Providing Tool-Support for Value-Based Decision-Making: A Usability Assessment

Vitor Freitas, Emilia Mendes, Burak Turhan
M-Group@M3S
Faculty of Information Technology and Electrical Engineering
University of Oulu
Oulu, Finland
{vitor.freitas, emilia.mendes, burak.turhan}@oulu.fi

***Abstract*— Numerous companies worldwide make their decisions related to software projects/products in a value neutral way, using only earned value systems, which represent short-term goals. Better decisions can be made using a value-based approach, achieving cost-effective results and reliable construction and maintenance of products. However, moving from a value-neutral to a value-based paradigm can be a challenge. We provide tool-support, which was co-created in collaboration with three software companies, to ease the paradigm shift. Our tool supports both individual and group-based decisions using several visualization mechanisms. Despite the co-creation process employed while developing the value tool, there are specific issues relating to its usability that must also be assessed in order to reduce any possible drawbacks for its adoption by industry. This paper details three usability studies that were carried out to assess the value tool's usability. The results also suggest that the tool is ready to be taken into use in the industry.**

Keywords—Value-based decision making; Value-based software engineering; tool; usability

I. INTRODUCTION

The software advent and its adoption in industry dramatically changed the way companies do business nowadays bringing new possibilities and better management. As a side effect, companies depend on multiple software solutions from operational level to strategic level in order to survive and to remain competitive. A business is a living organism, changing and evolving constantly, hungry for new technologies and software solutions, directly impacting upon the software industry, which in turn, works on its full potential to feed these businesses' needs [9].

As a result of the dynamic reality of enterprise businesses, many software projects are developed under time pressure dealing with limited budget and stakeholder conflicts. Decisions relating to the selection and prioritization of features/requirements/technologies are largely done using cost as the primary driver in a value neutral setting [1]. For most of the software companies the cost represents their reality and dictates what needs to be completed/delivered to achieve their goals. In a domain plagued by uncertainty, cost estimation is judged as a safer parameter upon which to base decisions [13]. This occurs because cost is a tangible and numerical measure, which can be employed to support decisions, via, for example, the comparison between available budget and the implementation costs of a set of features. In addition, it also differs from other less tangible factors, such as customer satisfaction or innovativeness, which are more subjective and difficult to measure, however play a very important role in decision-making. Overlooking such factors means ignoring long term value aspects that could enable better decisions to be taken, and in many cases even saving software projects from failure [1]. Many companies ignore those value aspects in their decision-making, and those that already use such aspects, do not use them in a systematic way [1].

The paradigm shift from a value-neutral decision-making to a value-based decision-making represents a challenge for software companies, because it involves consolidating views of value considerations by different stakeholders, modelling and quantification of uncertainty and utilizing decision-making models towards knowledge creation [2].

To ease such paradigm shift, we have developed a tool to support value-based decision-making within the context of product management, where products can either be themselves software, or use software intensive systems. This tool was co-created with three companies through an iterative process, using design-science research method, involving the elicitation of tool requirements, tool implementation and tool assessment. The design-science research addresses important unsolved problems in unique or innovative ways, and is suitable for the value tool development because it differentiates from a routine design, addressing a unsolved problem in Information Systems [3]. This process of co-creation was fundamental so to understand the companies' current value-based decision-making processes, and to offer a solution that would match as closely as possible their existing process. Further, such solution is in itself a contribution to the body of knowledge in value-based software engineering [1].

The expected benefits of the proposed value tool are many-fold: (1) representation of value considerations by different stakeholders; (2) explicit representation of value; (3) support for decision-making meetings; (4) individual assessment of stakeholders' decisions that can be done synchronously during a meeting or asynchronously before the decision-making meeting actually starts; (5) a rich dashboard with consolidated data from all stakeholders' assessments that can be used to compare, discuss and support group decisions; (6) means to record stakeholders' decisions and document past decisions; (7)

promote transparency on decisions; (8) promote organizational learning.

The value tool is the outcome of the design-science research, i.e., a purposeful IT artifact created to address value-based-decision making in software-intensive industry. The usability assessment of the value tool is part of the design-science research cycle. Although the value tool has been co-created and informally evaluated with three companies, there is also the need to formally assess the value tool's usability, so to minimize any possible barriers to its full adoption in industry. Therefore, we have conducted three usability case studies to assess and improve the value tool's usability. The first case study was a pilot run with six doctoral students in order to test and calibrate our usability measurement instruments and process. The second study was conducted with 27 graduate students as end-users. The third study was conducted with seven usability experts.

The research goal of this work is to formally evaluate the value tool in respect to its usability and detail these case studies, identifying its weaknesses in an early stage to improve the artifact design.

