

# Managing Information on Goals and Benefits during Development Initiatives

Sander Krøglid<sup>1</sup> and Jo E. Hannay<sup>2</sup>[0000-0002-8657-7593]

<sup>1</sup> University of Oslo, Dept. of Informatics, Pb. 1080 Blindern, 0316 Oslo, Norway  
Sander Krøglid sandek@ifi.uio.no

<sup>2</sup> Simula Metropolitan Center for Digital Engineering, Center for Effective Digitalization of the Public Sector, OsloMet, Pb. 4 St. Olavs plass, 0130 Oslo, Norway  
johannay@simula.no

**Abstract.** The main reason for initiating a software development initiative is to produce value for stakeholders. However, software projects vary in their degrees of success in delivering the intended value. Since software projects can consume a significant amount of resources, a lack of focus on delivering value can result in bad investments. Research suggests that the absence of frameworks and techniques to support organizations in benefits management activities may be the reason for the lack of focus on value creation in daily development life. To increase the motivation for, and the focus on, benefits management, we implemented a generalized version of a set of published benefit estimation and tracking techniques so that they are accessible in an extensively used development management platform (Jira). The techniques are for declaring goals and goal structures and for expressing estimates of the system's contribution to the goals. The usability of the system was tested with IT professionals, the preliminary results are promising for further development on the present functionality and for other pieces of benefits management functionality in the future.

**Keywords:** Goal Structures · Benefits Management · Software Development · Project Management · Jira

## 1 Motivation

Development initiatives are often complex. When the public sector or a private enterprise develops portfolios of services for citizens or customers, a host of concerns must be taken into consideration regarding the design of the services, their integration with existing services, the cost of development, deployment and maintenance, the timespan and the benefits of the services and which goals are to be met by the service portfolio. If digitalization comes into the picture, as it inevitably does—not only in pure information systems, but also in systems with physical infrastructure such as roads and other constructions, further concerns regarding digital inclusion, universal accessibility and automated service provision come into play.

Managing development initiatives to success has proven hard; especially for large information-technological systems. Even under the auspices of agile development and management which explicitly put value for the customer in the high seat, there is a strong focus on the traditional control metrics of time, cost and scope to the detriment of a focus on the system’s benefits toward fulfilling goals [1]. One of the basic ideas in improving on this situation is that benefit should be estimated, tracked and measured with the same rigor as one routinely does with cost. Several techniques and methods have been presented for associating benefit estimates to the product elements of development initiatives [2,3,4]; specifically, *Large Scale Scrum* (LeSS) [5], *Scaled Agile Framework* (SAFe) [6,7] and *Benefit Points* [8,9,10,11,12]—an analogy to story points for cost estimates.

It is essential, however, that professionals are able to use these techniques readily in daily work life. In project management tools there is usually functionality for recording and tracking cost estimates, but not so for benefit estimates for tracking progress toward goals. To overcome this shortcoming, an endeavor was initiated with the aim to integrate *benefit estimation and tracking* functionality in commonly used development management tools. Although the basic ideas of benefit estimation and tracking are simple, their more advanced use in calculations of benefit/cost ratios and in prioritizing which product elements to develop first, can quickly become intractable in practice in the midst of the complexity of development work—a plight shared by benefits management in general [13]. The postulate of the endeavor is that this is, to a substantial degree, a *human interface problem*, due to the lack of integrative tools that keep the simplicity of the concepts intact and hide complex calculations from users.

## 2 Benefit Estimation and Goal Management

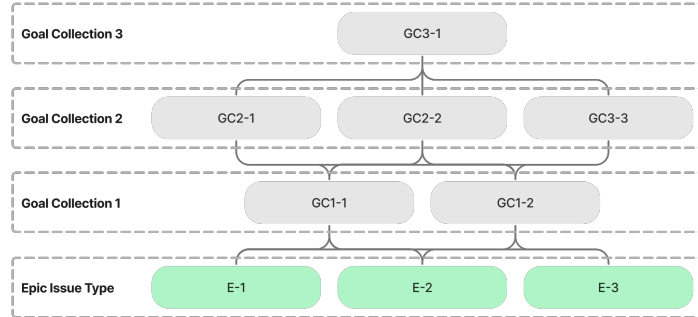
Here, we describe the design and development of a particular part of benefit estimation and tracking that concerns what we will call *Benefit Estimation and Goal Management* (BEGM). For full details, see [14]. The techniques in BEGM are generalizations of a core set of techniques for *Benefit/Cost-Driven Agile Development* (BCDAS) in [10], which is a compilation of ideas put forth in [8,9,12]. The BEGM functionality was implemented as an extension to Jira, a widely-used project management and issue tracking tool.<sup>3</sup>

Benefit estimation is more complex than cost estimation, in that benefit is assessed as the system under development’s potential to achieve a set of specified goals. Moreover, goals can be organized in tiered structures; examples being *Objectives and Key Results* (OKR) [15], *Lean Value Tree* (LVT) [16] and also more traditional project goal hierarchies.

