Benefit Considerations in Project Decisions

Sinan Sigurd Tanilkan[1][0000−0003−4216−5172] and Jo Erskine Hannay[1][0000−0002−8657−7593]

Simula Metropolitan Center for Digital Engineering, Oslo, Norway
{sinan,johannay}@simula.no

Abstract. Software project success is often characterized in terms of time, cost and scope – despite that delivering benefit is the main purpose of a project. In this paper, we explore 1) to what degree benefit considerations influence major project decisions, 2) to what degree a specific set of benefits management challenge are handled and influence major project decisions and 3) if there is any realization (over time) that benefit considerations should receive greater attention. We investigate influence in projects with four types of problem severity: completed projects with only minor problems, completed projects with major problems, projects that were disrupted but completed, and projects that were terminated before completion. We asked 45 software professionals to what degree time, cost, scope, benefit and benefit/cost, as well as benefits management challenges, influence major project decisions. Our findings indicate that time, cost and scope have a significantly higher degree of influence on project decisions than benefit and benefit/cost. However, practitioners think that benefit and benefit/cost *should* have significantly more influence on decisions than cost. The benefits management challenges are found to have less influence in the more severe projects. We argue that giving benefits considerations a stronger voice in project decisions would be in line with the desire of practitioners and the prime objective of delivering benefit to stakeholders. We conclude that it is important to understand how to handle benefits management challenges at different stages of project life and that handling such challenges should be integrated with other prime drivers of project success.

Keywords: Time · Cost · Scope · Benefit · Benefits management challenges · Software project severity

1 Introduction

A central public sector agency in Norway terminated its information technology modernization program prematurely after about one and a half years' development. The total budget was about EUR 400 million, to be spent over six years. The sunk cost at termination was about EUR 180 million, of which EUR 36 million was spent on functionality that was never to be used [33, 34]. Generally presented by the press as yet another information technology scandal, the termination of the program was applauded in professional circles as a remarkably

insightful decision [46]. When things went bad, program management took the bold decision to stop before further losses, thus countering the *escalation of commitment to a failing course of action* phenomenon [27] and *sunk cost effect* [1] otherwise so proliferant in high-stakes development initiatives. This, and other similar stories, give reasons for optimism; some programs and projects no longer simply spend up their allotted budget no matter what.

The reason, however, for taking action in the above program was, officially, a lack of cost control. Whether they were on track in delivering benefit was not explicitly evident in the decision to stop.

Delivering benefit is the prime reason for software development initiatives, and empirical studies suggest that organizations that engage in *benefits management* [43] perform better in terms of most success criteria [26]. Despite this, there is a tendency to focus on success understood as being on time, being on budget and delivering the specified functionality [18]. In other fields than software engineering, success measured in terms of time, cost and scope, does not correlate with client benefit and satisfaction [37]. This observation has lead researchers to call for further research on the relations between these dimensions for software projects [25].

In light of the above, we want to understand in more detail the extent to which considerations regarding benefit have, or should have, an impact on decisions to continue, disrupt or terminate projects, compared to the traditional control metrics time, cost and scope. To further understand how benefits considerations may play a role, we investigate the extent to which an identified set of benefits management challenges influence these project decisions and project flow. We also explore if there is a growing realization during projects that benefits considerations should have a greater influence. We investigate these topics in four types of projects, according to the severity of problems they encounter.

The next section presents relevant work for our discussion. We present our research questions in Section 3, the research method in Section 4 and the results in Section 5. After that, we discuss and conclude.

2 Background and Previous Work

Benefits Management, defined as "[t]he process of organizing and managing such that potential benefits arising from the use of IT are actually realized" [43], has been suggested to improve organizations' ability to successfully realize benefits of software investments [3, 8, 10, 14, 24, 42, 43], and benefits management practices have been reported to increase benefits realization [13, 23]. Notable characteristics of projects that professionals perceived as "successful" are (a) the application of benefits management practices before and during project execution, (b) the application of core agile practices of frequent delivery to the client and scope flexibility, and (c) that their clients were deeply involved in these practices [25].

The uptake of benefit management practices has been conspicuously slow in light of the existing evidence and general consensus among IT professionals of its relevance. There have been calls for research into what practices in bene-

fits management contribute to success, on how benefits management is actually performed and what challenges practitioners are facing [5, 25]. It seems particularly pertinent to investigate what it is that is hampering benefits management. There have been efforts to understand challenges in benefits management [4, 9, 12, 15, 16, 32] and barriers to benefits management [39]. However, there are few empirical studies on organizations applying benefits management in the context of software development [22], beyond professionals reporting a lack of methodological support for benefits management [25].

A recent in-depth analysis of public-sector projects revealed six sets of conceptual *benefits management challenges* [38]:

A: Identifying and describing the planned benefits of a solution
B: Ensuring that work in the project is aligned with the planned benefits
C: Ensuring the reception and acceptance of the planned benefits
D: Handling organizational issues related to realizing benefits
E: Maintaining an overview of whether the benefits can be realized by other solutions or mechanisms
F: Measuring and evaluating realized benefits

These challenges were uncovered in a *critical case study*: The investigated projects where critical cases [45] in that they had *explicit incentives* to employ benefits management practices. Benefits management challenges uncovered in these projects will arguably be accentuated in projects without such incentives. To increase our understanding of how these challenges influence project decisions, we use them in our further investigations in the next sections.

The so-called *iron triangle* of project management promotes time, cost and scope as control mechanisms to obtain technical quality. The *agile triangle* introduces benefit (extrinsic quality) as a goal together with technical quality; both of which are obtained by controlling, or constraining the bundle of time, cost and scope. However, it has been argued that benefit should be presented as a control mechanisms in its own right; not merely as a fuzzy goal to be obtained by controlling those other things [19, p. 17]. Further, the real control mechanism should be the ratio of benefit/cost, since the point is not to maximize benefit regardless, but to maximize benefit for the cost invested [21].