The remaining the paper is structured as follows: Section II we describe the value tool itself, its functionalities, the scenarios of use, the terminology and the value-based decision-making process supported by the tool. In Section III we present the research methods as well as the materials we used in the case studies. In Section IV we present and discuss the results. Finally, in Section V we present the conclusions and future works.

II. BACKGROUND

The solution we provide is a Web-based tool to be used by the key stakeholders participating in decision-making meetings relating to managing products, where products are either software or use software intensive systems. Because this tool (see Fig. 1) can be used in many different decision-making scenarios we adopted the terms **deliverable** to describe a product, a project, a release or a service, and **decision item** to describe a feature, a requirement, a use case, a bug, an idea or anything a decision is targeted at. For example, a deliverable can be a software product, and the decision items can be possible features to add into the product's next release. The explicit knowledge from stakeholders about what they understand as value is represented in the value tool via a set of company-specific **value factors**. The identification of value factors is an important step prior to using the value tool, and such identification can be achieved via qualitative as well as quantitative methods (e.g. questionnaires, interviews). Finally, each value factor is associated with a **measure**, which identifies how decision items will be assessed against each value factor. For example, assuming a decision item to be 'Feature A', and a value factor to be 'Customer Satisfaction', and measured using positive/neutral/negative impact. This means that when stakeholders are deciding upon each Feature, they will assess the impact (measure) that the implementation of the feature (decision item) will have upon Customer Satisfaction (value factor). This process is also illustrated in Fig 1.

At the top level, we have a repository of **deliverables** and each deliverable has a set of **stakeholders** and **decision items**. One of the stakeholders is the deliverable's manager and is responsible for setting up in the tool the data needed for the decision-making **meetings** (e.g. loading the information about each feature to be discussed, assigning stakeholders to a

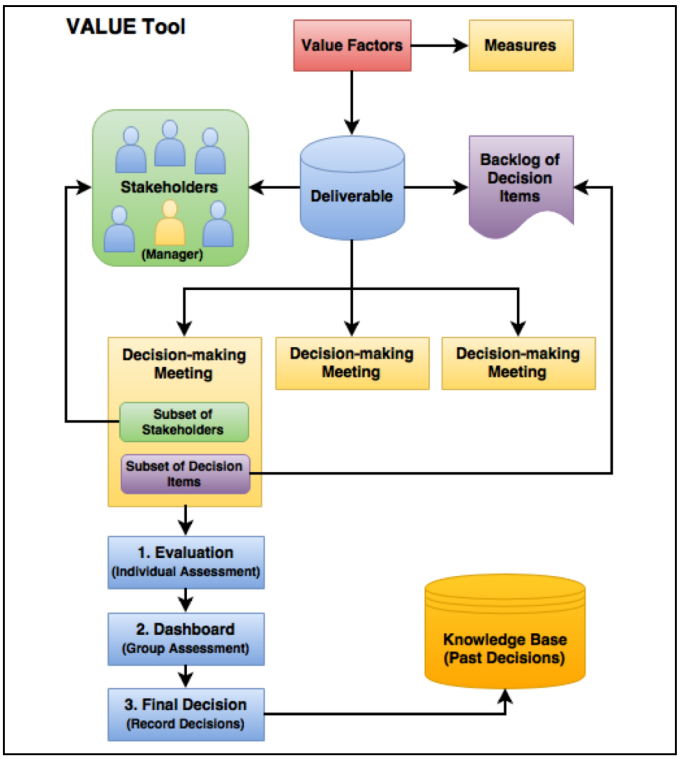

Fig. 1 - Value-based decision-making through the VALUE tool

decision-making meeting). The value tool organizes all the decision-making activities into meetings, i.e., decisions are stored per meetings. This means that a given deliverable may have several meetings associated with it. Each of the meeting is composed by stakeholders, decision items, and value factors & measures. During a meeting, stakeholders follow three sequential steps:

1. **Individual assessment (evaluation)**: The attending stakeholders input their individual assessment for each of the decision items being discussed in the meeting. The assessment is done using the value factors and the associated measures. In this step stakeholders can also input the rationale of their assessment.

2. **Group assessment (dashboard)**: After the individual assessment a rich dashboard (Fig. 2) is constructed aggregating all the stakeholders' assessments. It provides means to visualize all the stakeholders' opinions, compare views from different groups of stakeholders (e.g. Sales opinion against Technical option), compare the impact of the value factors, compose scenarios of decision items (e.g. a subset of five features with the best impact on the revenue growth against the five features with the best impact on the customer satisfaction). The dashboard is useful to guide group discussions as the value tool can identify conflicting views (e.g. a group of stakeholders assessed a given feature with an overall positive impact on the revenue growth at the same time other group assessed

with an overall negative impact), saving time going straight to the decision items that deserve more attention. In the data visualization illustrated by Fig. 2, the bar chart in the left shows the aggregated assessments of a given feature regarding to all the value factors used during the meeting. The radar chart in the right shows the same data but aggregated by the group of value factors (e.g. Customer group, aggregating Customer Satisfaction, Customer Retention and Customer's Perspective). The heat map chart in the background shows the value factors that contributed to the most for the positive, neutral and negative impacts of a given feature.