Figure 1 illustrates the BEGM goal structure in the Jira extension (J-BEGM), together with Epics (which are represented in Jira proper). The Epics (green) are high-level user stories specifying integral pieces of system functionality (*minimum marketable features* [17] or *minimum viable products* [18]). Above the epics

---

<sup>3</sup><https://www.atlassian.com/software/jira>



**Fig. 1.** Product elements (epics) and goal structure of BEGM as implemented in the Jira extension

are goals in collections on various tiers. The curved lines represent possible relationships for benefit assessments. Thus, the epics are assessed by assigning benefit points for their potential effects on each goal in Goal Collection 1 (which could be project objectives or effect goals expressing the desired effects of the system on the stakeholders’ business and life processes). The goals in Goal Collection 1 can, in turn, affect the higher-tier goals in Goal Collection 2 (which could be organizational goals), and these can, in turn, affect higher-tier goals still (for example, societal goals). The effects that goals have on higher-tier goals are also expressed giving benefits points. One can have as many tiers as one wants, but one must have at least one goal tier for evaluating the benefit of epics.

The lines in the figure shows all possible benefit relations; for example, all goals in Goal Collection 1 can be given benefit points according to their relative contribution to every goal in Goal Collection 2. However, some goal structures might be hierarchical (such as OKR), and elements (epics or goals) on a tier would relate to a single goal on the next tier, rather than to many. In any case, the total benefit of a goal or an epic is the result of a weighted sum of the benefit points assigned to that goal or epic. The goal collection at the uppermost tier must be given benefit assessments directly, since they are not given benefit points via their contribution to higher-tiered goals.

Benefit points are relative money-agnostic estimates designed to express relative contributions, rather than absolute calculations of monetary value. Monetary value can, however be assigned to goals if desired.

Figure 2 shows an example of how this looks like in J-BEGM, where three goals EFFE-1, EFFE-2 and EFFE-3 have been declared in an uppermost goal tier. The left-hand screen shows benefit points (weights) assigned, while the right-hand screen shows the choice to assign monetary values.

### Set values

Method:  Weight  Monetary Value

Prefix: %

Total weight: 100/100

EFFE-1	EFFE-2	EFFE-3
25 %	40 %	35 %
25	40	35

Total weight must be 100 %  
All fields need to be filled

Cancel **Set points**

### Set values

Method:  Weight  Monetary Value

Prefix: M\$

Total value: 120 M\$

EFFE-1	EFFE-2	EFFE-3
25 M\$	40 M\$	55 M\$
25	40	55

Monetary value cannot exceed 999  
All fields need to be filled

Cancel **Set points**

**Fig. 2.** Setting benefit for the uppermost goal tier: weights (left), monetary value (right)

Goals can be expressed in nominal, ordinal, interval and ratio terms. Examples are, respectively, *Receive media attention as the enterprise most successful in digital transformation* (nominal), *Increase customer satisfaction by one median point* (ordinal), *Increase drop-out age for user proficiency in senior category (60yrs–90yrs) by 10 years* (interval), and *Reduce current average case processing time of 150 hours by 30%* (ratio). With benefit points, also achievement on qualitative (non-financial) goals can be measured. Qualitative goals are often the main reason for initiating development initiatives—especially in the public sector—even though it is the financial goals that are often used, together with the investment cost, to justify an initiative. Financial goals tend to be efficiency goals. It is therefore important to include qualitative goals explicitly for benefit assessment in order to balance the abundance of efficiency goals off with “true” goals; e.g., public well-fare goals [19,20].

Figure 2 (leftmost table) then shows three epics BN-1, BN-2 and BN-3, which have been assigned benefit points for each of the three objectives. The monetary

EPIC	EFFE-1	EFFE-2	EFFE-3
BN-1	50	28	37
BN-2	25	43	39
BN-3	25	29	24

BENEFI...	SOCI-1	SOCI-2	SOCI-3
Effect Goals	100/100	100/100	100/100
EFFE-1	19	12	14
EFFE-2	12	14	21
EFFE-3	14	21	12

**Fig. 3.** Assigning benefit points to epics: Project (left), portfolio of two projects (right)

values of the three objectives are shown in the leftmost (purple) tags under each objective; here, monetary values of 25, 40 and 55 million, respectively, which gives weights  $25/120 = 0.21$ ,  $40/120 = 0.33$  and  $55/120 = 0.46$ , respectively. The leftmost (blue) tags for each epic show the total number of benefit points assigned to that epic (115, 107 and 78, respectively), while the rightmost (purple) tags show the weighted sum of benefit points for the epic; e.g.,  $50 * 0.21 + 28 * 0.33 + 37 * 0.46 \approx 37$  for BN-1. Figure 2 (rightmost table) shows two projects in a portfolio with shared goals.

### 3 Design Approach

The endeavor of developing tool support for benefit estimation and tracking follows the following design principles:

**Concreteness:** The techniques should be designed for performing concrete tasks; a lack thereof will leave professionals in the dark as to what to do, even if they grasp the general idea of benefit estimation and tracking.

**Noninvasiveness:** The techniques should be designed to be used in the existing process flow. If methods are too complex or too invasive in day-to-day work, they will not be employed.

**Satisficing:** The techniques should be designed to be *good enough* for the tasks at hand and in line with *satisficing* [21], rather than optimizing.

**Support for cognitive processes:** The techniques should be designed to suit the nature of the cognitive processes involved in assessment [22,23].

**Recognizability:** The techniques should be reminiscent of existing techniques of state of practice to facilitate adoption.

