and the contextual setting (what is valuable in one setting might not be valuable in another, or what one stakeholder considers to be of no value can be of high value to another stakeholder), we use the term Perceived Value. We understand that this is a way to measure value that is limited in its external and construct validity. However, this approach may help in finding early ways of indicating value [8].

In this paper, we analyze a set of projects conducted at a Belgian telecom company (referred to in this paper as BELTEL) and a Dutch software company (referred to in this paper as DUTCHCO) that provides billing software products and services (also largely to the telecom domain). We propose the following research question:

How do stakeholder satisfaction and perceived value relate to cost, duration, defects, size and estimation accuracy 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 26 industrial projects for which 83 key stakeholders provide stakeholder satisfaction and perceived value measurements in 133 completed surveys.
3. We contrast the resulting perceived value and stakeholder satisfaction statements with collected data on costs, duration, defects, size and estimation accuracy and look for links between them.

This paper is an extended journal version of an earlier published paper at the 20th International Conference on Evaluation and Assessment in Software Engineering (EASE 2016) [6]. Compared to the original paper the new contributions can be summarized as follows:

- We replicated the research performed in our original study in another company: DUTCHCO, a Dutch software company, specialized in delivering billing solutions to European telecom operators.
- Within DUTCHCO, we examined four (4) large software releases, representing an extension of 174% in project size. We collected detailed size, cost, time, and defects data from all releases. We performed electronic surveys on stakeholder satisfaction and perceived value among thirty (30) stakeholders within the DUTCHCO organization and INVEND, its provider of India-based development teams.

The remainder of this paper is structured as follows. In Section 2 related work and the background of the model that we use for analysis purposes are described. Section 3 outlines the research design. The results of the study are described in Section 4. We discuss the results in Section 5, and finally, in Section 6 we make conclusions and outline future work.

2. BACKGROUND AND RELATED WORK

Many studies include critical reviews of the Standish Chaos Report [2] [3] [4] [9] [10] [11] [12] [13]. The Standish Group reported in their 1994 CHAOS report that the average cost overrun of software projects was as high as 189%. Jørgensen and Molokken-Ostvold [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]. Gall [3] states that objective research study findings do not, in general, support those Standish conclusions [3].

Where in our research we measure value as perceived by stakeholders on four business related subjects, many different measures are used to identify value, and a clear and uniform definition is no question yet. Pekki [14] defines stakeholder value as the “usefulness of offering SPI to its key beneficiaries, so they are fully involved into SPI activities which increases the success of those activities”. Beck [15] indicates that value is about money and time, by saying we “need to make our software economically more valuable by spending money more slowly, earning revenue more quickly and increasing the probably productive lifespan of our project”. Dingsøyr and Lassenius [16] answer the question “What is value?” by saying that “the improvement trends are not specific on how they define value”. They come up with the argument that, proponents of agile development would argue that a development team needs to learn what external stakeholders value during a development project. In a way this matches our idea that besides internal stakeholders, especially external stakeholders should be involved in the value discussion.

Atkinson [17] argues that besides time, cost and quality, often referred at as the iron triangle, also stakeholder benefits should be taken into the equation. Besides that, he mentions the effect that quality is “an emergent property of people’s different attitudes and beliefs, which often change over the development life-cycle of a project”. Estimate the value of software is probably as challenging as predicting the cost of software [18]. Strand and Karlson [19] suggested to estimate value in the form of “benefit points”, as a kind of equivalent to story points. Cheng et al. [20] describe an architecture-based approach to discover value of software
engineering by using big data techniques. Although quite some research has been performed in the area of value estimation [21] [22] [23], and success criteria for software projects [24] [25], most of these approaches seem poorly adopted in industrial software project management settings. A good sign however, is that an increased focus on value improvement is seen in software development, mainly driven by agile development approaches [16].

Jørgensen [26] performed a survey among software professionals in Norway on the characteristics of projects with success in delivering client benefits. He mentions that a focus on client benefits as a success criterion is particularly important, because only weak correlations are found on other dimensions, such as “being on time” and “being on budget”. Besides that, he mentions that the traditional success factor “having the specified functionality” may even be in conflict with success in delivering client benefits.

3. RESEARCH DESIGN

The goal of this study is to understand the underlying reasons of stakeholder satisfaction and value of software projects. To achieve this, we contrast project metrics such as cost, duration, number of defects and estimation accuracy 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 explained in the introduction, the Standish criteria [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.

We performed a multiple case study in two different companies: BELTEL, a Belgian telecom company, and DUTCHCO, a Dutch software company that delivers billing solutions to European telecom companies. In the following two paragraphs we describe the industrial context of how both companies are included in our research.

3.1 BELTEL

BELTEL is a Belgian telecom company that can be characterized as a typical mid-sized information-intensive company with a mature software delivery organization that offers a mix of delivery approaches, ranging from plan-driven to agile (Scrum) [27]. 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 estimation accuracy of finalized software projects. This data has been used to analyze project performance at BELTEL, to benchmark project performance, and to continuously improve the software delivery process within BELTEL. In October 2015, 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 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 are 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 study data from 22 projects coming from four different releases.

3.2 DUTCHCO

DUTCHCO is a Netherlands based software company that offers billing solutions to a large variety of European telecom companies. Within this market DUTCHCO is a European market leader.

Unlike BELTEL, DUTCHCO does not structure its work into projects. All software development activities are organized into four large market releases each year. Driven by the desire of its customers to limit the number of deployments, DUTCHCO implements only four market releases a year. As a result, these four releases are usually quite large in size. Where BELTEL thus implements eight releases a year, each of which consist of a large number of small and medium-sized projects, DUTCHCO performs only four large releases, which are composed of many small user stories.

To build and test its software, DUTCHCO makes use of several development teams in India [28]. These teams are supplied and supported by INVEND, a Dutch consultancy company, specializing in agile software delivery. Activities such as preparation of releases, design, quality assurance, and overall management are performed by members of an onsite, Netherlands based team of DUTCHCO itself.

Based on the results of previous research within the organization, DUTCHCO pays considerable attention to communication between the different members of a development team. There is a virtual contact window that is constantly open to allow team members in different locations to contact colleagues, and substantial effort is put into reciprocal visits to the team sites.

All teams within DUTCHCO – including the development teams of INVEND in India – work according to the Scrum approach [27]. An enterprise backlog and sprint backlogs are maintained in Jira, bi-weekly sprints are performed, results are demonstrated to business stakeholders, and two-weekly retrospectives are performed. As such the DUTCHCO market releases contain a combination of about 6 to 7 (bi-weekly) Sprint deliveries. As these Sprints do deliver working tested software, one could also call these deliveries. However, these are only deployed in an acceptance test environment and not to
the market, we use the term ‘market release’ for those four releases each year.

The DUTCHCO teams are organized in a component-based way. One database-team (DB) is based in The Netherlands. Two teams are based in India: one portal and asset management-team of nine people (POR and AM), and one reporting-team (AR) of also nine people.

Table 1 summarizes the release approaches of both companies. BELLTELE runs single projects that are combined eight times per year for user acceptance testing and deployment. In the DUTCHCO case no projects are to be found; user stories are combined in releases that are deployed every three months.

<table>
<thead>
<tr>
<th>Table 1. Summary of Release Approaches</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BELTEL (N = 4)</strong></td>
</tr>
<tr>
<td><strong>Frequency of Release</strong></td>
</tr>
<tr>
<td><strong>Scope of Release</strong></td>
</tr>
<tr>
<td><strong>Average Size of Release</strong></td>
</tr>
<tr>
<td><strong>Average Cost of Release</strong></td>
</tr>
</tbody>
</table>

3.3 Challenges in Comparing both Companies

Looking at the large differences in the project size, staff count, budgets, geographic location of team and customer demands, we recognize major challenges in comparing software projects performed in a telecom company with a software company. To remedy this, we used a tool that we designed to address this challenge [29]. In previous research we built a model, the so-called Cost Duration Matrix (see Figure 2), 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] [7]. The model takes a project’s size, measured in function points (FPs) [30], as starting point and as a source for normalization that makes it possible to compare software projects with different settings. The model compares the actual costs normalized to a function point (in Euros per FP) and duration (in days per FP) for a project of this size to benchmarked data, taken from a set of 492 finalized software projects in the financial and telecom application domains. This is done using two power regressions conducted on the 492 projects, permitting the computation of the ‘expected’ cost and duration of a project of a given size (measured in function points) [5] [7].

3.4 Metrics

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

3.4.1 Project Metrics

Four project metrics are collected on each project that is subject of the case study: Project Cost (in Euros), Project Duration (in months), and the Number of Defects found during the project. Project Size is measured in function points, according to the IFPUG industry standard [30]. Based on this, we determine the Cost per Function Point, Days per Function Point, and Defects per Function Point, using in each case the size in function points as weighting factor.

3.4.2 Estimation Quality Factor

The Estimation Quality Factor (EQF) is a measure of the deviation of a forecast to the actual cost or duration. EQF is a forecasting metric that depicts the quality of forecasts made during a project. The measure was defined by DeMarco [31]. He defines EQF by:

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

We use the formalization proposed by Eveleens and Verhoeef [32]. We reiterate and correct the definition given there. Let \( a \) be the actual value (\( a > 0 \)), \( t_0 \) the time the actual is known and \( e(t) \) the value of the forecast at time \( t \) (\( 0 \leq t \leq t_a \)) in the project. Then, the EQF is represented by [32]:

\[
EQF = \frac{\int_0^{t_a} a \, dt}{\int_0^{t_a} |a - e(t)| \, dt} = \frac{\int_0^{t_a} 1 \, dt}{\int_0^{t_a} (1 - e(t)) / a \, dt}.
\]

EQF allows us to quantify the quality of forecasts. A low EQF value means that the deviation of the forecasts to the actual cost or duration is large. EQF is measured for both cost and duration.

3.4.3 Cost Duration Index

The Cost Duration Index is a measure of the relative position of a project within the Cost Duration Matrix (see Figure 2). 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:

\[
p = \sqrt{\frac{\text{Actual Duration}}{\text{Benchmark Duration}} \times \frac{\text{Actual Cost}}{\text{Benchmark Cost}}}
\]

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

\[
\text{Cost Duration Index} = \left( \frac{p_{\text{max}} - p}{p_{\text{max}}} \right) \times 100
\]
3.4.4 Stakeholder Satisfaction

Human satisfaction is a complex concept, involving many components such as physical, emotional, mental, social, and cultural factors [33] [34]. From behavioral science and consumerism multiple theories have emerged on psychometrically validated surveys on satisfaction (e.g. [35] [36]). Although extended handbooks are available on the setup of a satisfaction survey [37], we opted for a lean survey setup. The main reason to do so was a requirement from BELTEL executives to make the survey as short as possible in order to minimize disturbance to the daily work for employees. An important argument for this requirement was the fact that the survey was implemented as an integral part within the release process, meaning that some staff members had to fill it out several times during the release process (e.g. release managers that involved in more projects that were included in one release filled out a separate survey for each individual project), or within every release (e.g. team members of Scrum-teams). We assume that this light-weight requirement will apply for other companies too and therefore is a precondition for a successful metric.

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.

In both BELTEL and DUTCHCO surveys were answered by internal stakeholders of projects; e.g. project managers, developers, testers, product owners. In case external stakeholders were included, these were working for BELTEL or DUTCHCO as client or business analyst for a specific project. No external stakeholders in the meaning of end-users of a projects’ deliverables were involved in the surveys.

3.4.5 Perceived Value

Value of software projects is a complex metric to measure [18], and studies are not specific on how they define value [16]. It is difficult, if not impossible, to measure objectively and indisputable the real value as delivered by software projects to customers of BELTEL and DUTCHCO. Is real value about money and time? Does it mean financial value, as in studies indicated by Return Of Investment (ROI) [38]? Or is real value measured by Net Promotor Score (NPS), as other studies indicate [39] [40] [41]? Such holistic measurements on value are often difficult to make for a single project, and they cannot easily be related to single software projects, mainly because too many different factors are of influence for such measurements.

To approximate the real value, we measure Perceived Value as a qualitative measure of the perception of stakeholders of each project. This is based on the notion that in fact every measurement is an agreement on a measurement procedure that sufficiently approaches the actual value [38].

Perceived Value is measured for each stakeholder in a specific project, on four aspects: BELTEL’s or DUTCHCO’s customers, BELTEL or DUTCHCO’s financials, BELTEL or DUTCHCO’s internal process effectiveness, and BELTEL or DUTCHCO’s innovation. We base the use of the four perspectives Customer, Financial, Internal Process, and Innovation on the Balanced Scorecard [42]. Based on the results we developed a metric Perceived Value (overall) is calculated, with the number of measures (not counting the choice “Don’t know”) as weighting factor.