3. **Record decisions (final decision)**: The final step of a decision-making meeting is documenting the group's decision, which can be the prioritization or selection of the decision items discussed. All past decisions are stored in the tool, and can be used to support future decisions, and to go over past decisions too.

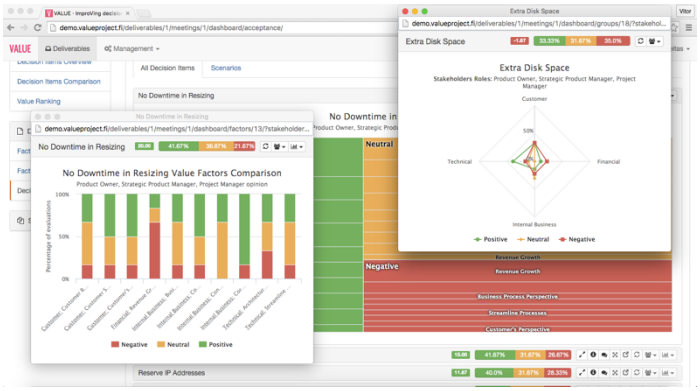

Fig. 2 - Decision support through data visualization

The development of the value tool started with a simple prototype representing our views of what would be a software solution to tackle the challenges of a value-based approach on decision-making. Once this prototype was completed, there were a series of meetings with the one of the industrial partners, later followed by other meetings with the two remaining industry partners. The focus of each meeting was to demonstrate the tool, and obtain feedback on additional functionality needed, usability improvements, and changes to existing functionality. In total, 16 meetings took place, adding to a total of 44 person-hours. All participants were experienced product owners or managers, and experienced project managers.

III. Materials and Method

The ISO Standard 9241 defines usability as "The **effectiveness**, **efficiency** and **satisfaction** with which specified users achieve specified goals in particular environments" [4]. The perceived usability of a software system has a significant impact on its reputation and customer satisfaction [5]. A study carried out in the domain of Web retailers demonstrated the general usability of a Web site plays an important role in the establishment of trust between the vendor and the customer [6]. In order to assess the usability of the value tool, easing its adoption by our business partners, three usability studies were carried out aiming at measuring its effectiveness, efficiency and user satisfaction.

Three usability case studies with 40 participants in total were carried out. The first study was a pilot study to assess the suitability of the instruments and mockup-example, in addition to making finer adjustments to the actual study's execution & process. The feedback from the pilot study was used to improve the instruments and mock-up example, which were later used in another two usability case studies. For the remainder of the paper the pilot study will be referenced as Study A, the second case study as Study B, and the third case study as Study C.

All the material used during the conduction of the case studies are available on http://valueproject.fi/usability/. Please refer to this link in case you want to see in great detail the contents of the questionnaires presented below.

A. *Questionnaires*

A pre-task and post-task questionnaires were prepared for the case study. In the pre-task questionnaire, we collected demographic data from the participants. They were asked to inform their age, gender, years of experience of software development in academia and industry, prior knowledge about Web development, usability and value-based decision making. The experience with software development was particularly important for us since all the subjects came from the academia and we wanted to analyze the data from those who had experience in industry separately, as their profile would closer to the end-users of the value tool.

The post-task questionnaire comprised questions related to the effectiveness and efficiency of completing the tasks that participants were asked to do during the case study. These tasks were related to the three most important functionalities of the tool, that is: the evaluation screen, the dashboard and the final decision screen. We also used a System Usability Scale (SUS) [7] to measure the level of usability of the tool. Even tough it is self-defined as a "quick and dirty" method, it is still nowadays the most used post-task questionnaire in usability studies, appearing in approximately 43% of them [8]. In the post-task questionnaire, the participants could also leave comments regarding the tasks and general comments about their experience with the tool.

The SUS contains 10 questions using a 5 points Likert scale (*1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree*) related to the usability and learnability of a software system, the original questions are presented below:

- *I think that I would like to use this system frequently.*
- *I found the system unnecessarily complex.*
- *I thought the system was easy to use.*
- *I think that I would need the support of a technical person to be able to use this system.*
- *I found the various functions in this system were well integrated.*
- *I thought there was too much inconsistency in this system.*
- *I would imagine that most people would learn to use this system very quickly.*

- *I found the system very cumbersome to use.*
- *I felt very confident using the system.*
- *I needed to learn a lot of things before I could get going with this system.*

The questionnaire was adapted as suggested by [10] and [11], but instead of changing the word "cumbersome" to "awkward", we preferred to replace it for "problematic" and every occurrence of the word "system" was replaced by "tool".