The central tenet in observing these design principles is that the practitioners should concentrate on assessing one benefit relation at a time, in a relative manner. For example, one should give benefit points to each epic for its relative contribution compared to the other epics, to one goal at a time in Goal Collection 1 and then move on to the next goal. The same procedure should be used for the relations between Goal Collection  $i$  and Goal Collection  $i + 1$ . Note that the relations between different tiers may be the domains of different stakeholders. The relations can be assessed disjointedly, and at different times, and updated when new information is available. All the weighted sums will be automatically recalculated with any change. Thus, each task is simple and the tool should enforce this simplicity. The totality, however, is complex, and the tool should automatically calculate and recalculate the benefit points of each element according to whatever the individual assessments are. In the end, each epic receives a single benefit points value adjusted for all the goal assessments, which can be used in calculations of benefit/cost ratios and in Earned Business Value Management [9] and benefits uncertainty assessments [12].

## 4 Evaluation

The objective for the type of tools discussed in this article is that they will aid practitioners in performing benefits management activities; in particular, benefit estimation and tracking activities. To evaluate the degree to which such tools meet this objective, we pose the following two propositions:

**Proposition 1:** *Support for benefit estimation and tracking activities can be implemented in project management tools.*

**Proposition 2:** *The support for benefit estimation and tracking activities as implemented in project management tools are useful in practice.*

The present Jira extension (J-BEGM) is a first minimal marketable feature MMF [24].<sup>4</sup> Several other MMFs are planned which will extend the functionality in the present MMF. The two propositions above will be applied to all of them.

## 5 Study

For the the evaluation of the present Jira extension, we operationalized the propositions above as follows:

**Proposition 1 Operationalization:**

- *J-BEGM faithfully implements BCDAS techniques for managing goal structures*
- *J-BEGM faithfully implements BCDAS techniques of benefit estimation*
- *The terminology used in J-BEGM is consistent with that of BCDAS*

**Proposition 2 Operationalization:**

- *Product elements and goal structures used in development initiatives can be expressed in Jira*
- *The terminology is comprehensible*
- *The added functionality in Jira will motivate benefit estimation and tracking*
- *The added functionality in Jira will assist in benefits management*

### 5.1 Method

The tool is designed to be used throughout the lifecycle of a software system under development, and the best approach would be to evaluate J-BEGM in a natural setting. However, this would require a production-ready version of J-BEGM and a fully prepared development project, which was not feasible due to time constraints. Therefore, an artificial controlled setting was chosen for this evaluation.

---

<sup>4</sup>In incremental development, a MMF is arguably a more suitable formulation of a small integral part of value-adding functionality than a Minimum Viable Product (MVP).

Usability testing [25] was selected as the method for evaluation. The testing concerns how the application is experienced and understood during use and was facilitated by the first author. The specific method used was expert/task-based usability testing. The usability test has two parts, one for each proposition. In the first part, IT professionals are asked to perform a set of defined tasks in J-BEGM that involve using J-BEGM's most important techniques and features. In the second part, IT professionals are asked to test a goal structure that they are using or have used in a project recently. To assess the application's usability during the usability testing, participants are observed, and potential miss-steps and struggles are noted.

## 5.2 Participants

The evaluation on Proposition 1 requires participants who have knowledge and experience in benefits management and, optimally, experience with BEGM. Since BEGM is a part of the current development (a generalization of BCDAS), no persons had experience with BEGM, but we asked two of the authors behind BCDAS, as well as a person involved in early stages of the development, to assess how faithful J-BEGM implements the techniques and terminology of BCDAS.

The evaluation on Proposition 2 requires participants who are familiar with Jira and who have used goal structures in their work, but preferably with no knowledge of BCDAS or BEGM (or J-BEGM). For this evaluation it was not possible to gather participants with these qualities, and we had to use the same persons to evaluate on both propositions.

Thus, three experienced IT professionals were obtained as participants for the evaluation. All the participants were picked because of their experience with the original framework, benefits management, and their experience in using Jira.

## 5.3 Materials

To conduct the usability testing, it was important to have an application at a sufficiently high level of maturity in which the user could perform a set of tasks. Tasks were designed so the participants could utilize all the techniques implemented from the BEGM framework as experiences through the most important features of J-BEGM. The set of tasks (translated from Norwegian) for Proposition 1 is given in Fig. 4. For Proposition 2, oral instructions were given to declare a user-defined goal structure in J-BEGM, without further detailed instructions.

In a real-life scenario, estimating the benefit in terms of distributing benefit points between epics and goals and between goals in different goal tiers can be time-consuming, and it also requires knowledge about the different goals to assess. Since the first part of the usability test (for Proposition 1) aimed at evaluating how well techniques of BCDAS were implemented, and not how well the techniques fit in a real-life scenario, a set of data was given for the tasks (Fig. 4), which the user could insert. The data was derived from the examples in [10] with some adjustments. Further, because projects and epics are managed in Jira proper (not in the J-BEGM extension), Jira projects and epics were

1. Click on **Benefit Estimation** under *Apps*
2. Initiate the project by selecting:
  - (a) **Epic** as *the Product Element*, and
  - (b) **Todo**, and **In progress** as *Issue Statuses*.
3. Create **Objectives** as a *Goal Collection* to the **Benefit Calculator** projects's *Goal Structure*:

Name	Description
Objectives	The system's intended effects on stakeholders' work/life-processes

4. Declare the following three objectives:

Obj1	Reduce the average case processing time by 30%
Obj2	Reduce the number of wrong case decisions by 90%
Obj3	Reduce the average client-case processor interaction time by 70%

5. Set benefit points to estimate the benefit of the seven epics for each objective:

	Obj1	Obj2	Obj3
E1	16	13	14
E2	25	35	11
E3	25	10	12
E4	10	13	7
E5	3	5	35
E6	6	11	9
E7	15	13	12