3.5 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 actual cost and duration of projects. Since 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.

3.6 Data Collection procedure

3.6.1 Collection of quantitative data

Within BELTEL, a major part of the data collection for our case study was performed within the measurement capability that was already operational within the software department of the organization. Data collection on Project Cost, Project Duration, Number of Defects, Project 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 [30]) by measurement staff of BELTEL’s main Indian supplier.

Different artifacts were used as a source for function point counting, depending from the availability per project (e.g. sets of functional documentation, user stories recorded in one of the Scrum backlog tools, architectural documents, project documentation, user manuals, or wireframes). All project data was stored in a measurement repository that was provided for our study. The lead author of the study was part of the BELTEL’s measurement team.

In the DUTCHCO case, a dedicated research project was performed in order to collect and analyze data of software releases. The lead author of this paper performed the size calculations in retrospect for DUTCHCO. Due to this, it was possible to replicate the study that we performed within BELTEL in exactly the same way in the DUTCHCO organization. All quantitative data was defined and collected in the same way. Function points were counted according to the same counting rules as used within BELTEL [30]. As a source for function point counting the user stories as recorded in the Scrum backlog tool were used.

Driven by the observations in our original study on correlations between Project Cost and Number of Defects on one hand, and Stakeholder Satisfaction on the other, we decided to collect data from finalized software releases within DUTCHCO in a more detailed way: cost data was categorized into a limited number of cost categories (e.g. design, build, test,
deploy, management overall, quality assurance), and defect data was collected per defect severity (e.g. blocking, critical, high, medium, low).

3.6.2 Collection of qualitative (survey) data

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 with stakeholders from BELTEL, and later from DUTCHCO. 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);
4. 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);
5. 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 with a mouse pointer over a question mark next to the text of each of the four aspects.

Within DUTCHCO we applied the same electronic survey for stakeholders of the finalized software releases, including team members from the INVEND teams located in India.

3.7 Analysis Procedure

To explore potential relationships between the collected metrics, we tested for association between paired samples. Because all sample data is not normally distributed (see Table 3 for details on skewness and kurtosis and the boxplots in Figure 1), we used a Spearman rank 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.

Following Hopkins [43], we prevent from Type I errors, e.g. finding a correlation by chance, simply because multiple comparisons are performed on the same dataset, by performing Benjamini-Hochberg corrections on all p-values. We used an alpha of 0.05/26 (the number of projects in scope of this study), meaning that we assume all p-values above 0.0019 as not significant [43]. We consider a significant correlation higher than 0.3 (or lower than −0.3) to be moderate, a significant correlation score higher than 0.5 (or lower than −0.5) to be strong, and a significant correlation above 0.9 (or lower than -0.9) to be very strong.

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 that were performed within BELTEL and DUTCHCO. We use the tool Qualyzer2 for this purpose. We applied open coding, breaking down the survey data into first level concepts and second-level categories. Coding was performed by the first author of the study, and reviewed by the other authors.

4. RESULTS

4.1 Description of the BELTEL Projects

Within the scope of our study we evaluated four software releases within BELTEL, covering a total of 22 software projects. Table 2 gives a brief description of each project, where the numbering of the projects indicates in which release each project was finalized and in which company a project or release was performed (e.g. BELTEL 6.4 is a BELTEL project that 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, however a formal Scrum-by-the-book approach was not in place (i.e. sprints where performed, a backlog was managed and prioritized in a backlog tool, a product owner was in place, however no retrospectives were performed, no Scrum master was in place).

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.

2 http://qualyzer.bitbucket.org
Table 2. The BELTEL projects in scope of the case study.

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

Table 3. Descriptive statistics of the BELTEL project data.

<table>
<thead>
<tr>
<th>Project Data (n = 22)</th>
<th>Cost</th>
<th>Duration</th>
<th>Project Cost</th>
<th>Project Size</th>
<th>Number of Defects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost Index</td>
<td>86.92</td>
<td>8,000</td>
<td>4.96</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>First Quartile</td>
<td>93.27</td>
<td>44,001</td>
<td>8.37</td>
<td>25</td>
<td>3</td>
</tr>
<tr>
<td>Median</td>
<td>97.28</td>
<td>66,209</td>
<td>10.18</td>
<td>39</td>
<td>9</td>
</tr>
<tr>
<td>Third Quartile</td>
<td>98.57</td>
<td>118,876</td>
<td>11.73</td>
<td>126</td>
<td>23</td>
</tr>
<tr>
<td>Maximum</td>
<td>99.78</td>
<td>296,000</td>
<td>19.03</td>
<td>324</td>
<td>223</td>
</tr>
<tr>
<td>Mean</td>
<td>95.90</td>
<td>99,615</td>
<td>10.20</td>
<td>79</td>
<td>29</td>
</tr>
<tr>
<td>Skewness</td>
<td>-1.03</td>
<td>1.27</td>
<td>0.78</td>
<td>1.71</td>
<td>3.19</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>0.06</td>
<td>0.77</td>
<td>1.43</td>
<td>2.71</td>
<td>10.89</td>
</tr>
<tr>
<td>St. Deviation</td>
<td>3.51</td>
<td>78,209</td>
<td>3.22</td>
<td>82</td>
<td>55</td>
</tr>
</tbody>
</table>

Table 4. Results from a Wilcoxon rank sum comparison of BELTEL releases (n = 22) with peer groups (n = 492).

<table>
<thead>
<tr>
<th>Project Data</th>
<th>Median</th>
<th>Median peer group</th>
<th>W</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Size</td>
<td>39</td>
<td>116</td>
<td>7148</td>
<td>0.0108</td>
</tr>
<tr>
<td>Project Cost</td>
<td>66,209</td>
<td>278,156</td>
<td>8003</td>
<td>0.0001</td>
</tr>
<tr>
<td>Project Duration</td>
<td>10.18</td>
<td>8.41</td>
<td>4008</td>
<td>0.0394</td>
</tr>
<tr>
<td>Number of Defects</td>
<td>9</td>
<td>72</td>
<td>2989</td>
<td>0.1028</td>
</tr>
<tr>
<td>Cost per FP</td>
<td>1612</td>
<td>2520</td>
<td>6952</td>
<td>0.0239</td>
</tr>
<tr>
<td>Days per FP</td>
<td>7.85</td>
<td>2.08</td>
<td>3058</td>
<td>0.0006</td>
</tr>
<tr>
<td>Defects per FP</td>
<td>0.22</td>
<td>0.16</td>
<td>2280</td>
<td>0.6621</td>
</tr>
</tbody>
</table>

If the data were sampled from a population with the median of the research repository, one would expect the sum of signed ranks (in the table reported as W) to be relatively small. The comparison shows that BELTEL significantly differs from the other projects in the repository on Project Cost, as well on Days per FP. On all other metrics no significance was found in the test. With regard to Project Cost we see this effect also in the boxplot in Figure 1; BELTEL clearly shows overall lower cost for its projects compared to the other companies in our research repository. Although not confirmed by the statistical tests, a similar effect can be seen for Project Size; the boxplots in Figure 1 indicate that overall size for BELTEL projects is smaller than that of the other companies. An explanation for differences in the outcomes of statistical tests and the boxplots in Figure 1 might be that the first only includes the 22 BELTEL projects that are in scope of this study, while the second includes all 157 BELTEL projects from our research repository.

Besides project metrics, we collected data of the BELTEL projects on Stakeholder Satisfaction and Perceived Value by sending an online questionnaire to applicable stakeholders of each software project once the technical go live was performed. The overall completion rate of all surveys within BELTEL 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.

4.2 Description of the DUTCHCO projects

Within the scope of our study we examined four DUTCHCO releases, all built from a large number of user stories. Table 5 gives a brief description of each release, where the numbering of the releases indicates in which company a release was performed; e.g. DUTCHCO 5.1 AM is a DUTCHCO release that was applicable to the asset management (AM) component of their billing solution.

Table 5. The DUTCHCO projects in scope of the case study.

<table>
<thead>
<tr>
<th>Project ID</th>
<th>Project Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DutchCo 5.1 AM</td>
<td>Release containing asset management (AM) user stories.</td>
</tr>
<tr>
<td>DutchCo 5.1 FOR</td>
<td>Release containing portal (FOR) user stories.</td>
</tr>
<tr>
<td>DutchCo 5.2</td>
<td>Release applicable on DutchCo’s general billing solution.</td>
</tr>
<tr>
<td>DutchCo 5.3</td>
<td>Release applicable on user stories for customer VFBK.</td>
</tr>
</tbody>
</table>

All project data of DutchCo is to be found in a Technical Report [73].
Unlike BELTEL, where the software portfolio includes a mix of projects of various business domains, delivery models and governance structures, the portfolio of DUTCHCO is more heterogeneous in nature. DUTCHCO implements only four releases each year to its customers. Due to that, these releases are usually quite large in size. All DUTCHCO releases relate to the same business domain, namely the billing solution it provides to its customers. However, it occurs that different sets of functionality are delivered to customers, due to differences in requirements.

DUTCHCO’s user stories are maintained in its backlog management tool, and continuously bundled in sprint backlogs. As a result, the governance structure of DUTCHCO is relatively simple. There are no projects, and there is a limited budget and planning activity. DUTCHCO has adopted a Scrum approach. Scrum teams are organized by functional component (e.g. Portal, Asset Management, Reporting, and Database). A large part of the Scrum teams is working from India, managed by INVEND.

Table 6 gives an overview of the descriptive statistics of the four DUTCHCO releases involved in this extended case study. The DUTCHCO subset exists of four releases, two relatively smaller ones (although still as large as BELTEL’s largest projects), and two large ones. As can be seen in Figure 2 and in Table 6 above, the two oldest DUTCHCO projects (DutchCo 5.1 - POR and DutchCo 5.1 – AM), are smallest in size (resp. 277 and 335 FPs). However, both are comparable to the largest projects from the BELTEL case. The two newest DUTCHCO releases are relatively large, compared to the projects in the Table 7. Results from a Wilcoxon rank sum comparison of DUTCHCO releases (n = 4) with peer groups (n = 492).

<table>
<thead>
<tr>
<th>Project Data (n=4)</th>
<th>Cost</th>
<th>Project Duration</th>
<th>Project Cost</th>
<th>Number of Defects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Index</td>
<td>(EUR)</td>
<td>(Months)</td>
<td>(FPs)</td>
</tr>
<tr>
<td>Minimum</td>
<td>95.03</td>
<td>125,827</td>
<td>3.78</td>
<td>277</td>
</tr>
<tr>
<td>First Quartile</td>
<td>96.12</td>
<td>192,040</td>
<td>5.90</td>
<td>321</td>
</tr>
<tr>
<td>Median</td>
<td>96.81</td>
<td>514,486</td>
<td>7.71</td>
<td>784</td>
</tr>
<tr>
<td>Third Quartile</td>
<td>97.41</td>
<td>835,343</td>
<td>9.18</td>
<td>1240</td>
</tr>
<tr>
<td>Maximum</td>
<td>98.22</td>
<td>896,788</td>
<td>10.29</td>
<td>1261</td>
</tr>
<tr>
<td>Mean</td>
<td>96.72</td>
<td>512,987</td>
<td>7.37</td>
<td>777</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.147</td>
<td>-0.002</td>
<td>-0.209</td>
<td>-0.003</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>-1.952</td>
<td>-2.404</td>
<td>-2.054</td>
<td>-2.432</td>
</tr>
<tr>
<td>St. Deviation</td>
<td>1.33</td>
<td>399,020</td>
<td>2.832</td>
<td>543</td>
</tr>
</tbody>
</table>

Table 7 gives an overview of the descriptive statistics of the four DUTCHCO releases involved in this extended case study. The DUTCHCO subset exists of four releases, two relatively smaller ones (although still as large as BELTEL’s largest projects), and two large ones. As can be seen in Figure 2 and in Table 6 above, the two oldest DUTCHCO projects (DutchCo 5.1 - POR and DutchCo 5.1 – AM), are smallest in size (resp. 277 and 335 FPs). However, both are comparable to the largest projects from the BELTEL case. The two newest DUTCHCO releases are relatively large, compared to the projects in the table.
BELTEL subset: DutchCo 5.2 is 1233 FPs in size, and DutchCo 5.3 is 1261 FPs in size. Apparently driven by a schedule of four deployments per year, combined with a tendency to bundle the user stories of all functional components of its system, DUTCHCO releases tend to grow relatively large.