B. Subjects

Regarding Study A, six Software Engineering Ph.D. students agreed to participate in the case study. Two are female. Their mean age was 33 years old with a Standard Deviation (SD) of 4.39. Table I shows participants' experience with software development. Most of the participants had more than 5 years of academic experience with software development and between 3 to 5 years of professional experience. Table II presents participants' prior knowledge on Web development, usability and value-based decision-making. The participants could select more than one option in the questionnaire, so the sum of each column can be greater than 6 (100%). Most of the participants had prior knowledge about software development acquired through academic/professional experience, some usability knowledge acquired through reading and attending to seminars related to usability inspection and none knowledge about value-based decision-making.

TABLE I. STUDY A - PARTICIPANTS EXPERIENCE WTH SOFTWARE DEVELOPMENT

	Experience with Software Development	
	Academic Experience	***Professional Experience***
None	1	1
Less than 1 year	0	1
Between 1-3 years	0	1
Between 3-5 years	2	2
More than 5 years	3	1

TABLE II. STUDY A - PARTICIPANTS PRIOR KNOWLEDGE

	Prior Knowledge		
	Web development	***Usability***	***Value-based Decision-making***
None	1	1	4
Some knowledge acquired through reading/seminars	0	4	1
Academic experience	4	2	1
Industrial experience	4	1	0

In relation to Study B, 27 graduate students attending the course Software Engineering Research participated in the case study. Note that this study was carried out as part of some additional lectures, and therefore students' attendance was an expectation. Six participants are female. Their mean age was 26.65 years old (SD = 6.43). The mean age was calculated based on the information provided by 26 participants, as one left it missing. Most of the participants had between 1 to 5 years of experience with software development in academia and none or less than 1 year of experience in industry (see Table III). Regarding to the participants' prior knowledge on Web development, usability and value-based decision-making (Table IV), most of the participants had knowledge about Web development acquired through reading or attending to seminars, academic experience with usability and none or some knowledge about value-based decision making acquired through reading or attending seminars.

TABLE III. STUDY B - PARTICIPANTS EXPERIENCE WTH SOFTWARE DEVELOPMENT

	Experience with Software Development	
	Academic Experience	***Professional Experience***
None	2	14
Less than 1 year	5	6
Between 1-3 years	8	3
Between 3-5 years	8	0
More than 5 years	4	4

TABLE IV. STUDY B - PARTICIPANTS PRIOR KNOWLEDGE

	Prior Knowledge		
	Web development	***Usability***	***Value-based Decision-making***
None	4	2	12
Some knowledge acquired through reading/seminars	12	12	12
Academic experience	9	14	1
Industrial experience	9	3	3

Study C was carried out within the context of a Usability Course. Seven students, who are already knowledgeable on usability, participated in the study. Note that participation in the study was compulsory for these students, as this was part of their coursework. One participant is female, and participants' mean age was 26.14 years old (SD = 5.75). Most of the participants' experience with software development came from the academia background (see Table V). Most of them had prior knowledge about web development. All of the participants had practical experience with usability inspection in academia or industry, and some knowledge about value-based decision making acquired through reading or attending to seminars.

TABLE V. STUDY C PARTICIPANTS EXPERIENCE WTH SOFTWARE DEVELOPMENT

	Experience with Software Development	
	Academic Experience	***Professional Experience***
None	0	3
Less than 1 year	2	3
Between 1-3 years	3	1
Between 3-5 years	1	0
More than 5 years	1	0

TABLE VI. STUDY C PARTICIPANTS PRIOR KNOWLEDGE

	Prior Knowledge		
	Web development	***Usability***	***Value-based Decision-making***
None	1	0	2
Some knowledge acquired through reading/seminars	3	3	4
Academic experience	4	5	1
Industrial experience	0	2	0

C. Procedure

In order to make the process of using the tool more comfortable for the participants, we elaborated a mock-up scenario where all the participants would play the role of a stakeholder of a software product called MusicFone – a music playing application for mobile phone [14]. The description of the scenario contained details about the application and context information such as release date, development schedule and team size. It also had a list of 11 features, where the participants would use the value tool to select four of them for the next release, using the following value factors: Customer Satisfaction, Delivery Time, Market Competitiveness and Revenue Growth. We also provided a description of each value factor and discussed about them during the case study, to make sure the participants would have a similar understanding of what each of the value factors was about.