6. Create a *Portfolio* named **The Benefits Management Agency**.
7. Connect the current project (**Benefit Calculator**) and the projects **dns.no** and **Benefits DB** to **The Benefits Management Agency** portfolio.
8. Create **Returns** as a *Goal Collection* to **The Benefits Management Agency's** *Goal Structure*:

Name	Description
Returns	The organization's long-term goals

9. Declare the following three returns (higher-level goals):

Ret1	Reduce number of man-hours
Ret2	Reduce number of compensations
Ret3	Improved public image of the organization

10. Set benefit points to estimate the benefit of the portfolio items (objectives) for each return:

Benefit Calculator	Ret1	Ret2	Ret3
Obj1	12	13	15
Obj2	9	11	6
Obj3	35	5	3

dns.no	Ret1	Ret2	Ret3
Obj1	7	13	10
Obj2	12	10	25

Benefits DB	Ret1	Ret2	Ret3
Obj1	11	35	25
Obj2	14	13	16

11. Set monetary values for each return
12. Go back to the project to see if the benefit points have changed

Fig. 4. Evaluation Tasks for Proposition 1



predeclared for the tasks. One of the tasks involved connecting three projects to a portfolio, and two projects were pre-declared for this task to save time.

The usability test was observed by the facilitator (the first author), and notes were taken when the participant struggled; e.g., if a participant struggled to perform a task, or if the participant asked to find a specific use case. As none of the participants had any experience using J-BEGM, some mis-clicks and wrong navigation were expected.

In addition to the observations by the facilitator, questionnaires (Google Forms) for the two parts of the usability test were given to the participants. The questionnaires addressed the operationalizations of the propositions directly; see Fig. 5 (translated from Norwegian). Further, to gather feedback on the application's general usability, *The System Usability Scale* was used [26]; see Fig. 5 (lowermost part).

Each question marked with \* was assessed on a scale from 1–5, where 1 signifies “Strongly Disagree”, while 5 signifies “Strongly Agree”.

**Questionnaire for Usability Test under Proposition 1**

1. To what degree are the techniques for goal structures reflected in the system?\*
  - Please explain your answer:
2. To what degree are the techniques for benefit estimation reflected in the system?\*
  - Please explain your answer
3. To what degree is the terminology in BCDAS reflected in the system?\*
  - Please explain your answer
4. Which aspects of the system did you appreciate?
5. Which aspects of the system can be improved?

**Questionnaire for Usability Test under Proposition 2**

1. To what extent can goal structures in your line of work be expressed in the system?\*
  - Please explain your answer:
2. To what extent can you perform benefits assessments that are relevant for your work in the system?\*
  - Please explain your answer:
3. To what extent is the terminology in the system comprehensible\*
  - Please explain your answer:
4. To what extent do you think the system can motivate the use of benefits assessments?\*
  - Please explain your answer:
5. To what extent do you think the system is helpful for assisting in benefits assessments?\*
  - Please explain your answer:

**The System Usability Scale**

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

**Fig. 5.** Usability Test Questionnaires

## 5.4 Procedure

The usability tests were arranged separately for each individual and split in two sessions; one for Proposition 1 and another for Proposition 2 and the SUS. One participant completed the test online via Microsoft Teams. Each session was scheduled to last about 60 minutes and was started by informing the participant about the focus for that session. When the participant was ready, the facilitator handed out the set of tasks and initiated the J-BEGM application with the appropriate context. If the participants experienced any confusion during the test, they were allowed to ask questions, and the confusions were then discussed at the end of the session.

In the first session, the participants were first instructed to initialize a project, set a goal collection with corresponding goals, and then estimate benefit points for the epics using the goals in the goal collection. Then, they were instructed to create a portfolio and connect it with the their project and two other projects declared by the facilitator. To test the estimation module's portfolio items mode, the participants were also instructed to create a goal collection with corresponding goals for the portfolio. Then, the participants could estimate the projects (portfolio items) using the newly created goal collection's goals. At the end, to test setting monetary value, the participants were asked to set some monetary value to the goals of the portfolio's goal collection. When the tasks were done, the participants were given the first questionnaire.

Then, the second part of the usability test was initiated, where the participant could freely use the application to set up a goal structure from a current or previous project they had knowledge about and assess it. When the participant felt done, or if the goal structure could not be reflected in J-BEGM, the second set of questionnaires were handed out, including the SUS.

At the end, an open discussion was initiated by the facilitator, where the participants could openly discuss their experiences and thoughts on J-BEGM. The participants were allowed to use the application further during the discussion to point out issues and for asking questions.

## 6 Results

We present the result for each proposition. Feedback from the questionnaire and the open discussions, including the notes from observation, will be discussed where relevant. In addition, some observed issues that do not relate to the propositions are discussed at the end. As the evaluation was conducted in Norwegian, the responses and feedback have been translated into English.

### 6.1 Results for Proposition 1

- *J-BEGM faithfully implements the BCDAS techniques for managing goal structures:* All the participants agreed that the BCDAS goal-managing techniques were faithfully implemented in J-BEGM. The scores on Question 1 in the

questionnaire for Proposition 1 (Fig. 5) given by the three participants were, respectively, 4, 5, and 5, resulting in an average score of 4.6. From the two participants who gave a rating of 5, there was no particular feedback other than “Don’t see anything missing”.