We performed a Wilcoxon rank sum comparison with Bonferroni corrections between the DUTCHCO subset and our research repository as a whole, holding data of 492 software projects from four different companies, to compare overall differences, and differences per size (see Table 7 and Figure 1). The comparison shows that DUTCHCO significantly differs from the other projects in the repository on Project Size, as well on Cost per FP and Days per FP. On all other metrics no significance was found in the test.

We observe two findings here. Firstly, DUTCHCO releases have on average a larger size than other projects in our repository, which is good. This leads to a positive effect, from benchmarking purposes: due to the larger size of DUTCHCO releases also Cost per FP and Days per FP are better than the values of the other companies in our research repository.

Although no statistical evidence is found for any differences between the Number of Defects of both distributions (see Table 7), the boxplot view in Figure 1 indicates that besides Project Size, DUTCHCO also deviates from its peer groups on Number of Defects. Based on this we assume that a good score on Project Size might be counterbalanced here by a bad score on Number of Defects.

Besides project metrics as described above, we collected data on Stakeholder Satisfaction and Perceived Value by sending an online survey to applicable stakeholders of each software release once the technical go live was performed. For this purpose, we used the same electronic survey that was used before within BELTEL. The overall completion rate of all surveys within DUTCHCO was 71%. Thirty (30) surveys were completed by 30 individual respondents of both DUTCHCO and INVEND. Due to the fact that the three first releases were measured relatively long after finalization of each release, only for the latest DUTCHCO release an electronic survey was performed.

4.3 Results of plotting on the Cost Duration Matrix

We used the model that we developed in previous research to compare a portfolio of projects to the benchmark, by means of a Cost Duration Matrix [5] [7], as shown in Figure 2 for the 26 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 take twice as long as expected from the benchmark.

As can be seen from the figure, most of the 26 projects in the portfolio are cheaper than 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:
Table 8. Matrix with test results of association between paired samples, using Spearman's rank correlation coefficient.

<table>
<thead>
<tr>
<th></th>
<th>Project Size</th>
<th>Project Cost</th>
<th>Project Duration</th>
<th>Number of Defects</th>
<th>Cost per FP</th>
<th>Days per FP</th>
<th>Defects per FP</th>
<th>Cost Duration Index</th>
<th>Stakeholder Satisfaction (Process)</th>
<th>Stakeholder Satisfaction (Result)</th>
<th>perceived Value (Proc. Value)</th>
<th>Perceived Value (Customer)</th>
<th>Perceived Value (Overall)</th>
<th>Perceived Value (Financial)</th>
<th>Estimation Quality Factor (Cost)</th>
<th>Estimation Quality Factor (Duration)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Cost</td>
<td>0.31</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project Duration</td>
<td>-0.13</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Defects</td>
<td>-0.08</td>
<td>0.000</td>
<td></td>
<td>-0.22</td>
<td>0.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost per FP</td>
<td>-0.70</td>
<td>0.000</td>
<td></td>
<td>0.87</td>
<td>0.35</td>
<td>-0.36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Days per FP</td>
<td>-0.34</td>
<td>0.000</td>
<td></td>
<td>-0.38</td>
<td>0.27</td>
<td>0.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defects per FP</td>
<td>-0.14</td>
<td>0.000</td>
<td></td>
<td>-0.50</td>
<td>0.21</td>
<td>0.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost Duration Index</td>
<td>0.09</td>
<td>0.000</td>
<td></td>
<td>0.00</td>
<td>0.10</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stakeholder Satisfaction (Process)</td>
<td>0.02</td>
<td>(0.93)</td>
<td></td>
<td>0.26</td>
<td>0.44</td>
<td>0.38</td>
<td>0.36</td>
<td>0.21</td>
<td>0.34</td>
<td>0.50</td>
<td>0.04</td>
<td>0.15</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stakeholder Satisfaction (Result)</td>
<td>-0.15</td>
<td>(0.09)</td>
<td></td>
<td>-0.16</td>
<td>0.47</td>
<td>-0.07</td>
<td>-0.27</td>
<td>-0.10</td>
<td>-0.23</td>
<td>-0.04</td>
<td>0.72</td>
<td>0.04</td>
<td>0.97</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived Value (Overall)</td>
<td>0.32</td>
<td>0.000</td>
<td></td>
<td>0.07</td>
<td>0.06</td>
<td>0.35</td>
<td>0.13</td>
<td>0.10</td>
<td>0.20</td>
<td>0.04</td>
<td>0.77</td>
<td>0.80</td>
<td>0.99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived Value (Customer)</td>
<td>0.34</td>
<td>(0.13)</td>
<td></td>
<td>0.10</td>
<td>0.06</td>
<td>-0.35</td>
<td>0.06</td>
<td>0.20</td>
<td>0.37</td>
<td>0.20</td>
<td>0.71</td>
<td>0.30</td>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived Value (Process)</td>
<td>0.26</td>
<td>0.000</td>
<td></td>
<td>-0.03</td>
<td>0.06</td>
<td>0.27</td>
<td>-0.15</td>
<td>0.01</td>
<td>0.60</td>
<td>0.20</td>
<td>0.58</td>
<td>0.35</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived Value (Financial)</td>
<td>0.32</td>
<td>(0.15)</td>
<td></td>
<td>-0.07</td>
<td>0.06</td>
<td>-0.35</td>
<td>0.13</td>
<td>-0.20</td>
<td>0.44</td>
<td>0.20</td>
<td>0.85</td>
<td>0.10</td>
<td>0.99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived Value (Innovation)</td>
<td>0.34</td>
<td>(0.13)</td>
<td></td>
<td>-0.09</td>
<td>0.06</td>
<td>-0.35</td>
<td>0.06</td>
<td>-0.20</td>
<td>0.43</td>
<td>0.14</td>
<td>0.85</td>
<td>0.20</td>
<td>0.99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimation Quality Factor (Cost)</td>
<td>-0.05</td>
<td>(0.06)</td>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimation Quality Factor (Duration)</td>
<td>-0.36</td>
<td>(0.10)</td>
<td></td>
<td>-0.43</td>
<td>-0.30</td>
<td>-0.18</td>
<td>-0.23</td>
<td>0.00</td>
<td>0.25</td>
<td>0.00</td>
<td>0.12</td>
<td>0.22</td>
<td>0.23</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The table above shows results from a test of association between paired samples of the 26 software projects from both case studies, using Spearman's rank correlation coefficient. Due to the fact that the DUTCHCO case study contained a limited number of four projects, we performed the association tests over the total set of 26 projects from both BELTEL and DUTCHCO. The overview shows for each test the correlation coefficient and between brackets the p-value. We corrected the problem of multiple comparisons by performing a Benjamini-Hochberg correction [69][70][71]. We compared each individual p-value to its Benjamini-Hochberg critical value, where i is the rank, m is the total number of tests, and Q is the false discovery rate. We used 0.10 as false discovery rate, according to [68]. The largest p-value that has p < (i/m)Q is significant, and all of the p-values smaller than it are also significant, even the ones that aren't less than their Benjamini-Hochberg critical value. A color indicates samples that are correlated: dark grey indicates a very strong (positive or negative) linear relationship (correlation coefficient higher than 0.70), moderate grey indicates a strong linear relationship (correlation coefficient between 0.50 and 0.70). Significant samples with a correlation coefficient lower than 0.50 are indicated in light grey.

1. **Good Practice** (top right): projects that score better than average for both cost and duration. In Figure 2, there are six projects in this quadrant, of which three of BELTEL (5, 3, 4, and 3.5) and three of DUTCHCO (5.1 - AM, 5.2, and 5.3).

2. **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 2.

3. **Bad Practice** (bottom left): projects that score worse than average for both cost and duration. In Figure 2, there are four projects in this quadrant, all from BELTEL.

4. **Time over Cost** (top left): projects that score better than average for duration, yet worse than average for cost. In Figure 2, there are no projects in this quadrant.

The overall performance of the portfolio is furthermore summarized through the two red 'median' lines: On average, projects in the subject portfolio take 34% more time than expected from the benchmark, yet are 51% cheaper. The Cost Duration Matrix provides a tool to compare two project portfolios in terms of Project Cost and Project Duration. Our comparisons are based on the benchmark of 492 projects from the finance and telecom industries, described in more detail in [5][7]. The benchmark of 492 projects contain 157 previous
projects from BELTEL, and 4 previous ones from DUTCHCO, making it a suitable benchmark to compare the new additional 26 projects against.

4.4 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 Spearman rank correlation coefficient. Because for only one DUTCHCO release data on Stakeholder Satisfaction and Perceived Value was measured (for only the latest release a survey was performed), we decided to test for associations on the BELTEL and DUTCHCO dataset as a whole. The results of these tests are shown in Table 8. 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. Results of the tests for association on the BELTEL projects only can be found in Table 3 of the original research paper [6].

However, a remark on the way we interpreted the results in Table 8 is in place. If results relating to the previous BELTEL analysis agree with the results including the DUTCHCO data, we assume that both organizations are exhibiting similar results. If the results are completely different when the DUTCHCO results are included, we conclude that the companies are behaving differently, and that further research is needed to establish whether the new combined results are valid.

A second warning is in place with regard to some of the metrics we use. As it is dubious practice to correlate metrics that have a functional relationship between them (e.g. Cost per FP and Days per FP), as likely spurious correlations are found [44], we do not valid any findings with regard to these metrics as reliable.

Analysis of the statistical tests for association between paired samples as depicted in Table 8 results in the following observations.

Observation 1: Strong positive correlations are found between Project Size, Project Cost, and Number of Defects.

In the first column of Table 8, it can be seen that Project Size, measured in function points, is strongly associated with Project Cost and Number of Defects. This effect is known from related studies [49] [5] and as such not a surprise in our research. The second column shows that also among themselves Project Cost and Number of Defects are strongly interrelated. However, where in many other organizations a clear correlation is found between Project Size and Project Duration, both BELTEL and DUTCHCO show an atypical pattern. Project Size and Project Duration are not related in any way.

This is the case when both BELTEL and DUTCHCO are analyzed in a combined way, like inventoried in Table 8 and plotted in Figure 3, and also when examined separately (though for DUTCHCO's four data points we have no statistical significance). In order to examine whether this effect is only

![Figure 3. Plot of Project Duration versus Project Size; BELTEL projects are indicated in open dots, DUTCHCO in closed dots.](image)

linked to the set of 26 projects in scope, or whether this effect goes for BELTEL as a whole, we perform the test also with the BELTEL projects that are not included in this paper, yet available in our repository. A test with all 157 BELTEL projects included shows a p-value of 0.002, and a correlation coefficient of 0.24, indicating that also in this case no correlation between Size and Duration is found. This outcome supports our observation that regardless the size of a project the duration is typically ten months. This is confirmed by a relatively low standard deviation for BELTEL’s Project Duration (see Table 3).

In spite of this atypical effect with regard to Project Duration, in the fifth row a strong correlation can be seen between Days per FP and Cost per FP. Besides that we observe a relation between Days per FP and Project Size and Project Cost. However, due to the functional relationship between both metrics we do not validate these findings as reliable [44].

Observation 2: Stakeholder Satisfaction for both process and result are strongly interrelated to each other. Stakeholder Satisfaction relates negatively with Project Duration.

Row nine of Table 8 shows that an observation that was already found in our original study [6] with regard to Stakeholder Satisfaction, remains intact. Both satisfaction ratings for process and product correlate strongly with each other. The fact that the same results are found for BELTEL alone, and also when the DUTCHCO data is added provides evidence that the observation applies to both companies and may represent a more general observation that high satisfaction ratings on process link with high ratings on the delivered product. However, the weak correlation between Project Duration and Costs per FP and Days per FP was not visible in the BELTEL data and has only occurred with the addition of the DUTCHCO data. This suggest the effect is due to the DUTCHCO data (based on a single large release) and calling for more research to investigate whether the effect is real.

Column three shows that Project Duration has a moderate negative relation with Stakeholder Satisfaction for both Process and Result. Longer project durations tend to lead to lower satisfaction rates. Furthermore Project Duration relates weakly with Cost per FP and Days per FP, indicating that longer project
durations lead to higher Cost per FP and a higher number of Days per FP. However, due to the functional relationship between both metrics we do not validate these findings as reliable [44].