Further, participants were also given a document containing tasks that they were supposed to do when using the tool. This document also contained practical instructions such as the tool's URL and a username and password to access the tool. The tasks were divided into two blocks: individual tasks and group tasks. The individual tasks asked the participants to open the MusicFone product in the tool, access the ongoing decision-making meeting and input their individual assessments using the **Evaluation** screen and to also add at least two rationales for their evaluations. During the group tasks, the participants were asked to use the **Dashboard** to analyze the aggregated data and discuss about the features to decide upon the 4 most valuable features to be implemented in the next release of the MusicFone product. After reaching a consensus, participants would access the **Final Decision** screen to record the meeting's decision. Participants were also asked to create a report containing the graphs that they believe contributed the most towards helping them decide upon the four top features to select.

TABLE VII. STEPS EXECUTED DUING THE CASE STUDIES

Step	Activity	Time (mins)
1	Provide print-out and give an overview of activity	15
2	Distribution, answering and gathering of pre-questionnaire	10
3	Distribution of MusicFone's and value factors descriptions plus reading time	15
4	Answering any doubts on features' and/or value factors' descriptions	5
5	Distribution and execution of individual tasks	20
6	Tell participants to familiarize themselves with the data visualizations provided in the dashboard	10
7	Each group manager leads discussion on group tasks	30
8	Distribution, answering and gathering of post-questionnaire	10

The individual tasks and group tasks were synchronized during the case study. All the participants had to finish the individual tasks before moving to the group tasks. In Studies A and C we used only one group where one of the participants would play the role of the meeting manager and the rest would play the role of a regular stakeholder. For Study B, as the number of participants was higher, participants were divided into groups of 2-3 persons each. The meeting manager was selected randomly among the participants, and their task scripts had additional tasks so to gather all the stakeholders together and guide the discussion using the dashboard.

All the steps we executed during the case study are presented in Table VII, and an estimated time for each of the activities was validated during the Pilot study.

IV. RESULTS AND DISCUSSION

A. Study A

The ten questions from the SUS Questionnaire were answered by the participants using a 5 points Likert Scale. To avoid biased answers, questions 2, 4, 6, 8 and 10 are reverse-scored, i.e., the median closer to 1 the better. Table VIII presents the results from the SUS Questionnaire, showing a high level of agreement among the participants that the tool is consistent (item 6, *median = 1*), easy to use (item 3, *median = 4*), well integrated (item 5, *median = 4*), quick learnability (item 7, *median = 4*) and felt confident using the tool (item 9, *median = 4*). Results also suggest that participants stayed neutral to answering if they would like to use the tool frequently (item 1, *median = 3*) or not. A possible explanation could be participants' lack of experience with value-based decision-making (see Table II) and also a lack of industrial background. Most of the participants chose 'neutral' to the question *"I needed to learn a lot of things before I could get going with this tool"* (item 10, *median = 3*). Given that they were exposed to just one part of the tool's functionality, they might have felt unsure as to what to answer, and thus settled for 'neutral'.

TABLE VIII. STUDY A - SUS QUESTIONNAIRE RESULTS

ID	Items	Median
1	I think that I would like to use this tool frequently	3
2	I found the tool unnecessarily complex (reverse score)	2
3	I thought the tool was easy to use	4
4	I think that I would need the support of a technical person to be able to use this tool (reverse score)	2
5	I found the various functions in this tool were well integrated	4
6	I thought there was too much inconsistency in this tool (reverse score)	1
7	I would imagine that most people would learn to use this tool very quickly	4
8	I found the tool very problematic to use (reverse score)	2
9	I felt very confident using the tool	4
10	I needed to learn a lot of things before I could get going with this tool (reverse score)	3

In addition to the SUS Questionnaire, participants were supposed to answer questions related to tasks' **ease of completing** and **amount of time for** tasks representing important functionalities of the value tool, i.e., the evaluation screen, dashboard and final decision screen. Results (see Table IX) suggest that participants were satisfied with the ease of completing and the amount of time it took for the tasks related

to the Evaluate screen (*median = 4*). On the other hand, participants stayed neutral (*median = 3*) regarding the Dashboard and the Final Decision screen tasks.