In the introductory anecdote, the program achieved project learning to the extent that it was possible to make an informed decision based on cost control in the midst of failure. The question arises as to what influence benefit has, or should have, and what influence do benefits management challenges have, both in plain sailing and when the going gets tough.

3 Research Questions

Our first objective is to study the extent to which considerations regarding benefit have, or should have, an impact on project decisions compared to the traditional control metrics time, cost and scope. We compare the standard control

metrics from the "iron triangle" and metrics explicitly involving benefit. The compared control metrics are *time*, *cost*, *scope*, *benefit* and *benefit/cost*.

Our second objective is to understand further how benefits considerations may play a role, and we investigate the extent to which the identified set of benefits management challenges (A–F above) influence project decisions.

A third objective is to see how the influence of both the control metrics and the benefits management challenges might vary according to project problems. For the purpose of this paper, we define four project *severity* types (S1–S4):

(S1) completed projects with only minor problems
(S2) completed projects with major problems
(S3) projects that were disrupted but completed
(S4) projects that were terminated before completion

These severity types are based on the work experience of the authors and three experienced software project professionals.

Based on the above elaborations, we pose the following research questions. Although partly exploratory, we also present expectations with rationales that are not yet founded in theory, but rather, based on anecdotal evidence.

RQa *To what degree do the control metrics time, cost, scope, benefit and benefit/cost influence decisions on termination and disruption in a project?*
Expectation: The measures time, cost and scope are more influential than the measures of benefit and benefit/cost. Moreover, they are more influential than the benefits management challenges, and more so for severe projects. There is a desire that benefit and benefit/cost should be more influential.
Rationale: There is still a focus on the "iron triangle" when controlling projects, and especially when things get difficult, where salvaging cost may be perceived as the better face-saver. There is currently an increased focus and awareness on benefits management that raises awareness that benefit should ideally be the more prominent argument.

RQb *To what degree do the benefits management challenges influence decisions on termination or disruption of a project?*
Expectation: The challenges have less influence on disruption or termination decisions in projects with more severe problems.
Rationale: Benefits management is not used in crises.

RQc *Are there differences in how well benefits management challenges are handled?*
Expectation: At early stages of a project, the challenges are handled less favorably, the more severe the project is.
Rationale: The lack of handling benefits management challenges might have an adverse effect on a project.

RQd *To what degree do practitioners improve their handling of benefits management challenges during projects?*
Expectation: The challenges are handled better at later stages than at early stages, and more so for severe projects.
Rationale: Failure can create an opportunity for learning.

4 Research Method

We conducted a survey with an online questionnaire. A full list of survey questions and responses can be found at: https://tinyurl.com/becipd. Below, we include a subset of the survey questions that are directly relevant to answering the research questions. To sample the participants' personal experience, we prompted them to choose one concrete project, among the four types of project severity, from their experience in software development, and answer the subsequent questionnaire items for that project. Based on the authors' knowledge of the Norwegian IT industry, we assumed that respondents would have fewer terminated or disrupted projects to report on, compared to finished projects. To increase the probability of receiving close to equal amounts of responses in each severity group, the project selection question was phrased to promote selection of disrupted and terminated projects. Respondents were also prompted for their role in the project, as well as for their professional experience in terms of years in software development and the number of projects they had participated in.

4.1 Survey Questions

The survey questions directly relevant to answering the research questions are listed in Table 1 in the order they were posed on the questionnaire. This order was designed for survey comprehension and differs from the (logical) order of the research questions above.

Respondents were prompted for each benefits management challenge A–F (Section 2), indicated by <benefits management challenges> in the questions, and for each control metric (time, cost, scope, benefit and benefit/cost), indicated by <control metrics>. Respondents were given variant phrases indicated by the text in square brackets, according to their choice of project severity type.

The survey was piloted prior to data collection on five respondents (on two research colleagues and three experienced managers from the IT industry). The pilot resulted in changes to the wording of questions in the survey for better comprehension and alignment with current terminology in the field. This applied in particular to the project selection question. Minor adjustments were also done to SQ1–SQ5. The project severity groups (S1–S4) were also finalized and validated for meaningfulness and relevance during the pilot.

Data was collected during a webinar titled "Failed digitalization projects: A source of learning and improvement?" in October 2021. In the webinar a selection of IT professionals presented experiences from failed projects, including lessons learned from these projects. A total of 71 professionals were present at the webinar when the survey was conducted. A link to the questionnaire was given as part of the opening remarks to the webinar, and participants were given ten minutes to complete the questionnaire. Fifty-seven persons participated in the survey, but twelve did not complete the survey, leaving 45 complete responses.

The number of software development projects in which the respondents had participated ranged from two to 100 (median: 15, mean: 20.31). The number of years of experience within development of digital solutions ranged from under

	In your opinion,	Answer options
SQ1	*how were the following* <benefits management challenges> *handled in the early phases of the project?*	seven-point ordinal (poorly 1–7 well)
SQ2	*to what extent did the following* <control metrics> *of the project influence* [*decisions along the way*], [*the decision to change course*], [*the decision to stop*]?	seven-point ordinal (minor 1–7 major)
SQ3	*to what extent should the following* <control metrics> *of the project have influenced* [*decisions along the way*], [*the decision to change course*], [*the decision to stop*]?	seven-point ordinal (minor 1–7 major)
SQ4	*to what extent did problems in the following* <benefits management challenges> *influence* [*decisions along the way*], [*the decision to change course*], [*the decision to stop*]?	seven-point ordinal (minor 1–7 major)
SQ5	*compared to the early stages of the project, how* [*were*], [*would*] *the following* <benefits management challenges> [*handled at later stages of the project*], [*handled after changing course*], [*have been handled if the project had continued*]?	seven-point ordinal (worse -3– +3 better)

Table 1: Survey questions

a year to 40 (median: 20, mean: 18.04). The number of years of experience as a manager within this field ranged from zero to 30 (median: 5, mean: 7.78). The project that each participant chose as a reference for the subsequent questions was owned by a public sector organization in (68.9%) of the cases and the private sector in (31.1%) of the cases. The distribution per project severity type was as follows: completed projects with only minor problems (24.5%), completed projects with major problems (22.2%), projects that were disrupted but completed (28.9%), projects that were terminated before completion (24.4%).