However, the participant who gave a rating of 4, expressed that the *Goal Tier* tab should have had a better introduction. However, once the participant got the hang of it, the flow increased for the rest of the session. The participant in question was, however, not alone in having issues with the *Goal Structure* and *Goal Tier* tabs. All participants had issues in finding where to insert the goals for the goal collection they created, although the other participants gave full scores on this question.

- *J-BEGM faithfully implements the BCDAS techniques of benefit estimation*: The participants also agreed that J-BEGM faithfully implemented BCDAS’s techniques for benefit estimation, as the scores given by the participants on Question 2 were the same as for the previous question. Two participants expressed joy in using the estimation table in J-BEGM. One of the participants mentioned the estimation table as one of the best aspects of the system: “I liked the estimation process the most”.

However, they had some suggestions for improvements. One of the participants suggested that the benefit estimates should be carried forward to Jira’s standard epic view. In the current state of J-BEGM, the benefit points are stored as a parameter for the issues but are not visible outside J-BEGM. Another participant suggested that, in addition to the table, a bar chart view for each goal (on which the epics are assessed) would be useful. According to the participant in question, this could perhaps visualize the differences between each epic to a greater extent than numbers in a table, and it would enable the user to distribute points to one goal before the next, which is the approach suggested by BCDAS [10, p. 24]. Beyond this, the participants seemed very satisfied with J-BEGM’s implementation of the techniques.

- *The terminology used in J-BEGM is consistent with that of BCDAS*: For the terminology, the participants were less satisfied; the scores given by the participants on Question 3 were 3, 4, and 5, resulting in an average score of 4. This is also reflected by the observations of completing the tasks and the participants’ justification of their scores. The terms that the participants commented on were *Goal Structure*, *-Tier*, *-Collection*, and *Portfolio Item*. Since these terms arose when generalizing the goal structure in BCDAS to BEGM, these are not fully consistent with the ones used in the BCDAS techniques in [10]. To improve on the situation, one of the participants suggested implementing profiles, or templates tailored to different contexts. The terms regarding the estimation module, however, were familiar to the participants.

## 6.2 Results for Proposition 2

- *Product elements and goal structures used in development initiatives can be expressed in Jira*: This was only answered by two of the participants. The scores on Question 1 in the questionnaire for Proposition 2 (Fig. 5) given by

the participants were 4 and 5, resulting in an average score of 4.5. One of the participants missed functionality for using non-financial goals in J-BEGM. The BCDAS framework supports setting monetary value to qualitative returns (non-financial gains) based on relative comparisons using a model for integrating soft and hard returns (MISHRI) [8][10, p. 31]. Since J-BEGM supports the setting of weights in addition to, or instead of, monetary value, combining financial and non-financial goals can be achieved using weights. Indeed, this generalizes the techniques for handling soft returns in [10], although this may not have been apparent to the participants. Other than that, both participants stated that they were able to reflect their goal structures in J-BEGM and assess on them.

- *The terminology is comprehensible*: As with the terminology statement for Proposition 1, the response on Question 3 for Proposition 2 on how the overall terminology of J-BEGM was understandable for more general use did not indicate satisfaction. The scores given by the participants were 3, 4, and 4, resulting in an average score of 3.6. Also here, the participants expressed confusion concerning the more general terminology introduced in the BEGM framework: “*Goal Tier* is a key concept and should have been given a more thorough introduction”.

One of the participants stated that the terms *Goal Structure*, *-Tier*, *-Collection*, and *Portfolio Item* seemed rather too generic and that these are not usually used in real-life contexts. Another participant argued that his lack of knowledge of the newly introduced terminology, which also increased his navigation problems, was linked to the fact that there was no introduction or explanation of these terms in J-BEGM. The participant suggested that J-BEGM could introduce and explain the new terms to make the system more understandable, and also suggested to combine the *Goal Structure* and *Goal Tier* tabs to avoid confusion.

- *The added functionality in Jira will motivate benefit estimation and tracking*. The response regarding whether J-BEGM can motivate to increase the use of benefit estimation and tracking received relatively high scores. The scores given by the participants on Question 4 were 4, 5, and 5, resulting in an average score of 4.6. However, one of the participants stated that the application’s current state might “require a user who is passionate about the idea”.

As the participants struggled with the introduced terminology from the BEGM framework, and to some degree with the goal managing practices as implemented in J-BEGM, one participant stated that there is a certain user threshold, but once overcome, the application should provide significant utility.

- *The added functionality in Jira will assist in benefits management* The response to Question 5 received the same scores as the previous question. In addition, Question 2 is relevant here, and the participants scored 4 and 5, resulting in an average of 4.5.

One participant enjoyed that J-BEGM was integrated with Jira, which enables a familiar workspace, where the backlog can be directly integrated. The latter was also expressed by another participant, who stated that the process enabled easier assessment than traditional Excel sheets, which offer no integration with Jira.

### 6.3 Observed issues

As stated previously, for some of the given tasks, all the participants struggled to navigate in J-BEGM to find where they could perform the tasks. Some of the most frequently occurring issues were declaring goals in a created goal collection, setting monetary value to the goals of a goal collection, and connecting portfolio items. The intuition of all participants was that goals might be created by clicking the action button of the *Goal Collection* in the *Goal Structure* tab. Only one of the participants was able to locate the *New Goal* button on their own. The rest of the participants needed guidance. The same intuition was at play when setting the monetary value of the goals. On this task, all participants needed guidance.

All the participants also needed guidance for connecting projects to portfolios. This involved a lot of navigation throughout the application. All participants checked the portfolio's *Goal Structure* tab, while some checked the project's *Settings* tab.