Two observations are related to Perceived Value. A weak uphill linear relationship between Project Size and Perceived Value (overall), as shown in our original study, is found here too, indicating that perceived value is higher for larger projects (in Function Points). Furthermore, several of the perceived value metrics show weak negative relations with Cost per FP and Days per FP, indicating that lower cost and duration per FP links with higher scores on perceived value. This effect is much reduced compared with the original study with BELTEL data only [6], suggesting that it is a BELTEL phenomenon.

![Figure 4. Plot of Perceived Value Overall versus Project Size; BELTEL projects are indicated in open dots, DUTCHCO in closed dots.](image)

A major limitation here is, that the DUTCHCO project for which Perceived Value and Stakeholder Satisfaction is measured, is significantly larger in size than all other BELTEL projects. Figure 4, with on the X-axis the Project Size in Function Points, and on the Y-axis the Overall Perceived Value rating of each project, clearly shows that a good comparison in fact is not yet possible in this context; more data is needed, especially from relatively larger projects. An additional remark on this phenomenon is that it may be that the correlations would have been weaker for BELTEL in the original paper if that analysis had been based on a more robust correlation coefficient.

We observe a striking correlation between all mutually Perceived Value measurements. We assume that the four aspects are measuring the same construct, or that the answers to those items were influenced in the same way. This effect was not measured this strongly with BELTEL data only. We assume the effect found now is an artefact of adding DUTCHCO data. Due to the fact that the results are unstable, we do not value these outcomes to high though.

Other observations with regard to perceived value, as mentioned in our original paper [6], seem to have vanished in this study. After adding the DUTCHCO data to the comparison, no relations with another project metric are observed.

A comparison of the results of the test for association which listed only BELTEL results (see Table 3 in the original study [6], with the results of the test in which both BELTEL as DUTCHCO projects are included (see Table 8), shows that the latter shows a clearer and more coherent pattern. Where the original, BelTel only, table shows a rather scattered pattern, the actual results focus on the three observations mentioned above. Especially the statistical power of function points as a measure for Project Size stands out. Besides that, we found indications for a positive relationship between both Stakeholder Satisfaction measures, and between Stakeholder Satisfaction for results and Project Duration. We did not find direct evidence for strong relations between Perceived Value. However, we do have expectations with regard to this for future research due to a very strong interconnection between the four Perceived Value measures. 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.

Observation 3: Weak correlations are found between Estimation Quality Factor for Duration on the one hand and Project Duration and Stakeholder Satisfaction on the other.

A final observation that results from the quantitative analysis is about the quality of estimations with regard to project duration (see the bottom horizontal row in Table 8). When compared to the initial BELTEL study, the only consistent observations are the negative correlation between EQF (Duration) and Project Duration and the positive correlation between Stakeholder Satisfaction and EQF (Duration). The first correlation suggests that shorter projects are less well estimated with regard to duration. However, this effect was not visible in the analysis of the BELTEL data, so must be due to the DUTCHCO data. The second suggests that stakeholders like accurate duration estimates, although in the initial study stakeholders were satisfied about the result, while after adding data stakeholders were satisfied about the process.

4.5 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 that were performed within BELTEL and DUTCHCO. See Table 9 and Table 10 for the outcomes of the coding of BELTEL and DUTCHCO free format text data. Both tables are ordered on the number of times a code was applied in the comments. We discovered seven main themes: In the following paragraphs we discuss these main themes, where we combined connected coding aspects into one theme. A subset of comments given by participants from the surveys is included in the following paragraphs, indicated by the letter “B” (for BELTEL) or “D” (for DUTCHCO) followed by a participant number.

4.5.1 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 were about good quality, however, a number had to do with low quality issues of deliverables.
A large number of negative comments given in the survey was related to the deployment of projects within a release into BELTEL’s or DUTCHCO’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, repeated rollback of releases, improvements to be made in the deployment process, and in solving issues.

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 often 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, for example:

‘A lot of discussion on how we need to test...’ (B39).
‘There were some defects in production’ (D30).

We note that deployment itself is not mentioned by any of DUTCHCO’s stakeholders. Maybe an explanation for this is the fact that (unlike the BELTEL approach with centralized deployment by a separate release team) DUTCHCO teams are themselves responsible for deployment of solutions. Summarizing, for both BELTEL and DUTCHCO a generic observation applies on quality:

Observation 4: Satisfied stakeholders tend to emphasize good quality, while dissatisfied stakeholders say testing and deployment need improvements.

4.5.2 Communication (A2)

The second most mentioned point on the stakeholder’s list in both studied companies is about communication. A number of remarks have to do with good communication between parties. A remarkable finding within BELTEL was that positive remarks all were related to external suppliers in the frontend development of website and app development, and not with the main strategic supplier INDSUP.

In the DUTCHCO case many positive remarks on communication had to do with team aspects, such as:

‘Improvements in the cooperation between teams. A lot of work done in a short amount of time’ (D02).

However, not all is well with communication. Besides the many positive remarks, there are also suggestions for improvement, sometimes related to the agile process:

‘Communication and involvement for agile items is limited to the bare minimum, so the added value of release management is not really large here. The whole agile process is still pretty blurry to most of its stakeholders, so this definitely needs to be improved’ (B48).

In the DUTCHCO case many of the negative responses on communication where about bad communication between teams. Overall, a generic observation can be made on communication:

Observation 5: Satisfied stakeholders emphasize good communication. Dissatisfied stakeholders say communication needs to be improved.

4.5.3 Requirements (A4)

Most of the comments related to requirements were about unclear requirements that hinder a project’s progress, such as:

‘Interpretation from requirements can be different and cause issues at testing phase’ (B40).

A limited number of comments were made on bad documentation, design problems and requirements creep, but also some comments were made on the availability of good requirements.

In the specific DUTCHCO case several remarks were made on unclear or hidden business rules as a cause for problems. Yet, a generic observation can be made with regard to requirements in both studied companies:

Observation 6: Dissatisfied stakeholders emphasize unclear requirements, bad documentation, hidden business rules, and requirements creep.

4.5.4 Stakeholder Satisfaction and Duration (A5, A6)

Many of the comments related to stakeholder aspects were about satisfied stakeholders. Most comments had to do with the quality of delivery and the time-to-market of delivery. Project duration and time-to-market was mentioned by many participants, where most comments are about on-time delivery. In the BELTEL case the following observation applied:

Observation 7: Satisfied stakeholders comment about good quality of duration estimates. Dissatisfied stakeholders comment about long duration and schedule overrun.

In the specific DUTCHCO case many respondents indicate they are satisfied with the product that was delivered to DUTCHCO’s customer, as for example stated by D01:

‘Despite the time constraints a decent product was delivered’.

With regard to this some respondents mentioned the inclusion of the customer as a positive factor, as stated by D10:

‘I am satisfied about the fact that we included the customer and got their feedback, which resulted in a better product’.

However, with regard to Project Duration a number of DUTCHCO respondents mention a high time pressure, especially towards the end of the release, leading to issues and defects in the last stages. As D13 says it:

‘Time issues caused several problems’.

Although duration was mentioned by stakeholders from DUTCHCO, they did this from the perspective of time pressure at the end of the release. In the DUTCHCO case no estimations were prepared with regard to duration. However, in the original BELTEL study we found that satisfied stakeholders comment about good quality of duration estimates, where dissatisfied stakeholders comment about long duration and schedule overrun. In the DUTCHCO study we found one additional observation with regard to time pressure:

Observation 8: Time pressure towards the end of a release leads to more defects.
4.5.5 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, as well as in the DUTCHCO case. Knowing this we argue that the low number of comments related to this aspect by BELTEL stakeholders (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.

However, some comments were related to the agile process itself that needed improvement, as stated for example by B48:

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

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, such as:

'No real feeling on the benefit of this project' (B45).

With regard to process aspects a limited number of comments were about needs for improvement, such as speeding up things and working in a more structured way. However, about as many comments were related to a lean and flexible process.

In the specific DUTCHCO case we observed that from the viewpoint of innovation stakeholders overall seem to be quite happy with the delivered result, although some respondents mention the fact that the release was only applicable to one specific customer, for example stated by D21:

'This was a big step in innovation, which would help us move further old applications to new technology stack' (D21).

Some stakeholders mention that not only the delivered product is innovative, but also the applied internal process:

'New technology used, new groundwork for new applications has been set. New way of working also (introduction of design street, and QA department)' (D10).

Observation 9: Satisfied stakeholders emphasize the delivery of good value to the customer.

However strongly related to remarks on to be improved internal communication, stakeholders feel that DUTCHCO's internal process should be improved, although people see things changing for the good too, as for example mentioned by D01:
The new structure of the company sort of has evolved on a better level.

Although value certainly was addressed by DUTCHCO’s stakeholders, no comments were made related to the agile delivery approach. Where in the specific BELTEL case an observation applied that the low number of comments related to agile processes does not reflect the company’s strategic choice for a new delivery approach, we argue that the DUTCHCO omission is only partially comparable with the BELTEL study; DUTCHCO stakeholders simply do not talk about agile because it is their only delivery approach. Thus, we’d like to adjust the BELTEL observation to the following generic one:

Observation 10: ‘Agile’ itself is not always a point of discussion in companies, even when they are agile.

4.5.6 Supplier Management (A11)

A number of comments were about issues with suppliers, where also BELTEL’s main supplier INDSUP was mentioned several times, such as for example by B14:

‘Very long delays and complete lack of knowledge and initiatives from INDSUP’.

In the DUTCHCO case no specific comments were found that related to supplier management. In a way this is not surprising, since besides the INDSUB teams no external parties are applicable within DUTCHCO.

4.5.7 Cost

A remarkable observation is that only once a comment is made related to cost of projects in the BELTEL study:

‘Implementation as per time, budget, and quality’ (B38).

No comments were made about the estimation accuracy with regard to project cost. The aspect of cost was mentioned only twice by DUTCHCO stakeholders.

‘We got paid quite a sum of money for this and I think we made some profit’ (D10).

‘Solution was completely paid by customer where we can use this functionality as generic functionality in our own solution’ (D03).

Although the backgrounds may be different, like BELTEL cost seems not a large issue for DUTCHCO stakeholders:

Observation 11. Cost does not seem an important issue for project stakeholders.

5. DISCUSSION

In order to validate the outcome of the quantitative analysis with the outcome of the qualitative analysis, we compared the observations from both analyses, as depicted in Table 11. We grouped the data into four themes that correspond to the horizontal rows in Table 8 that logically belong together; the core project metrics (Project Size, Project Cost, Project Duration, and Number of Defects), Stakeholder Satisfaction, Perceived Value, and the quality of duration estimations.

5.1 The Core Project Metrics

The strong positive correlations that we found between Project Size on one hand and Project Cost, Project Duration, and Number of Defects (observation 1), confirm what is already known from related work [45] [46] [47] [48] [49]. From this point of view, Project Size, measured in function points, can be considered as a very strong predictor of both cost and process quality.

Also the effect of project size as a risk factor is described earlier. Smaller projects tend to have lower cancellation rates [50] [51] [52]. Smaller projects tend to perform better in terms of quality, being on budget, and being on schedule [50] [51] [52] [53]. Project size is found to be an important risk factor for success [54] [55] [56] [57] [58] [59]. Note, however, that the literature does not match results from our study with regard to an economy-of-scale effect that larger projects in size are good, for Cost per FP and Days per FP (see paragraph 4.4). We argue that for most projects a trade-off is applicable between cost and duration on the one side, and risk on the other.

Despite the strong correlation, the use in practice of function points to measure project size suffers from shortcomings, such as additional training needed, subjective determination of complexity, and not considering the development environment [60].

In order to emphasize the effect of functional size (Function Points) as a normalizer we give an example related to the cheapest of all BELTEL projects versus the most expensive one. Project BelTel 6.3, a small (16 FPs) release-based enhancement on a CRM-application that was performed in a Scrum way as depicted above represents the minimum cost of 8,000 euro. This project scores good Cost Duration Matrix in Figure 2, and shows the highest score of all for Stakeholder Satisfaction for both process and result. To put things in perspective, the maximum cost of 296,000 euro is linked to Project BelTel 6.10, an implementation of a part of a new order management system. Yet, also this project scores on the upper side in the Cost over Time quadrant, mainly due to a 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 both projects were performed in a Scrum way as depicted above.

With regard to the strong correlations that we found between Project Size and other core project metrics such as cost and defects, it might be important to consider that Project Size was measured in Function Points, in a manual counting process. As described in the data collection approach, different artifacts were used as a source for function point counting, depending on the availability per project (e.g. sets of functional documentation, user stories recorded in one of the Scrum backlog tools, architectural documents, project documentation, user manuals, or wireframes). Manual counting was performed by different members of measurement teams of both companies, and reviewed by another member, to ensure proper use and interpretation of counting guidelines [30].