4.2 Analysis

Ordinal data from the questionnaire was analyzed using percentile box-plots for descriptive statistics.[1] We used related-samples Friedman's two-way analysis of variance by ranks for comparison with and across benefits management challenges and control metrics. We used independent samples Jonckheere-Terpstra non-parametric tests for comparisons across the four types of project severity. Both tests are specifically for ordinal data. The Jonckheere-Terpstra test assumes directional comparisons and is one-tailed: We are expecting responses to

[1] Data analyses were conducted using IBM SPSS Statistics version 27 using test-wise deletion of missing data.

be higher for one severity level than another; for example we expect handling to deteriorate from level S1 to level S4 (as described for RQc in Section 3). We accept statistical significance at $p \leq \alpha = 0.05$. That is, we accept a 5% chance of rejecting the null hypothesis when it is, in fact, true.

Traditionally, one performs an omnibus test, with ensuing pairwise comparisons if the omnibus test is significant. Our primary interest lies with the pairwise comparisons, and we perform the pairwise comparisons even when the omnibus test is not significant. There can be significant pairwise differences, even when the omnibus test is not significant [40]. We are interested in single comparisons; for example if a challenge is handled worse between two levels of severity. We are also interested in composed comparisons; for example if a challenge intensifies across a chain of severities. When composing multiple comparisons, the probability of rejecting the null hypothesis for any one in the group of comparisons increases. If one intends to draw conclusions on a composed comparison on the basis of any one constituent comparison, one should therefore use a stricter α_{adj} using, e.g., the Bonferroni adjustment. In the composed comparisons we are interested in, the null hypotheses for all tests in the composition have to be rejected, and using a stricter α_{adj} is not relevant [2, 17]. We do, however, also report the Bonferroni-adjusted probability (p_{adj}) to cater for other kinds of composed comparisons. For space reasons, we only display the significant results.

We wish to report effect sizes for the pairwise comparisons. The pairwise comparisons for the Jonckheere-Terpstra tests are based on the Mann-Whitney U statistic, and it is possible to report effect sizes estimates in terms of Cohen's d [30], where the following rules of thumb apply: <0.1 (very small), 0.1 – <0.3 (small), 0.3 – <0.5 (medium), 0.5 – <1.2 (large), 1.2 – <2.0 (very large) and >=2.0 (huge) [36]. Pairwise comparisons for the Friedman test are in terms of the Dunn-Bonferroni statistic with no straightforward effect size estimate, so for the Friedman tests, we report effect sizes in terms of Kendal's W for the omnibus test [41] in lack of anything better. Kendal's W ranges from 0 to 1, with the following rules of thumb for evaluating effect sizes: 0.1 – <0.3 (small), 0.3 – <0.5 (medium) and >=0.5 (large) [11].

With our small sample size, statistical power is expectedly low. That is, there is low probability of the data revealing (significant) effects, when, in fact, there are effects in the intended population, and the probability of revealing small effects is lower than that of revealing large effects. On the other hand, it is all the more promising for further studies if our data does reveal effects under low power. Given a sample size, one might calculate power for various effect sizes (small, medium, large) and see if the commonly acceptable level of $\beta = 0.8$ is achieved, but power calculations for non-parametric tests are not straightforward [35], and we omit them for this initial study.

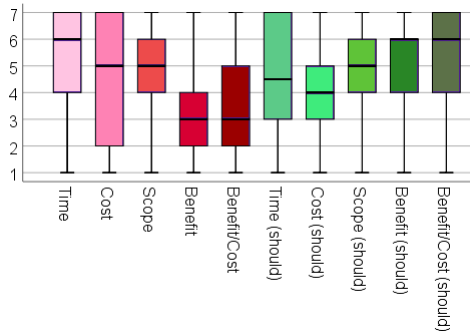
5 Results

The Friedman omnibus tests are all significant with effect sizes ranging from very small to small. None of the Jonckheere-Terpstra omnibus tests are significant,

while several of the pairwise comparisons are; and except for two of the results (on RQc), all significant results have large, very large or huge effect sizes. Larger effect sizes are generally more useful for practitioners [28]. Nevertheless, a study with higher statistical power would have a higher probability of finding significant results with also smaller effects sizes. This would be particularly interesting for establishing the expected linked relationships across all four severity types, which are only partially seen in our data. In the following, we report these, and other significant findings.