TABLE IX. STUDY A - ASSESSMENT OF INDIVIDUAL AND GROUP TASKS

Items	Median
I was satisfied with the **ease of completing** this task using the VALUE Tool **Evaluate** screen.	4
I was satisfied with the **amount of time** it took to complete the task using the **Evaluate** screen.	4
I was satisfied with the **ease of completing** this task using the **Dashboard** screen.	3
I was satisfied with the **amount of time** it took to complete the task using the **Dashboard** screen.	3
I was satisfied with the **ease of completing** this task using the **Final Decision** screen.	3
I was satisfied with the **amount of time** it took to complete the task using the **Final Decision** screen.	3

B. Study B

Table X presents the aggregated data from all the 27 subjects for the SUS Questionnaire. In this case study the number of participants was higher and some of them also had industrial experience with software development (note that their overall median values are presented separately). The question with the best result was item 4 (*median = 1*). The item 8 (*median = 1*), suggests that the tool was found easy to use and participants felt that they would not need technical support to be able to use the tool. Further, item 1, *"I think that I would like to use this tool frequently"*, had a good result (*median = 4*). It differentiates from the Study A because in the Study B 59.2% (16 out of 27) of the participants had some knowledge about value-based decision-making and 48.1% (13 out of 27) of the participants had previous experience in industry.

TABLE X. STUDY B - SUS QUESTIONNAIRE RESULTS

#	Items	Median (no industrial experience)	Median (industrial experience)
1	I think that I would like to use this tool frequently	3.5	4
2	I found the tool unnecessarily complex	2	2
3	I thought the tool was easy to use	4	4
4	I think that I would need the support of a technical person to be able to use this tool	2	1
5	I found the various functions in this tool were well integrated	4	4
6	I thought there was too much inconsistency in this tool	2	2
7	I would imagine that most people would learn to use this tool very quickly	4	4
8	I found the tool very problematic to use	1	1
9	I felt very confident using the tool	4	4
10	I needed to learn a lot of things before I could get going with this tool	2	1

The tool's perceived level of usability did not change much when comparing both groups. Table XI presents the results related to the individual and group tasks. The overall result was good, with the least median of 4, suggesting that the majority of the participants was satisfied with both the ease of completing and the amount of time that took to complete each of the tasks using the value tool. The amount of time regarding the 'Evaluate screen' received a median of 5, which is a good indicator that the tool did not require much time of the participants to input their evaluations.

TABLE XI. STUDY B - ASSESSMENT OF INDIVIDUAL AND GROUP TASKS

Items	Median
I was satisfied with the **ease of completing** this task using the VALUE Tool **Evaluate** screen.	4
I was satisfied with the **amount of time** it took to complete the task using the **Evaluate** screen.	5
I was satisfied with the **ease of completing** this task using the **Dashboard** screen.	4
I was satisfied with the **amount of time** it took to complete the task using the **Dashboard** screen.	4
I was satisfied with the **ease of completing** this task using the **Final Decision** screen.	4
I was satisfied with the **amount of time** it took to complete the task using the **Final Decision** screen.	4

Also important to note that the questions were related with the tasks that participants were asked to perform, as per the Tasks Script document. The results presented in the Table XI suggests that the main functionalities of the tool can be used with little effort and without spending too much time.

The analysis of 1188 database records of the features' assessment using the Evaluation screen provided the data used to calculate the average time to assess a feature. The average time represents the interval between each evaluation done in the tool by a given participant, during the Evaluation phase (individual assessment). A single evaluation in the tool means: for a given decision item (e.g. a feature), assess the impact (e.g. positive, neutral or negative) of the implementation of this feature based on a value factor (e.g. customer satisfaction or revenue growth).

The average time of a single evaluation, based on the data gathered by this experiment, is 14.01 seconds (SD = 36.47) per evaluation.

C. Study C

In the last case study, where participants had knowledge and experience with usability inspections, the items with the best result were items 6 and 8, both with median of 1, showing a consensus among the participants that the tool was consistent and not problematic to use (see Table XII). The results suggest a good perceived level of usability of the tool, because the smallest median of the questions with reverse-scored was 2 and for the other questions it was 4.

The results regarding the ease of completing and the amount of time for each of the tasks are presented in Table XIII. The overall median was quite similar that for Study B, suggesting that all the main functionalities of the value tool are easy to operate and do not require much time. In Study C, the overall

satisfaction with the amount of time to complete the tasks in the Evaluate screen also received a median of 5.

TABLE XII. STUDY C - SUS QUESTIONNAIRE RESULTS

#	Items	Median
1	I think that I would like to use this tool frequently	4
2	I found the tool unnecessarily complex	2
3	I thought the tool was easy to use	4
4	I think that I would need the support of a technical person to be able to use this tool	2
5	I found the various functions in this tool were well integrated	4
6	I thought there was too much inconsistency in this tool	1
7	I would imagine that most people would learn to use this tool very quickly	4
8	I found the tool very problematic to use	1
9	I felt very confident using the tool	4
10	I needed to learn a lot of things before I could get going with this tool	2