In agile environments, where usually no upfront artifacts such as functional and technical design documents are prepared, counting functional size can be challenging. However, we experienced in practice in both companies that...
Table 11. Summary of observations and implications for practice and research.

<table>
<thead>
<tr>
<th>Themes</th>
<th>Observations from the quantitative analysis</th>
<th>Observations from the qualitative analysis</th>
<th>Implications for practice and research</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The Core Project Metrics</strong></td>
<td>Strong positive correlations are found between Project Size, Project Cost, and Number of Defects (observation 1).</td>
<td>Cost does not seem an important issue for project stakeholders (observation 11).</td>
<td>The strong correlations between Project Size on the one hand, and Project Cost, Project Duration, and Number of Defects on the other are confirmed by related work [44] [5] [46] [47] [48] [49].</td>
</tr>
<tr>
<td><strong>Stakeholder Satisfaction</strong></td>
<td>Stakeholder Satisfaction for both process and result are strongly interrelated to each other. (observation 2). Stakeholder Satisfaction relates negatively with Project Duration (observation 2).</td>
<td>Satisfied stakeholders emphasize good quality (observation 4), good communication (observation 5), and good quality of duration estimations (observation 7). Dissatisfied stakeholders state that testing and deployment needs improvement (observation 4), they emphasize unclear requirements, bad documentation, hidden business rules, and requirements creep (observation 6), they say communication needs to be improved (observation 5), and they comment about long duration and schedule overrun (observation 7).</td>
<td>Additional research is needed to identify how good quality of deliverables, good communication, and reliable estimations for duration can be controlled in practice in order to create satisfied stakeholders. Additional research is needed to identify how issues with testing and deployment, unclear requirements, hidden business rules, bad communication, and schedule overrun can be controlled in order to mitigate stakeholder dissatisfaction.</td>
</tr>
<tr>
<td><strong>Perceived Value</strong></td>
<td>Although the different Perceived Value measures interrelate strongly with each other, we cannot draw general conclusions from this Weak correlations are found between Perceived Value and Project Size, Cost per FP, and Days per FP, we cannot draw general conclusions from this BéLTLe phenomenon.</td>
<td>Satisfied stakeholders emphasize the delivery of good value to the customer (observation 9) (DUTCHCO). ‘Agile’ itself is not always a point of discussion in companies, even when they are agile (observation 10).</td>
<td>Additional research is needed (especially on medium and big sized projects) to validate whether larger projects tend to lead to higher perceived value scores. If this effect is true, functional size might be a potential indicator for value, that can be used in practice by Product Owners to prioritize (enterprise) backlogs.</td>
</tr>
<tr>
<td><strong>Estimation Quality for Project Duration</strong></td>
<td>Weak correlations are found between Estimation Quality Factor for Duration and Project Duration, and Stakeholder Satisfaction (observation 3).</td>
<td>Satisfied stakeholders emphasize good quality of duration estimations (observation 7). Dissatisfied stakeholders comment about long duration and schedule overrun (observation 7). Time pressure towards the end of a release leads to more defects (observation 8) (DUTCHCO).</td>
<td>Much research is performed on Effort Estimation of software projects [72], yet very limited research is performed on the effects of the quality of Duration Estimation, because the outcomes of this study indicate correlations with both stakeholder satisfaction and perceived value.</td>
</tr>
</tbody>
</table>

In order to automate the data collection process where possible, we strongly felt a need for some form of automated function point count, if possible based on the code itself. However, a follow-up exploratory study of 336 functional size measurement specialists that was performed based on this hypothesis did showed that overall automated functional size measurement was considered as important, but also difficult to realize [61].

5.2 Stakeholder Satisfaction

Observation 2, indicates a moderate correlation between Stakeholder Satisfaction for both process and result and Project Duration. We found a moderate relation between Stakeholder Satisfaction for Estimation Quality Factor (Duration) too. This indicates that Stakeholder Satisfaction is related to interaction and being informed, yet also with conformance to planning and estimation of the delivery date. A strategy of ‘no last minute surprises’ as such helps better to increase Stakeholder Satisfaction, as well as giving attention to improvement of estimation and planning practices would.

Satisfaction of stakeholders with the development process and with the development outcome was studied before by Ferreira and Cohen [62]. They found “strong positive effects of agile practice (iterative development, continuous integration, collective ownership, test-driven design, and feedback) on stakeholder satisfaction with both development process and the project outcome”. A relation between stakeholder satisfaction and with agile software development in an Indian context was found by Nazir et al. [63].

Satisfaction of stakeholders with the development process and with the development outcome was studied before by Ferreira and Cohen [62]. They found “strong positive effects of agile practice (iterative development, continuous integration, collective ownership, test-driven design, and feedback) on stakeholder satisfaction with both development process and the project outcome”. A relation between stakeholder satisfaction and with agile software development in an Indian context was found by Nazir et al. [63].
Many participants mentioned communication to be important, while good communication is mentioned by satisfied stakeholders, and bad communication by dissatisfied ones. Approximately half of the comments were about good communication, such as good alignment between parties, good collaboration, and short feedback loops. The other half mentioned communication to be improved, such as provide information on processes and innovations (e.g., agile delivery), ongoing discussions, and miscommunication with suppliers. Unclear requirements, bad documentation, requirements creep, and bad quality of test and deployment resources are perceived by dissatisfied stakeholders as causes for bad quality of deliverables.

However, a warning is in place here: we notice that many positive comments on communication within BELTEL are linked to two specific Product Owners. 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” [64].

Contrary to the findings in our original study, we did not find evidence for correlations of Stakeholder Satisfaction with Number of Defects. However, in the qualitative analysis, we do find many comments that are in one way or another related to those aspects. Quality of the deliverables (both good quality and to be improved quality), in combination with testing aspects and deployment into the production environment, is commonly mentioned in comments by all participants.

5.3 Perceived Value

We assume that the relative absence of comments that are related to the ongoing innovation of implementing a more agile delivery process within BELTEL, in combination with the limited focus on value might be of importance here (observation 7). The low interest in agile innovation among BELTEL’s stakeholders in a way reflects our findings in the quantitative analysis too. No significant relation is found between Perceived Value and any other project metric of software project deliverables. However, the limited number of projects in scope of this study, combined with the diversity in project sizes (with many small and only one large project) can be a reason for replicating our study with more data in future.

Furthermore, what strikes with regard to the value measures are the almost perfect correlation coefficients (from 0.94 to 0.99) with p-value < 0.001 that we found between all four Perceived Value measures mutually. This is not a good sign when designing scales. It might imply that the four aspects, derived from well-known research on Kaplan and Norton’s Balanced Scorecard [42], are measuring the same construct, or that the answers to those items were influenced in the same way. Besides that a warning is in place here with regard to functional relations between the perceived Value metrics. As it is dubious practice to correlate metrics that have a functional relationship between them, as likely spurious correlations are found [44], we do not validate the findings on interrelated Perceive Value metrics as reliable.

5.4 Estimation Quality for Duration

The comments given in the surveys confirm the more or less company specific observations with regard to Project Duration. Stakeholders of projects are satisfied when delivery of results is in-time, where we assume this relates to good quality of duration estimates. However, it needs to be said that the words estimate or estimation are never used in the comments. Dissatisfaction of stakeholders is often linked with too late delivery and long project durations (long waiting time).

In the DUTCHCO case we do not find direct evidence for this, although time pressure towards the end of a release, causing issues and defects at a late stage in the release, are mentioned as a source for problems by many respondents.

5.5 Success or failure: complex relations

Looking at the seventh row and eighth column in Table 8, the test results of association between paired samples, it strikes that no correlation is found at all between Cost Duration Index and any of the other samples. Apparently no relation exists between the position of a project in our Cost Duration Matrix and the measure of Stakeholder Satisfaction and perceived Value of that project. Based on this, we conclude that success and failure apparently are more complex than cost, duration and defects only: stakeholders can be satisfied or have the perception of much value delivered, even when a project is in the so-called Bad Practice quadrant.

We suspect, based partly on recent ongoing research, that the limitation in the current study to internal stakeholders of projects can play a limiting role here. For external stakeholders, usually the customers of the software organizations that actually pay for the software and use it in practice, cost, duration and number of errors seem to be a very important factor for success or failure. Besides that we suspect that Perceived Value needs to be measured on a lower level than a project (e.g., at user story or at epic level). We argue that additional research is needed to unfold these complex relations.

5.6 Agile and Cost were not mentioned

Except for the last four themes that resulted from the quantitative and qualitative analyses, we found two issues that were not mentioned in the test free format of the surveys, and that were not found in the tests for association between paired samples.

Firstly, we found a very low number of comments in both studies, that are related to the concept of “agile” itself. A number of observations that were applicable for BELTEL did not apply to DUTCHCO, and the other way around. No evidence was found within DUTCHCO on the relation between satisfied stakeholders and good quality of estimates. DUTCHCO does not produce estimates as such for its software delivery activities. Like BELTEL, within DUTCHCO no specific remarks were made about the agility of its process. However, an agile development approach within DUTCHCO is widely assumed as the only form of process, no other approaches (e.g., plan-driven) are applicable. Thus, we assume that no comments were made about this just because it is “business as usual.”

Besides the development method used, we assume that also differences in test and deployment (e.g. release)
approaches of both software companies influence the portfolio performances. The fact that DUTCHCO collects all of its features in large three-monthly releases, while BELTEL runs a variety of software projects sequentially with eight combined releases per year, should be taken into account when comparing the performances of both companies.

A second finding was the fact that cost seems not an important issue for stakeholders within both BELTEL and DUTCHCO. Only one comment by a BELTEL stakeholder is made related to this. This finding applies to DUTCHCO too; not much attention is given in the comments on cost either. We think that this might be caused by the effect that in more or less agile organizations, the focus tends to shift from time and cost driven controlling towards scope and value driven steering. Agile teams tend to stay in place for longer periods, and budgeting often needs to be done only once a year, instead of many times per year in a pre-project phase in plan-driven organizations.

In a way this is not a big surprise in the DUTCHCO organization, since development teams are fixed and stay together for long periods of time. Due to this cost is just a derivative of effort spent by these teams. No budget estimations upfront are applicable in the DUTCHCO organization. Although this cannot be understood in a way that DUTCHCO is not interested in cost at all; DUTCHCO’s management team is highly interested in cost reduction based on the effects of shortening learning curves [28] and efficiency improvements based on outsourcing approaches.

5.7 Implications

The outcomes of both our case studies might not simply be generalized to other environments. We identify a number of take-away-messages that apply to research and practice in other software companies too.

The first one relates to the from related work already known strong correlations between Project Size on the one hand, and Project Cost, Project Duration, and Number of Defects on the other, indicating the power of Project Size (measured in Function Points) as an indicator for cost, duration, and quality. However, a link with agile development is poorly covered in research [61], and in practice many agile software companies tend to see (manual) counting of functional size as waste. We argue that also agile practitioners and researchers should rethink and embrace Project Size.

Secondly, we argue on the one hand that good quality of deliverables, good communication, and reliable estimations for duration can be used to increase stakeholder satisfaction, and on the other that issues with testing and deployment, unclear requirements, hidden business rules, bad communication, and schedule overrun increase stakeholder dissatisfaction, and should therefore be mitigated when possible. Assuming that agile development methods might play a role here, additional research is needed to identify the backgrounds and ways to control these findings in a practical context.

Thirdly, we recognize a need for additional research (especially on medium and large sized projects) to validate whether larger projects tend to lead to higher perceived value scores. We think that measuring perceived value on a lower level than a software project, e.g. on user stories or epics, might result in other outcomes. If a strong positive effect is found in future, functional size might be a potential indicator for value, that can be used in practice by Product Owners to prioritize (enterprise) backlogs.

Finally, a fourth take-away-message relates to a need for additional research in a practical context on the effects of the quality of Duration Estimation, because the outcomes of this study indicate correlations with both stakeholder satisfaction and perceived value. Good quality estimation of a project’s delivery date seems very important for stakeholders, and relates to the perception of value that is delivered.

5.8 Threats to Validity

5.8.1 Construct 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, which 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 Promotor Score) cannot easily be related to software projects, mainly because too many different factors are of influence for such measurements.