RQa: To what degree do the control metrics time, cost, scope, benefit and benefit/cost influence decisions on termination and disruption in a project? This research question is answered using the responses from survey questions SQ2 and SQ3. Figure 1a shows descriptive statistics, regardless

(a) Influence of control metrics on project decisions – actual (red), should (green):



(b) Sign. diff. in influence – actual: Omnibus test n: 41, p: .000, W: .19

Pair-wise two-sided tests			p	p_{adj}
Time	>	Benefit	.000	.002
	>	Benefit/Cost	.000	.003
Cost	>	Benefit	.025	.254
	>	Benefit/Cost	.028	.278
Scope	>	Benefit	.000	.005
	>	Benefit/Cost	.001	.005

(c) Sign. diff. in influence – should: Omnibus test n: 41, p: .000, W: .14

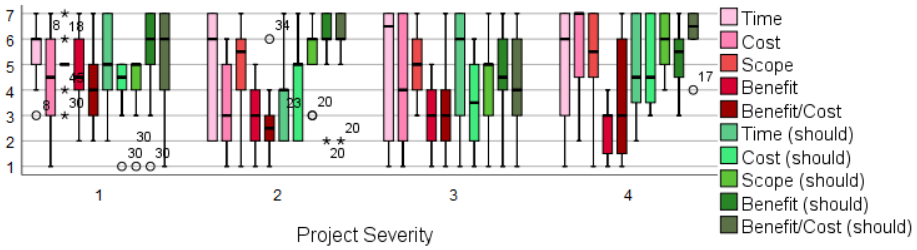
Pair-wise two-sided tests			p	p_{adj}
Benefit	>	Cost	.006	.058
Benefit/cost	>	Time	.010	.098
	>	Cost	.000	.002
Scope	>	Cost	.043	.428

(d) Sign. diff. in influence – actual versus should: Omnibus test n: 38, p: .000, W: .17

Pairwise two-sided tests			p	p_{adj}
Benefit actual	<	Benefit should	.000	.000
Benefit/Cost actual	<	Benefit/Cost should	.000	.000

Fig. 1: Analysis for RQa of SQ2 and SQ3 – the influence of control metrics on project decisions – actual and should. Friedman tests.

(a) Influence of control metrics on project decisions – by project severity:



(b) Sign. diff. in influence of control metrics across severity (one-sided):

Actual	Severity	p	p_{adj}	d	Should	Severity	p	p_{adj}	d
Cost	S2<S4	.006	.039	1.250	Scope	S1<S2	.016	.096	1.017
Benefit	S1>S2	.011	.067	1.129		S1<S4	.035	.210	0.813
	S1>S3	.023	.136	1.297		S2>S3	.035	.208	0.821
	S1>S4	.017	.105	0.558		S3<S4	.041	.245	0.768
Benefit/Cost	S1>S2	.032	.194	0.888	Benefit	S2>S3	.045	.270	0.756
					Benefit/Cost	S2>S3	.019	.114	0.949
						S3<S4	.005	.028	1.264

Fig. 2: Analysis for RQa of SQ2 and SQ3 – the influence of control metrics on project decisions – by project severity. Jonckheere-Terpstra tests.

of project severity, for the influence the control metrics time, cost, scope, benefit and benefit/cost (red shades) are reported to have on project decisions, and the influence practitioners report that the control metrics "should have" (green shades). It is immediately apparent that benefit and benefit/cost were perceived as less influential than the iron-triangle metrics (significant for time, cost and scope – Fig. 1b), but that the respondents thought that benefit and benefit/cost should have had more influence (significant for cost and time – Fig. 1c). Figure 1d shows significance for comparisons between the influence that a metric has, versus should have. The desire that benefit and benefit/cost should have more influence shows up highly significantly in the data.

Figure 2a shows descriptive statistics again, but now per project severity (S1–S4). One can see how respondents perceive that time, cost and scope were highly influential in terminating projects, and that benefit and benefit/cost considerations were not very influential (red boxplots for severity S4) in terminating projects. However, respondents think benefit and the benefit/cost ratio *should* have been highly influential when considering terminating the project (green boxplots for severity S4). Similar remarks hold for severities S3 and S2. For projects with minor problems, benefit and benefit/cost are perceived to have had more influence, with a desire to increase that influence further.

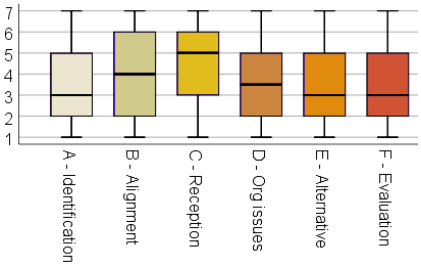
The expectation that benefit is less influential the more severe the project, is supported for project severities S1>S2, S1>S3 and S1>S4 (Fig. 2b). The notion

of benefit/cost ratio looses influence for S1>S2. In contrast, cost is perceived to have greater influence the more severe the project (S2<S4). Thus, the difference in influence between cost and benefit clearly increases to the disadvantage of benefit, over project severities.

RQb: To what degree do the benefits management challenges influence decisions on termination or disruption of a project? For this, we analyzed responses for SQ4. From Fig. 3a and b, we see that challenge C has significantly more influence on decisions, regardless of severity, than challenges A, D, and F. Also, challenge B is significantly more influential than challenge A.

Comparing across project severity (Fig. 3c), we see that influence is uniformly higher for severity S1 (minor issues). Pairwise comparisons (Fig. 3d) support the expectation of less influence the more severe the project, with significant

(a) Influence of benefits management challenges on project decisions:

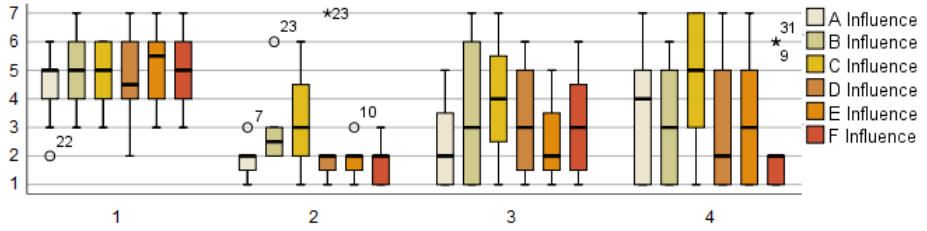


(b) Sign. diff. in influence of benefits management challenges:

Omnibus test n: 38, p: .011, W: .078

Pairwise two-sided tests			p	p_{adj}
B (Alignment)	>	A (Ident.)	.034	.516
C (Reception)	>	A (Ident.)	.008	.115
	>	D (Org issues)	.050	.746
	>	F (Evaluation)	.027	.409

(c) Influence of challenges on project decisions – by project severity:



(d) Sign. diff. in influence of challenges across severities (one-sided):

Chal.	Severity	p	p_{adj}	d	Chal.	Severity	p	p_{adj}	d	Chal.	Severity	p	p_{adj}	d
A	S1>S2	.000	.002	2.185	C	S1>S2	.030	.183	0.872	F	S1>S2	.000	.001	2.911
	S1>S3	.002	.015	1.415	D	S1>S2	.012	.074	1.132		S1>S3	.010	.057	1.130
B	S1>S2	.017	.104	1.017	E	S1>S2	.001	.005	1.896		S1>S4	.030	.183	0.910
	S1>S4	.021	.123	0.974		S1>S3	.002	.011	1.564		S2<S3	.045	.272	0.764
						S1>S4	.039	.231	0.849					
						S2<S3	.045	.272	0.764					

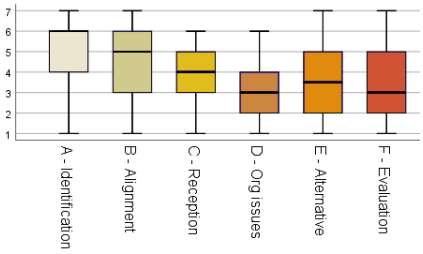
Fig. 3: Analysis for RQb of SQ4– influence of benefits management challenges on project decisions. Friedman tests (a, b), Jonckheere-Terpstra tests (c, d).

differences for all challenges in the expected direction. (The data exhibits the opposite direction, across severities S2 and S3, for challenge F.)

RQc: Are there differences in how well benefits management challenges are handled? To answer this research question, we used responses to SQ1. Figure 4a indicates that challenges C, D, E and F are handled poorer (at early stages) than A and B. Looking at Fig. 4b, we see that challenge A is handled significantly better than all the other challenges. Also, challenge B is handled significantly better than challenges D, E and F.

Our expectation that challenges are handled less favourably the more severe the project, is supported to some extent: Visual inspection of the boxplots (Fig. 4c) suggests a general tendency of decreasing early handling of challenges

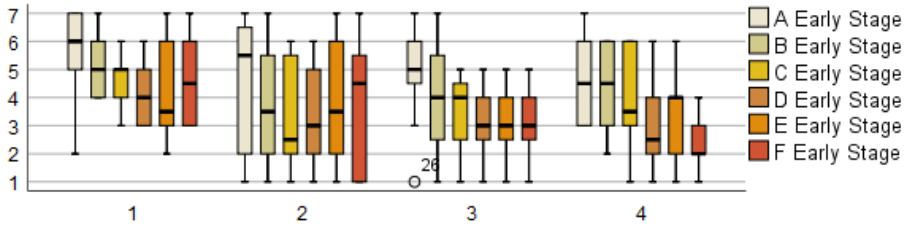
(a) How well benefits management challenges were handled early:



(b) Sign. diff. in how well the benefits management challenges were handled early:
Omnibus test n: 39, p: .000, W: .27

Pairwise two-sided test		p	p_{adj}
A (Ident.)	> B (Alignment)	.032	.475
	> C (Reception)	.000	.003
	> D (Org issues)	.000	.000
	> E (Alternative)	.000	.000
	> F (Evaluation)	.000	.000
B (Alignment)	> D (Org issues)	.001	.018
	> E (Alternative)	.018	.274
	> F (Evaluation)	.009	.139

(c) How well benefits management challenges were handled early – by project severity:



(d) Sign. diff. in handling of each benefits management challenge across project severities (one-sided):

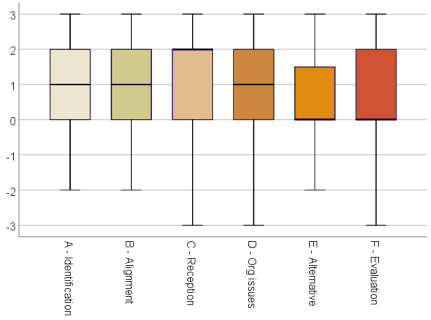
Chal.	Severity	p	p_{adj}	d	Chal.	Severity	p	p_{adj}	d	Chal.	Severity	p	p_{adj}	d
A	S1>S4	.046	.277	0.747	C	S1>S4	.038	.226	0.799	D	S1>S4	.015	.089	1.002
B	S3<S4	.034	.202	0.214		S2<S4	.042	.251	1.956	F	S1>S4	.002	.0125	1.558
						S3>S4	.022	.131	0.275		S2>S4	.009	.052	2.437

Fig. 4: Analysis for RQc of SQ1 – handling of benefits management challenges in early stages of projects. Friedman tests (a, b), Jonckheere-Terpstra tests (c, d).

as project severity increases, which is supportive of our expectation. Pairwise comparisons (Fig. 4d) give significant differences for challenges A, B (small effect size), C, D and F, where early handling is better for severity S1 than for a variety of higher severities. Still, the data does not give evidence of a steadily decreasing trend through severities.