TABLE XIII. STUDY C - ASSESSMENT OF INDIVIDUAL AND GROUP TASKS

Items	Median
I was satisfied with the **ease of completing** this task using the VALUE Tool **Evaluate** screen.	4
I was satisfied with the **amount of time** it took to complete the task using the **Evaluate** screen.	5
I was satisfied with the **ease of completing** this task using the **Dashboard** screen.	4
I was satisfied with the **amount of time** it took to complete the task using the **Dashboard** screen.	4
I was satisfied with the **ease of completing** this task using the **Final Decision** screen.	4
I was satisfied with the **amount of time** it took to complete the task using the **Final Decision** screen.	4

D. Discussion

An important point to note while analyzing the median score for the SUS Questionnaire's answers (see Tables VIII, X and XII) is that items 2, 4, 6, 8 and 10 are reverse-scored. The odd items are phrased positively while the even items are negative phrases. It is done this way to avoid extreme response bias. Even though some studies [12] suggests the alternate tone in the SUS Questionnaire does more harm than good, in these three studies we observed a consistency in participants' responses.

The statement *"I found the tool very problematic to use"* appeared in all three studies as one of the highest median. As it is reverse-scored (*1 = Strongly Disagree, 5 = Strongly Agree*), and with such a low mean it suggests the tool already provided a smooth experience for the end-user.

The analysis of the qualitative data provided by some of the participants in the comments section of the post-task questionnaire, suggested that some of the participants felt the dashboard visualizations were difficult to interpret. The reason of the perceived complexity might be associated with their lack of experience in value-based decision-making, and also the range of different types of visualizations provided in the Dashboard. In addition to the numerical data, participants could also leave comments about their experience using the tool. The following comments can support their thinking:

- *"It was clear, however more information about analyzing the graphs is beneficial".*
- *"Without guide, it will take some time to figure out the functions and the process of transfer 'ongoing' to 'analyze' is not obvious, it will take some time to learn how to operate".*
- *"Some graphs bring challenges to be compared and analyzed well".*

This last bullet point brought up an issue that can be addressed via some training, or a short tutorial. As part of the study we wanted the participants to explore the tool by themselves trying to complete the tasks in the tasks script, without showing them where to click and how to reach the desired results (e.g. add a rationale in the tool, compare two features in the dashboard).

Analyzing the comments left by the participants in all the three studies, some general issues were found. Some of the participants missed a save button in the evaluation screen (Fig. 3). Following, some of the comments:

- *"Save button for evaluation would be good".*
- *"Tell me you have saved my decision".*
- *"At first I didn't save my evaluation".*

This issue can easily be addressed by adding a "placebo save button" to the screen, as the evaluations are already saved on each of the user's interaction with the screen. A save button would then be added just to make the user feel safe.

The medians calculated from the questions related to the tasks executed by the participants during the experiment (Tables IX, XI and XII) suggests that for the most important functionalities of the tool that is, Evaluation screen, Dashboard and Final Decision screen, are easy to operate and requires a low amount of time.

E. Threats to Validity

In terms of threats to the validity of this research, the major issue was with the selection of subjects, as we used a convenient sample in all three studies, i.e., the university's graduate and Ph.D. students. The reason why we used students instead of practitioners was threefold: (i) our industrial partners are very busy and always working under time pressure; (ii) it is hard to arrange a meeting with a significant number of stakeholders to participate in a case study, as it demands almost 2 hours; (iii) we wanted to make the tool ready before they take it into use.

Regarding the construct validity of the case studies, we tried to minimize the possible issues adapting the SUS Questionnaire as suggested by [10] and [11], as no participant was a native English speaker and all the questionnaires were presented in English. Also the alternate tone in questions in the SUS Questionnaire is demonstrated to cause more confusions in the participants' answers than actually avoiding highly biased answers [12]. Even though the results cannot be generalized, the trends in the three case studies shows a good overall level of perceived usability by its users.

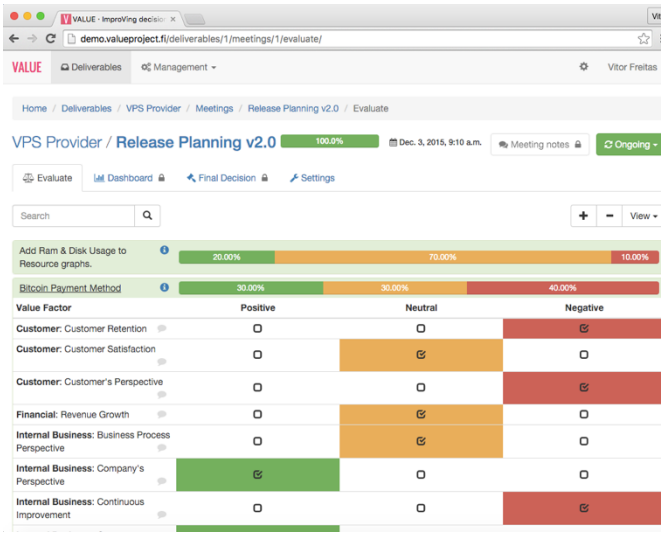

Fig. 3 - Evaluation Screen of the VALUE tool

V. Conclusions

In this paper three usability studies of the value tool were presented. The value tool was co-created with the participation of three software companies. The goal of the tool is to assist companies in the software domain to make a paradigm shift towards value-based decision-making, providing means to explicitly represent stakeholders' value considerations and also providing a mechanism to conduct meetings using the tool.