As explained in paragraph 5.3, two limitations are in place with regard to the setup of our electronic survey. Adopting any of the existing validated measurement instruments on customer / stakeholder satisfaction, which are available from the behavioral science, and economics and management theory, might be helpful for continuation of the survey in future research. Secondly, the almost perfect correlations between the four aspects of Perceived Value indicate that the aspects are measuring the same construct, or that the answers to those items were influenced in the same way. We argue that it would be good to adjust the survey with regard to these aspects for future research.

5.8.2 Internal Validity

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. We limited this effect by making Benjamini-Hochberg corrections for all p-values that we used in the multiple comparisons (see Table 8). Furthermore, 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 (from the organizations involved) that did not know the subject projects in the assessment of perceived value. However, the study was performed in an operational context within BELTEL and subsequently DUTCHCO.
Answering surveys, subsequent a release, was implemented as an operational capability. When designing the study we considered that it was undesirable to interfere with 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.

A limitation is in place regarding the summarized themes in Table 9 and Table 10; the included key concepts are defined loosely on free text concepts provided by respondents in which they later were categorized.

5.8.3 External Validity

Concerning external validity, the extent to which the results of our study can be generalized to other companies than BELTEL and DUTCHCO is difficult to answer because we performed multiple case studies in these two specific companies. Not all findings that occurred in the BELTEL case were found in the DUTCHCO case too. 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 [11] in a way we performed for this study within both BELTEL and DUTCHCO is a precondition for mature improvement within other companies too.

5.8.4 Study Reliability

A threat related to the study’s reliability lies in the fact that the lead author of this paper was a member of the measurement team within BELTEL, and carried out the functional size measurements within DUTCHCO. However, we tried to prevent from bias by ensuring that the BELTEL measurement team and DUTCHCO measurement expert are independent and objective in their collection of data. Moreover, the size measurements were made before the analyses were made and performed along the functional size measurement procedures that are repeatable independently from the actual measurement expert [30].

A link that we did not study, but that is mentioned in other studies, is the relation between stakeholder satisfaction and requirements [65] [66] and documentation [67].

6. CONCLUSIONS AND FUTURE RESEARCH

The outcomes of our multiple case 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) and optimal collaboration between teams, has a positive effect on satisfaction of stakeholders. On the other hand, too late delivery and long project durations, and many defects dissatisfy them. Our study does not provide any evidence that steering on costs helped to improve the satisfaction of stakeholders.

A novelty in the results of our study is that we linked Perceived Value to a set of project metrics, among others functional size of projects. As an answer to our research question - how do stakeholder satisfaction and perceived value relate to cost, duration, defects, size and estimation accuracy of software projects? – we found the following five take-away-messages:

- **Stakeholder Satisfaction** can be improved by steering on good quality, good communication, and good quality of duration estimations. Satisfied stakeholders emphasize the delivery of good value to the customer.
- **Stakeholder Satisfaction** goes down when issues occur with testing and deployment, unclear requirements, bad documentation, hidden business rules, and requirements creep, when communication is bad, and in case of long project duration and schedule overrun.
- **Perceived Value** correlates (however weakly) with **Project Size**, **Cost per FP**, and **Days per FP**, indicating that functional size might be an indicator for value, however more research is needed to confirm this finding.
- We identified two themes that did not apply to stakeholders or value: ‘Agile’ itself is not always a point of discussion in companies, even when they are agile, and **Project Cost** seems not an important issue for stakeholders.
- The study also confirmed an effect known from related work: **Project Size** strongly correlates with the other core project metrics **Project Cost**, **Project Duration**, and **Number of Defects**.

Our final question is how we and others build on the main findings of this study. We see the following four aspects to be important for further research:

- How can good quality of deliverables, good communication, and reliable estimations for duration be controlled to create satisfied stakeholders?
- How can issues with testing and deployment, unclear requirements, hidden business rules, bad communication, and schedule overrun be controlled to mitigate stakeholder dissatisfaction?
- Do larger projects (in functional size) lead to higher perceived value scores?
- How can quality of project duration estimations be used to improve stakeholder satisfaction and perceived value?

ACKNOWLEDGMENTS

We thank BELTEL, DUTCHCO, and INVEND for their 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. We thank Tableau for allowing us to use their BI solution to build our Performance Dashboard.

REFERENCES


[33] C. Wu, M. Naqibuddin and L. Fleisher, "Measurement of patient satisfaction as an outcome of regional anesthesia and analgesia:


The Effects of Perceived Value and Stakeholder Satisfaction on Software Project Impact


7. Appendix A

7.1 R-script used for this research paper

```r
setwd("C:/Users/Hennie/OneDrive/Documents/Promotieonderzoek/Elsevier_IST")

Repository <- read.csv("MeasurementRepository.csv", header = TRUE, sep="", dec=".", fill = TRUE)
EBSPM <- read.csv("MeasurementRepositoryAll.csv", header = TRUE, sep="", dec=".", fill = TRUE)

RepositoryBeltel = subset(Repository, Repository$Organization == "Beltel")
RepositoryDutchco = subset(Repository, Repository$Organization == "DutchCo")
RepositoryBeltelAll = subset(EBSPM, EBSPM$Organization == "Beltel")
RepositoryDutchcoAll = subset(EBSPM, EBSPM$Organization == "DutchCo")

summary(RepositoryBeltel)
summary(EBSPM)
summary(RepositoryDutchco)
str(RepositoryDutchco)
str(RepositoryBeltel)
str(RepositoryBeltelAll)
str(RepositoryDutchcoAll)
str(EBSPM)

cor.test(RepositoryBeltel$Cost, RepositoryBeltel$Size,
  use="pairwise", alternative = c("two.sided", "less", "greater"),
  method = c("spearman"),
  exact = NULL, conf.level = 0.95, continuity = FALSE)

cor.test(RepositoryBeltelAll$Duration, RepositoryBeltelAll$Size,
  use="pairwise", alternative = c("two.sided", "less", "greater"),
  method = c("spearman"),
  exact = NULL, conf.level = 0.95, continuity = FALSE)

cor.test(RepositoryDutchcoAll$Duration, RepositoryDutchcoAll$Size,
  use="pairwise", alternative = c("two.sided", "less", "greater"),
  method = c("spearman"),
  exact = NULL, conf.level = 0.95, continuity = FALSE)

# Statistical tests for comparing overall differences
wilcox.test(EBSPM$Size, Repository$Size)
wilcox.test(EBSPM$Cost, Repository$Cost)
wilcox.test(EBSPM$Duration, Repository$Duration)
wilcox.test(EBSPM$Defects, Repository$Defects)

# Statistical tests for comparing differences per size
wilcox.test(EBSPM$Cost/EBSPM$Size, Repository$Cost/Repository$Size)
wilcox.test(EBSPM$Duration/EBSPM$Size, Repository$Duration/Repository$Size)
wilcox.test(EBSPM$Defects/EBSPM$Size, Repository$Defects/Repository$Size)
```

The Effects of Perceived Value and Stakeholder Satisfaction on Software Project Impact
The Effects of Perceived Value and Stakeholder Satisfaction on Software Project Impact

7.2 Results of the statistical analysis

7.2.1 Summary of the BelTel subset

```r
> summary(RepositoryBeltel)

  ID Organization ProjectID
  Min. :390.0  Beltel:22  Project 3.1: 1
  1st Qu.:424.2  DutchCo: 0  Project 3.2: 1
  Median :429.5
  Mean   :426.4
  3rd Qu.:436.5
  Max.   :444.0

ProjectDescription
Adapt T&C Procedure DOB Wmode Platform Small enhancement to an existing system Plan-driven delivery approach. Performed by a team of the strategic partner (Indian, partly onsite, partly off
```
<table>
<thead>
<tr>
<th>Size</th>
<th>Duration</th>
<th>Cost</th>
<th>Cost PerFP</th>
<th>Days PerFP</th>
</tr>
</thead>
<tbody>
<tr>
<td>M.n.</td>
<td>12.00</td>
<td>4.96 M.n.</td>
<td>8000 M.n.</td>
<td>434.0 M.n.</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>24.25</td>
<td>8.95 1st Qu.</td>
<td>46013 1st Qu.</td>
<td>836.8 1st Qu.</td>
</tr>
<tr>
<td>Median</td>
<td>38.50</td>
<td>10.29 Median</td>
<td>66209 Median</td>
<td>1612.0 Median</td>
</tr>
<tr>
<td>Mean</td>
<td>80.77</td>
<td>10.22 Mean</td>
<td>99562 Mean</td>
<td>1932.2 Mean</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>127.25</td>
<td>11.50 3rd Qu.</td>
<td>135342 3rd Qu.</td>
<td>2790.2 3rd Qu.</td>
</tr>
<tr>
<td>Max.</td>
<td>324.00</td>
<td>19.03 Max.</td>
<td>296000 Max.</td>
<td>4976.0 Max.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Defects Per FP</th>
<th>CostDurationIndex</th>
<th>Quadrant</th>
<th>EQF Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>M.n.</td>
<td>0.0200</td>
<td>Bad Practice</td>
<td>6 M.n.</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>0.0975</td>
<td>Cost over Time: 14 1st Qu.</td>
<td>2.123</td>
</tr>
<tr>
<td>Median</td>
<td>0.2150</td>
<td>Good Practice: 2 Median</td>
<td>3.520</td>
</tr>
<tr>
<td>Mean</td>
<td>0.3728</td>
<td>Mean</td>
<td>4.321</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>0.6000</td>
<td>3rd Qu.</td>
<td>7.992</td>
</tr>
<tr>
<td>Max.</td>
<td>1.4400</td>
<td>Max.</td>
<td>9.140</td>
</tr>
<tr>
<td>NA's</td>
<td>4</td>
<td>NA's</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EQF Duration</th>
<th>Multi Application Release</th>
<th>Release Based</th>
<th>Once Only Project</th>
</tr>
</thead>
<tbody>
<tr>
<td>M.n.</td>
<td>1.890</td>
<td>M.n.</td>
<td>0.0000 1st Qu.</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>2.393</td>
<td>1st Qu.</td>
<td>0.0000 1st Qu.</td>
</tr>
<tr>
<td>Median</td>
<td>3.600</td>
<td>Median</td>
<td>0.0000 Median</td>
</tr>
<tr>
<td>Mean</td>
<td>7.273</td>
<td>Mean</td>
<td>0.2727 Mean</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>6.327</td>
<td>3rd Qu.</td>
<td>0.7500 3rd Qu.</td>
</tr>
<tr>
<td>Max.</td>
<td>27.570</td>
<td>Max.</td>
<td>1.0000 Max.</td>
</tr>
<tr>
<td>NA's</td>
<td>2</td>
<td>NA's</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Single Application Dependencies with other applications Migration Project</th>
</tr>
</thead>
<tbody>
<tr>
<td>M.n.</td>
</tr>
<tr>
<td>1st Qu.</td>
</tr>
<tr>
<td>Median</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>3rd Qu.</td>
</tr>
<tr>
<td>Max.</td>
</tr>
</tbody>
</table>
7.2.2 Summary of the EBSPM repository

> summary(EBSPM)

    ID Organization ProjectID
  Min. :1.0  Banking A:206  Mode:logical
  1st Qu.:123.8  Banking B:125  NA's:492
             NA's :2

7.2.2 Summary of the EBSPM repository
### The Effects of Perceived Value and Stakeholder Satisfaction on Software Project Impact