RQd: To what degree do practitioners improve their handling of benefits management challenges during projects? We analyzed responses on SQ5. The boxplots in Fig. 5a indicate that respondents perceive a weak improvement of the handling of the benefits management challenges at later stages, thus

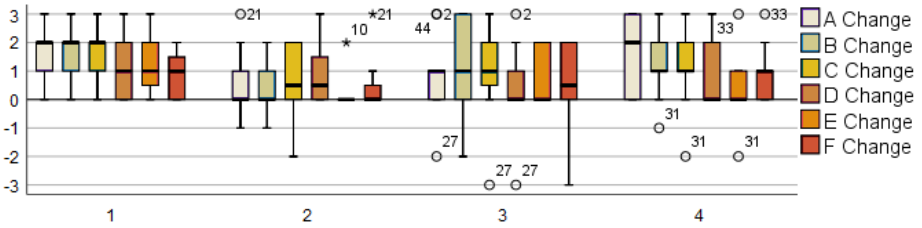
(a) Improvement in handling of benefits management challenges:



(b) Sign. diff. in improvement of handling benefits management challenges:
Omnibus test n: 36, p: .018, W: .076

Pairwise two-sided test			p	p_{adj}
A (Ident.)	>	E (Altern.	.044	.657
B (Alignment)	>	E (Altern.)	.038	.565
C (Reception)	>	E (Altern.)	.044	.657

(c) Improvement in handling of benefits management challenges – by project severity:



(d) Sign. diff. in improvement of handling benefits management challenges across project severities (one-sided):

Challenge	Severity	p	p_{adj}	d
A	S1>S2	.020	.121	1.090
	S1>S3	.048	.290	0.727
B	S1>S2	.015	.090	1.160
C	S1>S2	.027	.164	0.923
E	S1>S2	.035	.209	0.867

Fig. 5: Analysis for RQd of SQ5 – improvement in handling of benefits management challenges. Friedman tests (a, b), Jonckheere-Terpstra tests (c, d).

supporting the expectation of project learning on these challenges. Improvement is similar across challenge types, except for challenge E, where practitioners report significantly less improvement compared to challenges A, B and C (Fig 5b).

When comparing improvement across severity types (Fig. 5c), pairwise comparisons (Fig. 5d) reveal that for some of the challenges, improvement in handling challenges was greater for severity S1 than for severity S2 and severity S3. There are also some indications in the data that improvement was lowest for severity S2 and that improvement increases from S2 to S3 and S4 (challenges A, B, E), in line with our expectations of increased improvement for severe projects, but these latter observations fail to be significant at our chosen level.

6 Discussion

The results above suggest clearly that practitioners think that more emphasis should be placed on benefit and benefit/cost when making project decisions, compared to the iron-triangle metrics of time, cost and scope. But given that practitioners seem to be aware of the importance of increasing the influence of benefit and benefit/cost considerations, the question arises as to how to make this happen. To provide actionable guidance to practitioners, we must understand what is keeping practitioners from prioritizing the right factors when making decisions. We observed that benefit considerations have less influence in more severe projects, and that challenges A, C and F are handled less favorably in more severe projects. These challenges may therefore be a good starting point to understanding why benefit does not get the attention it should in project decisions. In particular, it seems important to understand better the characteristics of challenges A, C and F that can affect practitioners ability to manage benefit, and therefore to employ considerations of benefit in project decisions.

The deterioration of the early handling of challenge A (Identifying and describing the planned benefits of a solution) as project severity increases, could be due to difference in predictability of benefits identification. Several papers have reported that practitioners find it challenging to identify all benefits before project initiation [3, 14, 25, 38], but it is reasonable to assume that the identification of benefits of some projects are more predictable than others. As such, less favorable handling of challenge A in more severe projects, might be due to greater challenges in identifying benefits, rather than poor handling of the challenge. Also, it is reasonable to expect that some projects are more aware of the need to update planned benefits during the project [38]. If unpredictable changes to benefits is the underlying problem, then measures to handle such unpredictability are called for. The incremental and agile approach of not overplanning early then applies also to benefit, and it becomes correspondingly more important to open up the project to changes to planned benefits in addition to time, cost and scope. Techniques to declare and update benefits and monitor the progress in developing beneficial code may be useful [20, 21, 29]. Keeping track of realized benefit in terms of beneficial code, as one keeps track of the cost of code, can then aid in early and sound decisions on continuation, disruption or

termination of projects. However, even with such techniques, organizational issues may add to the challenge, because updating the business case of the project often requires effort from people outside the project organization [39]. An agile approach to business cases is called for.

Challenge C (Ensuring the reception and acceptance of the planned benefits) is likely to be affected both by a difference in difficulty of the challenge and a difference in practitioners handling, which can help explain the deterioration in handling as severity increases. Difference in difficulty is likely to occur as a result of varying resistance to new solutions [8] and a varying interest in benefits themselves [9]. It is unclear if it is possible to predict the degree of difficulty that will be encounter in the reception and acceptance of benefits in different projects. There seems to be differences between benefits that are internal and external to the organization, but further research is needed to understand the characteristics of benefits and stakeholders, that affect the challenges of reception and acceptance of benefits [38]. Previous empirical research has documented a difference in effort put into handling of challenge C. Even when practitioners are aware that more work is needed in order to realize benefits, the extra needed effort is sometimes not spent [38]. Influencing stakeholders to receive and accept new solutions and benefits is a topic where we have not found any research in the field of benefits management, and it is likely that practitioners do not have much empirical or actionable guidance available, other than normative guidance to the effect that the problem is important to keep in mind [42].

Challenge F (Measuring and evaluating realized benefits) has gained much attention in research on benefits management [6, 31], but the challenge still remains very much alive for practitioners [38]. Here we discuss three issues of measuring and evaluating realized benefits that are relevant to project decisions and the handling of challenge F.

First, while measures of time, cost and scope are fairly standardized, and fits nicely into business decisions, measures of benefits varies largely. The characteristics used to describe benefits, such as qualitative/quantitative [23] and financial/non-financial [44] is one example. It is reasonable that financial measures are easier to include in business decisions than qualitative evaluations.

Secondly, measures must be taken after benefits realization has started. This is usually in late phases of a project or after a project has ended. Hence, the necessary data is often not available to be used for project decisions. One mitigation is the use of leading measures [24] which are measurement of indicators that are available early. The problem with indicators, is that they are not measures of the actual benefit, and might not give a true representation of the benefits.