Based on the individual SUS items' medians, the strength in its usability is related to the tool being consistent and not being problematic to use. The weaknesses of its usability are related to unnecessary complexity. A way to tackle this problem is investing in training and providing help material before taking the tool into use. For the companies starting to use the tool, we have organized workshops using data from real products/projects. During the workshops we guide the stakeholders assisting them through all the steps in the tool, from individual assessment, dashboard analysis, how to interpret the data and how to record their final decisions.

Even though the SUS does not provide a diagnostic for the usability problems identified by its analysis, the comments left by the participants gave good insights on how to address some of the issues raised Several interface enhancements were done based on the participants' experience with the tool.

Using the qualitative data provided by the participants during the studies we learn that:

- The user interface of the evaluation screen (individual assessment) was smooth and comfortable to use.
- Even though the dashboard is visual clear and readable, some graphs bring challenge to be properly analyzed without guidance.
- The final decision screen is easy to use and self-explanatory. Also no participant had difficult using the screen to record the meeting decision.
- The "add rationale" functionality was not clear for a significant number of participants, needing more clarifications in the tool.
- Many participants stated that the tool helped reaching a fast consensus on the decisions

Also important to note that the main goal of these studies was to assess the usability of the value tool, as it plays an important role towards technology adoption. The results shown herein suggest that no major problem was identified. However, more empirical studies are needed, and planned to be take place as part of our future work in order to assess further the tool's usability.

Acknowledgment

The research project is funded by the Finnish Funding Agency for Technology and Innovation (Tekes), under the FiDiPro VALUE project.

References

[1] S. Biffl, A. Aurum, B. Boehm, H. Erdogmus, and P. Grünbacher, Value-Based Software Engineering. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006.

[2] E. Mendes, B. Turhan, P. Rodríguez, and V. Freitas, "Estimating the Value of Decisions Relating to Managing and Developing Software-intensive Products and Projects," in Proceedings of the 11th International Conference on Predictive Models and Data Analytics in Software Engineering - PROMISE '15, 2015, vol. 2015-Octob, pp. 1–4

[3] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information systems research," MIS Q. Manag. Inf. Syst., vol. 28, no. 1, pp. 75–105, 2004.

[4] ISO 9241-11:1998 Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability.

[5] L. Casaló, C. Flavián, and M. Guinalíu, "The role of perceived usability, reputation, satisfaction and consumer familiarity on the website loyalty formation process," Comput. Human Behav., vol. 24, no. 2, pp. 325–345, Mar. 2008.

[6] M. C. Roy, O. Dewit, and B. A. Aubert, "The impact of interface usability on trust in Web retailers," Internet Res., vol. 11, no. 5, pp. 388–398, 2001.

[7] J. Brooke, "SUS: A Quick and Dirty Usability Scale," in *Usability Evaluation in Industry*, London, United kingdom, 1996, pp. 189–194.

[8] J. Sauro and J. R. Lewis, Quantifying the user experience: Practical statistics for user research. Elsevier, 2012.

[9] A. McAfee and E. Brynjolfsson, "Investing in the IT that makes a competitive difference," *Harv. Bus. Rev.*, vol. 86, no. 7–8, 2008.

[10] K. Finstad, "The System Usability Scale and Non-Native English Speakers," J. Usability Stud., vol. 1, no. 4, pp. 185–188, 2006.

[11] A. Bangor, P. T. Kortum, and J. T. Miller, "An Empirical Evaluation of the System Usability Scale," *Int. J. Hum. Comput. Interact.*, vol. 24, no. 6, pp. 574–594, Jul. 2008.

[12] J. Sauro and J. R. Lewis, "When designing usability questionnaires, does it hurt to be positive?," Proc. 29th SIGCHI Conf. Hum. Factors Comput. Syst., pp. 2215–2223, 2011.

[13] B. Boehm and K. Sullivan, "Software economics: status and prospects," Inf. Softw. Technol., vol. 41, no. 14, pp. 937–946, Nov. 1999.

[14] D. Fucci, B. Turhan, N. Juristo, O. Dieste, A. T. Misirli and M. Oivo, "Towards an Operationalisation of Test-driven Development Skills: An Industrial Empirical Study", Information and Software Technology Journal, Vol.68, pp.82-97, 2015.