<table>
<thead>
<tr>
<th>Project Description</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remove duplicate activation in bundle area</td>
<td>2 M.n.</td>
</tr>
<tr>
<td>Replacement of the roaming tariff solution</td>
<td>2 Median</td>
</tr>
<tr>
<td>S&amp;V Renewal Search Engine (GICT)</td>
<td>1 Mean</td>
</tr>
<tr>
<td>(PCK) Plan Releases (GICT)</td>
<td>1 3rd Qu.</td>
</tr>
<tr>
<td>1-day porting (legal/BIPT requirement)</td>
<td>1 Max.</td>
</tr>
<tr>
<td>(Other)</td>
<td>483</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Duration</th>
<th>Cost</th>
<th>CostPerFP</th>
<th>DaysPerFP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>0.900</td>
<td>329</td>
<td>82</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>5.397</td>
<td>70796</td>
<td>1236</td>
</tr>
<tr>
<td>Median</td>
<td>8.411</td>
<td>278156</td>
<td>2520</td>
</tr>
<tr>
<td>Mean</td>
<td>8.950</td>
<td>592802</td>
<td>3618</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>11.070</td>
<td>699527</td>
<td>4546</td>
</tr>
<tr>
<td>Max.</td>
<td>26.840</td>
<td>6802466</td>
<td>35818</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DefectsPerFP</th>
<th>CostDurationIndex</th>
<th>Quadrant</th>
<th>EQFCost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>0.000</td>
<td>6.19</td>
<td>Bad Practice</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>0.080</td>
<td>86.69</td>
<td>Cost over Time</td>
</tr>
<tr>
<td>Median</td>
<td>0.160</td>
<td>91.72</td>
<td>Good Practice</td>
</tr>
<tr>
<td>Mean</td>
<td>0.351</td>
<td>89.19</td>
<td>Time over Cost</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>0.360</td>
<td>95.46</td>
<td></td>
</tr>
<tr>
<td>Max.</td>
<td>5.340</td>
<td>99.14</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EQFDuration</th>
<th>MultiApplicationRelease</th>
<th>ReleaseBased</th>
<th>OnceOnlyProject</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode: logical</td>
<td>Mode: logical</td>
<td>Mode: logical</td>
<td>Mode: logical</td>
</tr>
<tr>
<td>NA's: 492</td>
<td>NA's: 492</td>
<td>NA's: 492</td>
<td>NA's: 492</td>
</tr>
</tbody>
</table>

| Single.application Dependencies.with.other.applications Migration.Project |
|-------------------------|-------------------------|-------------------------|
| Mode: logical | Mode: logical | Mode: logical |
| NA's: 492 | NA's: 492 | NA's: 492 |

| PackageOfTheShelf SteadyHeartbeat FixedExperiencedTeam BusinessDriven |
|------------------------|------------------------|------------------------|
| Mode: logical | Mode: logical | Mode: logical |
| NA's: 492 | NA's: 492 | NA's: 492 |

<table>
<thead>
<tr>
<th>RulesAndRegulationsDriven</th>
<th>Pilot</th>
<th>POC Y e a rGoL i ve</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode: logical</td>
<td>Mode: logical</td>
<td>M.n.</td>
</tr>
<tr>
<td>NA's: 492</td>
<td>NA's: 492</td>
<td>1st Qu.</td>
</tr>
<tr>
<td>Median</td>
<td>2011</td>
<td>Mean</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>2014</td>
<td>Max.</td>
</tr>
</tbody>
</table>

TUD-SERG-2017-001
### Business Domain and Development Method

<table>
<thead>
<tr>
<th>Business Domain</th>
<th>Development Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client and Account Management</td>
<td>Plan-driven: 385</td>
</tr>
<tr>
<td>Internet &amp; Mobile</td>
<td>RUP: 1</td>
</tr>
<tr>
<td>Finance &amp; Risk</td>
<td>Scrum: 106</td>
</tr>
<tr>
<td>Payments</td>
<td></td>
</tr>
<tr>
<td>Billing</td>
<td></td>
</tr>
<tr>
<td>Savings &amp; Loans</td>
<td></td>
</tr>
<tr>
<td>(Other)</td>
<td></td>
</tr>
</tbody>
</table>

### Stakeholder Satisfaction Process and Stakeholder Satisfaction Product

- Mode: logical
- NA's: 492

### Perceived Value

<table>
<thead>
<tr>
<th>Mode: customer</th>
<th>Mode: financial</th>
<th>Mode: internal process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode: logical</td>
<td>Mode: logical</td>
<td>Mode: logical</td>
</tr>
<tr>
<td>NA's: 492</td>
<td>NA's: 492</td>
<td>NA's: 492</td>
</tr>
</tbody>
</table>

### Perceived Value Innovation and Overall

<table>
<thead>
<tr>
<th>Mode: logical</th>
<th>Mode: logical</th>
<th>Defects</th>
<th>Errors First Month</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA's: 492</td>
<td>NA's: 492</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Development Classification

- Conversion (<5% new): 28
- Major Enhancement (25-75% new): 156
- Minor Enhancement (5-25% new): 126
- New Development: 182

### Summary of the DutchCo subset

```r
> summary(RepositoryDutchco)

ID   Organization ProjectID
Min. :446.0    0     DutchCo 5.1 - POR: 1
1st Qu.:446.8  DutchCo: 4 DutchCo 5.1 - AM: 1
Median :469.0   DutchCo 5.2 : 1
Mean  :469.0    DutchCo 5.3 : 1
3rd Qu.:491.2  Project 3.1: 0
Max.  :492.0    Project 3.2: 0
                ( Other ) : 0
```

### Project Description

- DutchCo Release 5.2: 3
- DutchCo Release 5.3 (Baykom): 1

Adapt T&C Procedure DOB Whole Platform Small enhancement to an existing system Plan-driven delivery approach. Performed by a team of the strategic partner (Indian, partly onsite, partly off site). Project manager of the Telecommunications (onsite): 0
Admin Fee for SoHo Customers. Rules- and regulations driven small enhancement to existing CRM system. Small project team of onsite Project Manager and Business Analyst and offshore Indian developers.

Apple Store SSO and QR Code. New development in the frontend (including mobile) environment by a team of the fixed supplier for frontend development (offsite, Belgium). Project manager of the telecom company onsite.

BSCS CDR Cycling Control. Small enhancement on a Billing application. Performed by a project team (globally distributed; partly offshore, Indian supplier) for Billing system enhancements. Performed as a fixed price project.

<table>
<thead>
<tr>
<th>Size</th>
<th>Duration</th>
<th>Cost</th>
<th>Cost Per FP</th>
<th>Days Per FP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>277.0</td>
<td>3.780</td>
<td>125827</td>
<td>376.0</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>320.5</td>
<td>5.902</td>
<td>192040</td>
<td>589.8</td>
</tr>
<tr>
<td>Median</td>
<td>784.0</td>
<td>7.710</td>
<td>514486</td>
<td>686.0</td>
</tr>
<tr>
<td>Mean</td>
<td>776.5</td>
<td>7.372</td>
<td>512897</td>
<td>630.2</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>1240.0</td>
<td>9.180</td>
<td>835343</td>
<td>726.5</td>
</tr>
<tr>
<td>Max.</td>
<td>1261.0</td>
<td>10.290</td>
<td>896788</td>
<td>773.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DefectsPerFP</th>
<th>CostDurationIndex</th>
<th>Quadrant</th>
<th>EQFCost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>0.0400</td>
<td>M.n. : 95.03</td>
<td>Bad Practice : 0</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>0.1075</td>
<td>1st Qu. : 96.12</td>
<td>Cost over Time : 0</td>
</tr>
<tr>
<td>Median</td>
<td>0.2200</td>
<td>Median : 96.72</td>
<td>Good Practice : 4</td>
</tr>
<tr>
<td>Mean</td>
<td>0.2100</td>
<td>Mean : 96.72</td>
<td>Mean : NaN</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>0.3225</td>
<td>3rd Qu. : 97.41</td>
<td></td>
</tr>
<tr>
<td>Max.</td>
<td>0.3600</td>
<td>Max. : 98.22</td>
<td>Max. : NA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EQFDuration</th>
<th>MultiApplicationRelease</th>
<th>ReleaseBased</th>
<th>OnceOnlyProject</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>0.8100</td>
<td>M.n. : 1</td>
<td>M.n. : 0</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>0.9075</td>
<td>1st Qu. : 1</td>
<td>1st Qu. : 0</td>
</tr>
<tr>
<td>Median</td>
<td>1.0050</td>
<td>Median : 1</td>
<td>Median : 0</td>
</tr>
<tr>
<td>Mean</td>
<td>1.1025</td>
<td>Mean : 1</td>
<td>Mean : 0</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>1.2000</td>
<td>3rd Qu. : 1</td>
<td>3rd Qu. : 0</td>
</tr>
<tr>
<td>Max.</td>
<td>1.3000</td>
<td>Max. : 1</td>
<td>Max. : 0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SingleDependencies</th>
<th>with other applications</th>
<th>M iration</th>
<th>Project</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>1</td>
<td>M.n. : 0</td>
<td>M.n. : 0</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>1</td>
<td>1st Qu. : 0</td>
<td>1st Qu. : 0</td>
</tr>
<tr>
<td>Median</td>
<td>1</td>
<td>Median : 0</td>
<td>Median : 0</td>
</tr>
<tr>
<td>Mean</td>
<td>1</td>
<td>Mean : 0</td>
<td>Mean : 0</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>1</td>
<td>3rd Qu. : 0</td>
<td>3rd Qu. : 0</td>
</tr>
<tr>
<td>Max.</td>
<td>1</td>
<td>Max. : 0</td>
<td>Max. : 0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PackageOffTheShelf</th>
<th>SteadyHeartbeat</th>
<th>FixedExperiencedTeam</th>
<th>BusinessDriven</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>0</td>
<td>M.n. : 1</td>
<td>M.n. : 1</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>0</td>
<td>1st Qu. : 1</td>
<td>1st Qu. : 1</td>
</tr>
<tr>
<td>Median</td>
<td>0</td>
<td>Median : 1</td>
<td>Median : 1</td>
</tr>
<tr>
<td>Mean</td>
<td>0</td>
<td>Mean : 1</td>
<td>Mean : 1</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>0</td>
<td>3rd Qu. : 1</td>
<td>3rd Qu. : 1</td>
</tr>
<tr>
<td>Max.</td>
<td>0</td>
<td>Max. : 1</td>
<td>Max. : 1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RulesAndRegulationsDriven</th>
<th>Pilot</th>
<th>POC</th>
<th>YearGoLive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>0</td>
<td>M.n. : 0</td>
<td>M.n. : 2016</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>0</td>
<td>1st Qu. : 1</td>
<td>1st Qu. : 1</td>
</tr>
<tr>
<td>Median</td>
<td>0</td>
<td>Median : 1</td>
<td>Median : 1</td>
</tr>
<tr>
<td>Mean</td>
<td>0</td>
<td>Mean : 0</td>
<td>Mean : 2016</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>0</td>
<td>3rd Qu. : 0</td>
<td>3rd Qu. : 0</td>
</tr>
<tr>
<td>Max.</td>
<td>0</td>
<td>Max. : 0</td>
<td>Max. : 2016</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BusinessDomain</th>
<th>DevelopmentMethod</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Descriptive statistics of the DutchCo subset

```r
> str(RepositoryDutchco)
'data.frame': 4 obs. of 39 variables:
$ ID : int 446 447 491 492
$ Organization : Factor w/ 2 levels "Beltel","DutchCo" : 2 2 2 2
$ ProjectID : Factor w/ 26 levels "DutchCo 5.1 - POR",... : 2 1 3 4
$ ProjectDescription : Factor w/ 24 levels "Adapt T&C Procedure DOB Wmode Platform Small enhancement to an existing system Plan-driven delivery approach. Performed by a | truncated,... : 5 5 5 6
$ Size : int 335 277 1233 1261
$ Duration : num 10.29 6.61 8.61 8.81
$ Cost : int 125827 214111 814861 896788
$ CostPerFP : int 376 773 661 711
$ DaysPerFP : num 0.34 1.13 0.16 0.21
$ DefectsPerFP : num 0.04 0.36 0.31 0.13
$ CostDurationIndex : num 98.2 95 97.1 96.5
$ Quadrant : Factor w/ 3 levels "Bad Practice",... : 3 3 3 3
$ EQFCost : num NA NA NA NA
$ EQFDuration : num 0.81 1.2
$ MultiApplicationRelease : int 0 0 0 0
$ ReleaseBased : int 1 1 1 1
```

#### 7.2.4 Development classification

- **Conversion (<5% new):** 0
- **Major Enhancement (25-75% new):** 4
- **Minor Enhancement (5-25% new):** 0
- **New Development:** 0
The Effects of Perceived Value and Stakeholder Satisfaction on Software Project Impact