Third, it seems practitioners do not prioritize evaluation. Organization may put little emphasis on evaluation because their limited IT resources would be better spent on new projects, rather than on evaluating old projects [7].

7 Limitations

Statistical Conclusion Validity. The low statistical power decreases the probability of the data exhibiting effects where there might, in fact, be effects in the population. The opposite threat of the data exhibiting effects, when there might, in fact, be none, is handled by the significance tests. Even with this small a sample, the data exhibits significant effects. Replicating studies that use larger samples may find additional effects.

External Validity. Based on the characteristics of the sample itself one can generalize the results to similar groups (populations) of interest. This can be problematic for our sample for the following reasons:

1. The sampling strategy was designed to increase the number of disrupted and terminated projects (compared to the population of IT-projects).
2. We have limited information on the characteristics of the projects in the study. Some characteristics, such as the technology applied or the type of solution created, are less likely to influence the studied topics. Other characteristics, such as project size and organization, are likely to have an influence, but were not collected, due to time limitations duration the webinar.
3. It is possible that practitioners attending a seminar in Norway on failed projects have different experiences than other practitioners.

Experiences from this, and other similar webinars and seminars, suggests that participants represent a varied selection of IT professionals that together have a broad experience in many types of software project. We therefore hold that our results are genealizable to the situation of termination and disruption decisions, as long as one takes the above threats into consideration.

Construct Validity. Although response rates and pilots of the survey suggest that participants were able to relate to the given challenges, we do not know to what degree the challenges occurred or how difficult they were to handle. This is likely to have caused different perceptions of the challenges as concepts and of the concepts of "good" and "poor" handling. In retrospect, asking for success in handling, could have mitigated part of this problem. There are similar issues with the influence that challenges have on project decisions (e.g., a challenge that did not occur, is likely to be reported to have low influence on project decisions). Not asking about the occurrences of challenges was a conscious choice when designing the study, because we did not want to confuse the respondents with too many similar questions. However, this issues should be dealt with in later studies. We hold that our findings are relevant as a basis for further study and as initial advice to practitioners in software projects.

8 Conclusion and Further Research

We conclude that benefit and benefit/cost considerations should have more influence in project decisions than they currently have. This would help align project

decisions with the primary objectives of projects – to deliver benefits. However, the characteristics of benefits together with benefits management challenges, seems to make considerations of benefit more difficult than time, cost and scope. As a result, we propose three topics for further improvement and research:

- Guidance to help practitioners handle changes to understanding of benefits
- Explore how practitioners can influence others to ensure benefits realization
- Improved guidance to practitioners on benefits evaluation

Further research in these topics is needed in order to understand the difficulties of benefit considerations and how practitioners can gain the information they need to make timely project decisions, influenced by benefit considerations.

Acknowledgments The authors are grateful to experienced IT professionals Kjetil Strand, Hans Christian Benestad and Bjørn Olav Salvesen for feedback on the questionnaire. The authors are further grateful to the survey respondents.

References

1. Arkes, H.R., Ayton, P.: The Sunk Cost and Concorde Effects: Are humans less rational than lower animals? Psychological Bulletin **25**(5), 591–600 (1999)
2. Armstrong, R.: When to use the bonferroni correction. Ophthalmic and Physiological Optics **34** (04 2014)
3. Ashurst, C., Doherty, N.F.: Towards the formulation of a 'best practice'framework for benefits realisation in it projects. Electronic Journal of Information Systems Evaluation **6**(2), 1–10 (2003)
4. Askedal, K., Flak, L.S., Aanestad, M.: Five challenges for benefits management in complex digitalisation efforts–and a research agenda to address current shortcomings. Electronic Journal of e-Government **17**(2), 64–78 (2019)
5. Aubry, M., Boukri, S.E., Sergi, V.: Opening the black box of benefits management in the context of projects. Project Management Journal **52**(5), 434–452 (2021)
6. Ballantine, J.A., Galliers, R.D., Stray, S.J.: Information systems/technology evaluation practices: evidence from uk organizations. Journal of Information Technology **11**(2), 129–141 (1996)
7. Berghout, E., Nijland, M., Powell, P.: Management of lifecycle costs and benefits: Lessons from information systems practice. Computers in Industry **62**(7), 755–764 (2011)
8. Bradley, G.: Benefit Realisation Management: A practical guide to achieving benefits through change. Routledge (2016)
9. Breese, R.: Benefits realisation management: Panacea or false dawn? Int'l J. Project Management **30**(3), 341–351 (2012)
10. Breese, R., Jenner, S., Serra, C.E.M., Thorp, J.: Benefits management: Lost or found in translation. International Journal of Project Management **33**(7), 1438–1451 (2015)
11. Cohen, J.: Statistical power analysis for the behavioral sciences. Routledge, second edn. (1988)
12. Coombs, C.R.: When planned IS/IT project benefits are not realized: a study of inhibitors and facilitators to benefits realization. International Journal of Project Management **33**(2), 363–379 (2015)