One of the participants, who conducted the usability testing through Microsoft Teams, had a low screen resolution, which resulted in a too-low browser height limit. Because of this, during the first estimation the participant had to scroll down, which hid the navigation bar making it difficult find the way back. One solution to this could be to use a fixed position on the navigation bar. The navigation bar is important for navigating through the different functionalities in the application.

### 6.4 The System Usability Scale

To calculate the System Usability Scale score (SUS score), an individual's ratings (scale positions) are converted into a *score contribution*. This is done as follows: For scores on an odd-numbered question (questions 1, 3, 5, 7, and 9), the score contribution is calculated by the scale position minus one. For even-numbered questions, the score contribution is calculated by five minus the score position. Then, the individual's score contributions are added and multiplied by 2.5. After doing this for each participant's answers, one can calculate the average score, which is the resulting SUS score. Figure 6 shows a spreadsheet used to calculate the SUS score of the usability test.

A score of above 68 is considered above average.<sup>5</sup> Our SUS score of 72.5 is slightly above average which is a positive result for the first evaluation of the application.

### 6.5 Results Summary

The evaluation results indicate mostly positive feedback regarding the statements in the propositions and for the usability of J-BEGM. However, the feedback provided by the participants identifies some aspects of J-BEGM that are

<sup>5</sup><https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>

Questions	Participant:	Scale Position			Score Contribution		
		A	B	C	A	B	C
I think that I would like to use this system frequently		5	3	3	4	2	2
I found the system unnecessarily complex		2	2	2	3	3	3
I thought the system was easy to use		4	4	2	3	3	1
I think that I would need the support of a technical person to be able to use this system		1	1	2	4	4	3
I found the various functions in this system were well integrated		5	4	4	4	3	3
I thought there was too much inconsistency in this system		2	2	1	3	3	4
I would imagine that most people would learn to use this system very quickly		4	2	3	3	1	2
I found the system very cumbersome to use		2	2	2	3	3	3
I felt very confident using the system		3	3	4	2	2	3
I needed to learn a lot of things before I could get going with this system		1	2	2	4	3	3
<b>Sum of score contributions</b>					<u>33</u>	<u>27</u>	<u>27</u>
<b>Sum multiplied with 2.5</b>					<u>82.5</u>	<u>67.5</u>	<u>67.5</u>
<b>Average score</b>							<u>72.5</u>

Fig. 6. SUS Results

confusing, and that can be improved in the next iteration of J-BEGM’s development. Nevertheless, since this was the first application evaluation, the results indicate that J-BEGM has the potential to motivate and be a helpful assistance in benefits management.

## 7 Discussion

Research, e.g., [1] and anecdotal evidence suggest that practitioners think that benefits considerations do not play an important enough role in project decisions and that considerations of time, cost, and scope take up too much space in comparison. Initiatives have been, and are, under way to guide digitalization agencies and funding bodies in integrating benefits considerations to a greater extent in governmental guidelines. The BCDAS framework has been given as part of a course for IT professionals over several years, but although participants have expressed enthusiasm over guidelines and frameworks, few practitioners have adopted the techniques in daily work; a tendency also observed years ago [13].

Making benefits management techniques accessible in Jira could enable practitioners to adopt and adapt benefits management activities into their daily work routines, and hopefully make them more likely to perceive themselves as successful in delivering benefits [27]. The current development is intended to be a first step in that direction for benefit estimation and tracking activities. Overall, the usability tests suggest that the Jira extension contributes to making techniques accessible in daily work, but improvements and enhancements must be done.

As stated by one of the participants, adopting benefits management practices with the Jira extension as it is now might require users who are, at the outset, passionate about the idea of estimating and monitoring benefits. Hence, passionate benefits management practitioners might be the *early adopters* of the application, while most organizations and their potential users may take the role of the *late majority* who might not use the application “unless everyone else in

the business does so”. Early adopters, on the other hand, might use an application even when it is not optimized in usability or not even fully functional [28]. To enable easier adoption and to encourage the late majority to use the J-BEGM, shortcomings and issues regarding the current state of the application must be resolved.

In addition to the shortcoming already mentioned, the participants suggested further development along the following lines:

*Templates and information views.* To assist navigation and to explain the concept in the tool, templates might help; e.g., for various goal structures (OKR, LVT and more traditional structures). Information views could be available on each tab to explain the concepts and their use.

*Enhanced visualization.* To help the user to distribute points for one goal at a time [10, p. 24] and to include an enhanced visualization of the differences in points distributed for each goal to be estimated, one participant suggested a view in which the estimation tables were split for each goal.

*Benefit-cost index and ranking of the backlog.* For now, the only way to see the benefit points is through J-BEGM, and there is no way of seeing them in the actual backlog or in Jira’s standard issue view. Therefore, to fully integrate the benefit points in the Jira backlog, there is still some work to be done. Since Jira has a ranking system for issues, the estimated benefit points could be used together with the story points to calculate a benefit-cost index [10, p. 29] to prioritize backlog elements based on benefit and cost estimates.

## 8 Threats to validity

We discuss threats to internal and external validity and reliability, as those most relevant for this usability study.<sup>6</sup>

### 8.1 Internal Validity

Internal validity in the context of usability testing refers to the degree to which the study design biases participants toward a certain response or behavior.