7.2.5 Descriptive statistics of the BelTel subset

```r
> str(RepositoryBeltel)
'data.frame': 22 obs. of 39 variables:
  $ ID             : int 390 393 403 411 423 424 425 426 427 428 ...
  $ Organization  : Factor w/ 2 levels "Beltel","DutchCo": 1 1 1 1 1 1 1 1 1 1 ...
  $ ProjectID     : Factor w/ 26 levels "DutchCo 5.1 - POR",..: 9 6 8 10 7 ...
  $ ProjectDescription: Factor w/ 24 levels "Adapt T&C Procedure DOB Wmode Platform Small enhancement to an existing system Plan-driven delivery approach. Performed by a"...: 20 4 7 24 9 14 2 10 13 11 ...
  $ Size          : int 232 34 177 35 128 13 12 63 130 28 ...
  $ Duration      : num 12.59 9.69 9.69 19.03 8.81 ...
  $ Cost          : int 100752 25092 143575 110643 62964 45817 54500 ...
  $ CostPerFP     : int 434 738 811 3161 492 4542 1746 1618 1338 ...
  $ DaysPerFP     : num 1.65 8.67 1.67 16.55 2.09 ...
  $ DefectsPerFP  : num 0.22 0.09 0.8 0.86 0.13 0.92 0.33 0.16 0.02 0.21 ...
  $ CostDurationIndex: num 99.8 97.5 86.9 98.8 97.3 ...
  $ Quadrant      : Factor w/ 3 levels "Bad Practice",..: 2 2 1 2 1 1 3 2 ...
  $ EQFCost       : num NA NA NA 3.94 NA NA ...
  $ EQFDuration   : num 1.89 2.68 3.99 3.11 NA ...
  $ MultiApplicationRelease: int 1 1 1 1 0 1 0 0 0 ...
  $ ReleaseBased : int 0 0 0 1 1 0 1 0 1 1 ...
  $ OnceOnlyProject: int 0 0 0 0 0 0 0 0 1 0 ...
  $ Single.application: int 0 0 0 0 0 0 0 0 0 0 ...
```

TUD-SERG-2017-001
7.2.6 Descriptive statistics of all BelTel projects in the EBSPM repository

> str(RepositoryBeltelAll)
' data.frame': 157 obs. of 39 variables:
$ ID           : int 312 313 314 315 316 317 318 319 320 321 ...
$ Organization : Factor w/ 4 levels "Banking A","Banking B", ...
3 3 3 3 3 3 3 3 3 3 ...
$ ProjectID    : logi NA NA NA NA NA NA ...
$ ProjectDescription : Factor w/ 489 levels "&V: Renewal Search Engine (GICT)" ...
...: 95 66 123 471 262 122 80 460 302 116 ...
$ Size         : num 17 24 1089 341 238 ...
$ Duration     : num 13.74 15.42 24.16 14.45 4.13 ...
$ Cost         : num 113040 382790 1600000 241400 140000 ...
$ CostPerFP    : int 6649 15950 1469 708 588 563 3127 5431 3195 2600 ...
$ DaysPerFP    : num 24.6 19.55 0.68 1.29 0.53 ...
$ DefectsPerFP : num NA NA 0.13 NA NA 0.07 0.24 NA NA ...
$ CostDurationIndex : num 78.1 57.9 87.8 92.6 97.8 ...
$ Quadrant     : Factor w/ 4 levels "Bad Practice", ...
2 2 ...
$ EQFCost      : logi NA NA NA NA NA NA ...
$ EQFDuration  : logi NA NA NA NA NA NA ...
$ MultiApplicationRelease : logi NA NA NA NA NA NA ...
$ ReleaseBased : logi NA NA NA NA NA NA ...
$ OnceOnlyProject : logi NA NA NA NA NA NA ...
$ Single.application : logi NA NA NA NA NA NA ...
$ Dependencies.with.other.applications : logi NA NA NA NA NA NA ...
$ Migration.Project : logi NA NA NA NA NA NA ...
$ PackageOffTheShelf : logi NA NA NA NA NA NA ...
$ SteadyHeartbeat : logi NA NA NA NA NA NA ...
$ FixedExperiencedTeam : logi NA NA NA NA NA NA ...
$ BusinessDriven : int 1 1 1 0 1 1 1 1 1 1 ...
$ RulesAndRegulationsDriven : int 0 0 0 1 1 0 0 0 0 1 ...
$ BusinessDomain : Factor w/ 11 levels "Billing","Call Center Solutions", ...
...
7.2.7 Descriptive statistics of all DutchCo projects in the EBSPM repository

```r
> str(RepositoryDutchcoAll)
data.frame: 4 obs. of 39 variables:
$ ID : int 446 447 491 492  
$ Organization : Factor w/ 4 levels "Banking A","Banking B",..: 4 4 4 4  
$ Project ID : logi NA NA NA NA  
$ Project Description : Factor w/ 489 levels 
$ Size : num 335 277 1233 1261  
$ Duration : num 3.78 10.29 6.61 8.81  
$ Cost : num 125827 214111 814861 896788  
$ CostPerFP : int 376 773 661 711  
$ DaysPerFP : num 0.34 1.13 0.16 0.21  
$ DefectsPerFP : num 0.04 0.36 0.31 0.13  
$ CostDurationIndex : num 98.2 95.9 97.1 96.5  
$ Quadrant : Factor w/ 4 levels "Bad Practice",..: 3 3 3 3  
$ EQF Cost : logi NA NA NA NA  
$ EQF Duration : logi NA NA NA NA  
$ MultiApplicationRelease : logi NA NA NA NA  
$ ReleaseBased : logi NA NA NA NA  
$ OnceOnlyProject : logi NA NA NA NA  
$ Single.application : logi NA NA NA NA  
$ Dependencies.with.other.applications : logi NA NA NA NA  
$ SteadyHeartbeat : logi NA NA NA NA  
$ FixedExperiencedTeam : logi NA NA NA NA  
$ BusinessDriven : logi NA NA NA NA  
$ RulesAndRegulationsDriven : logi NA NA NA NA  
$ PilotOrPOC : logi NA NA NA NA  
$ YearGoLive : int 2015 2015 2015 2015  
$ BusinessDomain : Factor w/ 11 levels "Billing","Call Center Solutions",..: 4 4 4  
$ DevelopmentMethod : Factor w/ 3 levels "Plan-driven",..: 1 1 1 1 1 1 1 1 1 1  
$ StakeholderSatisfactionProcess : logi NA NA NA NA NA NA NA NA NA NA NA NA  
$ StakeholderSatisfactionProduct : logi NA NA NA NA NA NA NA NA NA NA NA NA  
$ PerceivedValueCustomer : logi NA NA NA NA NA NA NA NA NA NA NA NA  
$ PerceivedValueFinancial : logi NA NA NA NA NA NA NA NA NA NA NA NA  
$ PerceivedValueInternalProcess : logi NA NA NA NA NA NA NA NA NA NA NA NA  
$ PerceivedValueInnovation : logi NA NA NA NA NA NA NA NA NA NA NA NA  
$ PerceivedValueOverall : logi NA NA NA NA NA NA NA NA NA NA NA NA  
$ Defects : int 15 99 386 163  
$ ErrorsFirstMonth : int NA NA NA NA NA NA NA NA NA NA NA NA  
$ DevelopmentClassification : Factor w/ 4 levels "Conversion (<5% new)",..: 2 2 2 2  
```

7.2.8  Descriptive statistics of the EBSPM repository

```r
> str(EBSPM)
data.frame': 492 obs. of 39 variables:
  $ ID : int 1 2 3 4 5 6 7 8 9 10 ...
  $ Organization : Factor w/ 4 levels "Banking A","Banking B",...
  $ ProjectID : logi NA NA NA NA NA ...
  $ ProjectDescription: Factor w/ 489 levels "&V: Renewal Search Engine (Gl CT)"
  $ Size : num 47 75 128 131 93 15 107 56 48 469 ...
  $ Duration: num 4.9 3.81 7.48 4.38 4.19 ...
  $ Cost : num 489620 343250 989570 1115640 187880 ...
  $ CostPerFP: int 10417 4577 7731 8516 2020 24924 1983 8185 6862 238 ...
  $ DaysPerFP: num 3.17 1.55 1.78 1.02 1.37 ...
  $ DefectsPerFP: num NA NA NA NA 0.02 NA NA NA NA NA ...
  $ CostDurationIndex: num 77.9 90.8 80.7 79.9 95.8 ...
  $ Quadrant: Factor w/ 4 levels "Bad Practice",...
  $ EQFCost: logi NA NA NA NA NA ...
  $ EQFDuration: logi NA NA NA NA NA ...
  $ MultiApplicationRelease: logi NA NA NA NA NA ...
  $ ReleaseBased: logi NA NA NA NA NA ...
  $ OnceOnlyProject: logi NA NA NA NA NA ...
  $ Single.application: logi NA NA NA NA NA ...
  $ Dependencies.with.other.applications: logi NA NA NA NA NA ...
  $ Migration.Project: logi NA NA NA NA NA ...
  $ PackageOffTheShelf: logi NA NA NA NA NA ...
  $ SteadyHeartbeat: logi NA NA NA NA NA ...
  $ FixedExperiencedTeam: logi NA NA NA NA NA ...
  $ BusinessDriven: logi NA NA NA NA NA ...
  $ RulesAndRegulationsDriven: logi NA NA NA NA NA ...
  $ PilotOrPOC: logi NA NA NA NA NA ...
  $ BusinessDomain: Factor w/ 11 levels "Billing","Call Center Solutions",...
  $ DevelopmentMethod: Factor w/ 3 levels "Plan-driven",...
  $ StakeholderSatisfactionProcess: logi NA NA NA NA NA ...
  $ StakeholderSecurityProduct: logi NA NA NA NA NA ...
  $ PerceivedValueCustomer: logi NA NA NA NA NA ...
  $ PerceivedValueFi nancial: logi NA NA NA NA NA ...
  $ PerceivedValueInternalProc: logi NA NA NA NA NA ...
  $ PerceivedValueInnovation: logi NA NA NA NA NA ...
  $ PerceivedValueOverall: logi NA NA NA NA NA ...
  $ Defects: int NA NA NA NA 2 NA NA NA NA ...
  $ ErrorsFirstMonth: int NA NA NA NA NA NA NA ...
  $ DevelopmentClassification: Factor w/ 4 levels "Conversion (<5% new)",...
```

7.2.9  Example of the outcome of correlation testing between BelTel Cost and BelTel Size

```r
> cor.test(RepositoryBelTel$Cost, RepositoryBelTel$Size,
+          use="pairwise", alternative = c("two.sided", "less", "greater"),
+          method = c("spearman"),
+          exact = NULL, conf.level = 0.95, continuity = FALSE)
```
Spearman's rank correlation rho

data:  RepositoryBeltel$Cost and RepositoryBeltel$Size
S = 467.53, p-value = 9.442e-05
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.7360092

Warning message:
In cor.test.default(RepositoryBeltel$Cost, RepositoryBeltel$Size, :
  Cannot compute exact p-value with ties

7.2.10 Example of the outcome of Wilcoxon test between EBSPM Size and Repository Size

> wilcox.test(EBSPM$Size, Repository$Size)

Wilcoxon rank sum test with continuity correction

data:  EBSPM$Size and Repository$Size
W = 7367.5, p-value = 0.1917
alternative hypothesis: true location shift is not equal to 0

7.2.11 Outcome of hist(Repository$Cost)

> hist(Repository$Cost)

7.2.12 Outcome of hist(Repository$Size)

> hist(Repository$Size)
7.2.13 Plot of Size versus Cost

```r
> plot(Repository$Size, Repository$Cost)
```

7.2.14 Plot of Duration versus Size

```r
> plot(Repository$Size, Repository$Duration, 
+      pch = c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,16,16,16,16), 
+      xlab = "Project Size", ylab = "Project Duration")
```
7.2.15 Plot of RepositoryBeltelAll$Size, RepositoryBeltelAll$Duration

> plot(RepositoryBeltelAll$Size, RepositoryBeltelAll$Duration)

7.2.16 Plot of RepositoryDutchcoAll$Size, RepositoryDutchcoAll$Duration

> plot(RepositoryDutchcoAll$Size, RepositoryDutchcoAll$Duration)
7.2.17 Plot of Repository\$Size, Repository\$Stakeholder\$SatisfactionProcess

> plot(Repository$Size, Repository$Stakeholder$SatisfactionProcess)

7.2.18 Plot of Repository\$Size, Repository\$Stakeholder\$SatisfactionProduct

> plot(Repository$Size, Repository$Stakeholder$SatisfactionProduct)
7.2.19 Plot of Repository\$Size, Repository\$PerceivedValueOverall

```r
plot(Repository$Size, Repository$PerceivedValueOverall,
    pch = c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,16,16,16,16),
    xlab = "Project Size", ylab = "Perceived Value Overall")
```

7.2.20 Shapiro Wilk Normality test

```r
shapiro.test(Repository$Cost)
```

```
Shapiro-Wilk normality test
```

---

The Effects of Perceived Value and Stakeholder Satisfaction on Software Project Impact
data:  Repository$Cost
W = 0.60549, p-value = 3.423e-07

> shapiro.test(Repository$Size)

Shapiro-Wilk normality test
data:  Repository$Size
W = 0.54617, p-value = 7.492e-08