13. Doherty, N.F., Ashurst, C., Peppard, J.: Factors affecting the successful realisation of benefits from systems development projects: Findings from three case studies. Journal of Information Technology **27**(1), 1–16 (2012)
14. Farbey, B., Land, F., Targett, D.: The moving staircase – problems of appraisal and evaluation in a turbulent environment. Information Technology & People **12**(3), 238–252 (1999)
15. Fernandes, G., O'Sullivan, D.: Benefits management in university-industry collaboration programs. International Journal of Project Management **39**(1), 71–84 (2021)
16. Flak, L.S., Eikebrokk, T.R., Dertz, W.: An exploratory approach for benefits management in e-government: Insights from 48 norwegian government funded projects. In: Proc. 41st Annual Hawaii International Conference on System Sciences (HICSS) (2008), article no. 210
17. Gigerenzer, G.: Mindless statistics. The Journal of Socio-Economics **33**, 587–606 (2004)
18. Gingnell, L., Franke, U., Lagerström, R., Ericsson, E., Lilliesköld, J.: Quantifying success factors for it projects-an expert-based bayesian model. Information Systems Management **31**(1), 21–36 (2014)
19. Hannay, J.E.: Benefit/Cost-Driven Software Development with Benefit Points and Size Points. Simula Springer Briefs, Springer (2021)
20. Hannay, J.E., Benestad, H.C., Strand, K.: Benefit points—the best part of the story. IEEE Software **34**(3), 73–85 (2017)
21. Hannay, J.E., Benestad, H.C., Strand, K.: Earned business value management—see that you deliver value to your customer. IEEE Software **34**(4), 58–70 (2017)
22. Hesselmann, F., Mohan, K.: Where are we headed with benefits management research? Current shortcomings and avenues for future research. In: Proc. 22nd European Conference on Information Systems (ECIS) (2014)
23. Holgeid, K.K., Jørgensen, M.: Benefits management and agile practices in software projects: how perceived benefits are impacted. IEEE 22nd Conference on Business Informatics (CBI) **2** (2020)
24. Jenner, S.: Managing Benefits: Optimizing the Return from Investments. The Stationery Office, APMG-International (2014)
25. Jørgensen, M.: A survey of the characteristics of projects with success in delivering client benefits. Information and Software Technology **78**, 83–94 (2016)
26. Jørgensen, M., Mohagheghi, P., Grimstad, S.: Direct and indirect connections between type of contract and software project outcome. International J. Project Management **35**(8), 1573–1586 (2017)
27. Keil, M., Mann, J., Rai, A.: Why software projects escalate: An empirical analysis and test of four theoretical models. MIS Quarterly **24**(4), 631–664 (2000)
28. Kirk, R.E.: Practical significance: A concept whose time has come. Educational and psychological measurement **56**(5), 746–759 (1996)
29. Leffingwell, D.: Agile Software Requirements: Lean Requirements Practices for Teams, Programs and the Enterprise. Addison Wesley (2011)
30. Lenhard, W., Lenhard, A.: Computation of effect sizes (2016), retrieved from: https://www.psychometrica.de/effect_size.html. Psychometrica. DOI: 10.13140/RG.2.2.17823.92329
31. Lin, C., Pervan, G.: Is/it investment evaluation and benefits realisation issues in a government organisation. In: ACIS 2001 Proceedings. 49 (2001)
32. Lin, C., Pervan, G.: The practice of IS/IT benefits management in large Australian organizations. Information & Management **41**(1), 13–24 (2003)

33. Lystad, J.: Det er ingen skam å snu – erfaringer fra Mattilsynet og NAV. Presentation given at Conf. of the Agency for Public Management and eGovernment (Difi), Dec. 6 (2017)
34. Olaussen, S., Tendal, Ø., Johansen, S., Sem, V., Bråthen, S., Bremnes, H., Grubbmo, E., Ræder, A.D.: KSP-rapport nr. 1 for Modernisering av IKT i NAV – Rapport til Finansdepartementet og Arbeids- og sosialdepartementet, Versjon: 1.0 (2015)
35. Rabbee, N., Coull, B.A., Mehta, C.: Power and sample size for ordered categorical data. StatisticalMethods in Medical Research **12**, 73–84 (2003)
36. Sawilowsky, S.S.: New effect size rules of thumb. Journal of Modern Applied Statistical Methods **8**(2), 596–599 (2009)
37. Shenhar, A.J., Dvir, D., Levy, O., Maltz, A.C.: Project success: A multidimensional strategic concept. Long range planning **34**(6), 699–725 (2001)
38. Tanilkan, S.S., Hannay, J.E.: Perceived challenges in benefits management—a study of public sector information systems engineering projects. In: 2022 IEEE 24th Conference on Business Informatics (CBI) (06 2022)
39. Terlizzi, M.A., Albertin, A.L., de Oliveira Cesar de Moraes, H.R.: IT benefits management in financial institutions: Practices and barriers. Int'l J. Project Management **35**(5), 763–782 (2017)
40. Tian, C., Manfei, X., Justin, T., Hongyue, W., Xiaohui, N.: Relationship between omnibus and post-hoc tests: An investigation of performance of the F test in ANOVA. Shanghai Arch Psychiatry **30**(1), 60–64 (2018)
41. Tomczak, M., Tomczak, E.: The need to report effect size estimates revisited. an overview of some recommended measures of effect size. Trends in Sport Sciences **1**(21), 19–25 (2014)
42. Ward, J., Daniel, E.: Benefits Management: How to increase the business value of your IT projects. Wiley, second edn. (2012)
43. Ward, J., Taylor, P., Bond, P.: Evaluation and realisation of IS/IT benefits: an empirical study of current practice. European Journal of Information Systems **4**, 214–225 (1996)
44. Williams, T., Vo, H., Bourne, M., Bourne, P., Cooke-Davies, T., Kirkham, R., Masterton, G., Quattrone, P., Valette, J.: A cross-national comparison of public project benefits management practices – the effectiveness of benefits management frameworks in application. Production Planning & Control **31** (2020)
45. Yin, R.K.: Case Study Research: Design and Methods, Applied Social Research Methods Series, vol. 5. Sage Publications, third edn. (2003)
46. Zachariassen, E.: Nav stanser IT-prosjekt til 3,3 milliarder – Moderniseringsprogrammet var feil metode. Nav får skryt fra statlig ekspert, article published Oct. 25, 2013