*Questions.* The questions in the questionnaires ask directly whether the application succeeds in fulfilling desired objectives. As the facilitator was present while some participants were filling out the forms, and because the participants knew that the facilitator would read their answers, a potential threat is that they gave a high score out of politeness (acquiescence bias, social acceptability bias, researcher bias, and interview bias). Another threat is that the first part of the usability test, along with the questionnaire, took much longer than anticipated. The questions were quite comprehensive, leading participants to investigate the application more thoroughly before providing answers. This could pose a threat to the evaluation’s internal validity as the work of participants regarding the first questionnaire might be tiresome, which could result in less motivation for the second part of the usability test.

<sup>6</sup><https://www.nngroup.com/articles/internal-vs-external-validity/>

*Tasks.* Because the participants were first instructed to complete a set of tasks, the participants inhabited knowledge and user experience before they were asked to express a goal structure of a current or previous project of their knowledge. The first usability test might ease the participants' ability to express their goal structure compared to a study without those tasks.

## 8.2 External Validity

External validity refers to the extent to which results apply to relevant behaviors or situations in the target audience, different from those applied in the study.

*Time.* Because the evaluation was scheduled to last about 60 minutes, it is important to note that the participants, although experts in benefits management and therefore expected to detect any deviations, might have overlooked some features that do not faithfully implement the goal managing and estimation techniques of BCDAS. Another consideration is that one of the participants took the test after a long day at work and found it hard to concentrate after completing the first part of the test. Further, that particular test instance was conducted online through Microsoft Teams, and the participant was provided access to the Jira test site where the tasks were performed. This enabled the participant to complete the questionnaire at will in spare time. This might have made it easier for the participant to be more thorough in his investigation. Because the response from the evaluation arrived some days after the initial session, the participant might have forgotten some aspects of the tool. While this also has bearings on internal validity (above), the fact that the study was conducted with somewhat primed participants in an artificial setting compared to a daily work situation might pose a threat to external validity.

*Use of goal structures.* For the second part of the study, participants were asked to evaluate the goal structure of a project they had previously worked on or were working on. However, one of the participants had no prior experience using goal structures in their work. Therefore, the second part of the usability test was skipped, and the first two questions of the second questionnaire were omitted since they required the participant to reflect on their work setting while using J-BEGM. This highlights the question as to whether J-BEGM should be used to encourage the use of goal structures in settings where an organization currently does not use goal structures, or whether J-BEGM should only be used to express goal structures already in place in organizations. The external validity of this study depends, in part, on what standpoint one takes on these questions.

*Timing of the SUS.* According to guidelines, the SUS should be used after the respondents have had an opportunity to use the evaluated system and before any debriefing or discussion takes place. The latter is to prevent "... the respondents from thinking about items for a long time", and most likely to avoid a facilitator influencing their answers. However, the SUS was, in fact, introduced as the last questionnaire. Because the usability test was divided into two parts, we wanted the participants to spend more time using the system, giving them more experience with the application that could be useful when filling out the SUS. However, as two questionnaires and some discussion took place in between, the



results from the SUS might contain some bias, which poses a threat to external validity, but also to internal validity (above).

### 8.3 Threats to Reliability

Reliability concerns how well a study design produces the same results on separate occasions under the same circumstances [29].

*Question interpretations.* In the first question of the questionnaire, one participant gave a lower score than the other participants because of issues related to the goal tier term and tab. However, all the participants discussed these issues giving the impression of similar perceptions. A reason for unequal scores under seemingly equal perceptions could be misinterpretations of the questions. This might suggest that the tasks and questions were not clearly enough defined, which may introduce variability in answers and poses a threat to the reliability of the evaluation. This could potentially have been avoided if the questions were broken down into smaller sub-questions that were more easily understood. However, the writing of well-written, non-biased questions can be challenging [30].

*Time.* As stated in the external validity section, one participant was tired after a long day of work, which resulted in more time spent to complete the evaluation. This poses a threat to the reliability of the evaluation, as the time given might have influenced the participant's answers. However, as this was the initial evaluation of J-BEGM, getting feedback was prioritized over reliability.

*Number and nature of participants.* Because the number of practitioners of BCDAS is unknown, it was difficult to get in touch with further participants for the evaluation. Because the participants used to evaluate the solution had knowledge of the BCDAS framework, the results for Proposition 1 would seem reliable. The results for Proposition 2, however, may not be reliable, specifically regarding J-BEGM's ability to reflect various goal structures used in development initiatives. As the second part of the usability test was completed only by two participants, the test only shows J-BEGM's ability to reflect two different goal structures compared to the variety of those that might exist in different organizations and development initiatives. Because of this, Proposition 2 needs more investigation.

## 9 Conclusion

Further research and development on the human-computer interface for benefits-management tool support is needed. Nevertheless, our preliminary findings, together with preliminary findings from prototyping related functionality [11] (also based on [10]), indicate a certain enthusiasm for the usefulness of such tool support. Benefits management involves activities that are inherently complex. The present study therefore tests the effect of designing and implementing a human interface to a complex totality, according to design principles that are geared toward helping practitioners to keep the task simple.

Future research and development should include easier navigation in the present tool along with enhanced visualization features and even deeper integration with the standard features of Jira. Further development is underway by several teams, and the hope for the future is that J-BEGM and other extensions that implement other techniques will be complete products that can be distributed on the Atlassian Marketplace to allow organizations and users to install them on their Jira site. This would enable IT professionals to visualize their project's benefits potential and to prioritize with a benefits/cost perspective. Our society depends on the successful management of development projects in terms of goals that reflect the true societal values for its citizens.

## Acknowledgments

The authors are grateful for the efforts and feedback of the practitioners who participated in the usability tests. No sensitive information was gathered, and national regulations concerning consent and anonymity were adhered to.

## References

1. S. S. Tanilkan and J. E. Hannay, "Benefit considerations in project decisions," in *Proc. Int'l Conf. Product-Focused Software Process Improvement (PROFES 2022)*, 2022, pp. 217–234.
2. S. Biffl, A. Aurum, B. Boehm, H. Erdogmus, and P. Grünbacher, Eds., *Value-Based Software Engineering*. Springer, 2006.
3. B. Boehm and L. G. Huang, "Value-based software engineering: A case study," *Computer*, vol. 36, no. 3, pp. 33–41, March 2003.
4. J. Karlsson and K. Ryan, "A cost-value approach for prioritizing requirements," *IEEE Software*, vol. 14, no. 5, pp. 67–74, September/October 1997.
5. C. Larman and B. Vodde, *Practices for Scaling Lean & Agile Development: Large, Multisite, and Offshore Product Development with Large-Scale Scrum*. Addison Wesley, 2010.
6. D. Leffingwell, *Agile Software Requirements: Lean Requirements Practices for Teams, Programs and the Enterprise*. Addison Wesley, 2011.
7. D. Reinertsen, *Principles of Product Development Flow: Second Generation Lean Product Development*. Celeritas Publishing, 2009.
8. J. E. Hannay, H. C. Benestad, and K. Strand, "Benefit points—the best part of the story," *IEEE Software*, vol. 34, no. 3, pp. 73–85, 2017.
9. —, "Earned business value management—see that you deliver value to your customer," *IEEE Software*, vol. 34, no. 4, pp. 58–70, 2017.
10. J. E. Hannay, *Benefit/Cost-Driven Software Development with Benefit Points and Size Points*, ser. Simula Springer Briefs. Springer, 2021.
11. M. Haaber and P. Grøhøj, "Benefit points in scrum: A design science study," Dept. of Computer Science, Aalborg University, Tech. Rep., 2018.
12. J. E. Hannay, H. C. Benestad, and K. Strand, "Agile uncertainty assessment for benefit points and story points," *IEEE Software*, vol. 36, no. 4, pp. 50–62, 2019.
13. R. Breese, S. Jenner, C. E. M. Serra, and J. Thorp, "Benefits management: Lost or found in translation," *Int'l J. Project Management*, vol. 33, no. 7, pp. 1438–1451, 2015.

14. S. Krøglid, “Building and evaluating a web-based tool for software benefits estimation and management,” 2023, Master thesis, University of Oslo, Dept. of Informatics.
15. R. G. Greenwood, “Management by objectives: As developed by Peter Drucker, assisted by Harold Smiddy,” *The Academy of Management Review*, vol. 6, no. 2, 1981.
16. J. Highsmith, L. Luu, and D. Robinson, *EDGE: Value-Driven Digital Transformation*. Addison-Wesley Professional, 2019.
17. M. Denne and J. Cleland-Huang, *Software by Numbers: Low-Risk, High-Return Development*. Prentice Hall, 2003.
18. E. Ries, *The Lean Startup*. Board Book, 2011.
19. H. H. Rydén, S. Hofmann, and G. Verne, “The self-serving citizen as a co-producer in the digital public service delivery,” in *Electronic Government*, I. Lindgren, C. Csáki, E. Kalampokis, M. Janssen, G. Viale Pereira, S. Virkar, E. Tambouris, and A. Zuiderwijk, Eds. Cham: Springer Nature Switzerland, 2023, pp. 48–63.
20. I. Heggertveit, I. Lindgren, C. Østergaard Madsen, and S. Hofmann, “Administrative burden in digital self-service: An empirical study about citizens in need of financial assistance,” in *Electronic Participation, Proc. 14th IFIP WG 8.5 International Conference, ePart, 2022*, pp. 173–187.
21. H. A. Simon, *The Sciences of the Artificial*, 3rd ed. MIT Press, 1996.
22. G. Gigerenzer and D. G. Goldstein, “Reasoning the fast and frugal way: Models of bounded rationality,” *Psychological Review*, vol. 103, no. 4, pp. 650–669, 1996.
23. G. Klein, “Developing expertise in decision making,” *Thinking & Reasoning*, vol. 3, no. 4, pp. 337–352, 1997.
24. M. Denne and J. Cleland-Huang, “The incremental funding method: Data-driven software development,” *IEEE Software*, vol. 21, no. 3, pp. 39–47, May/Jun. 2004.
25. J. Nielsen, *Usability engineering*. Morgan Kaufmann, 1994.
26. J. Brooke, “SUS – a quick and dirty usability scale,” in *Usability Evaluation in Industry*, P. W. Jordan, B. Thomas, B. A. Weerdmeester, and I. L. McClelland, Eds. Taylor & Francis, 1996.
27. M. Jørgensen, “A survey of the characteristics of projects with success in delivering client benefits,” *Information and Software Technology*, vol. 78, pp. 83–94, 2016.
28. E. M. Rogers, *Diffusion of Innovations*, 5th ed. Free Press, 2003.
29. J. Preece, Y. Rogers, and H. Sharp, *Interaction design: beyond human-computer interaction*, fourth edition ed. Wiley.
30. J. Lazar, *Research methods in human computer interaction*, 2nd ed. Elsevier.