Suggestions for Improving Measurement Plans: First results from a BMP application

Reiner Dumke, Alain Abran, Luigi Buglione

Abstract

"Tracking and control" activities in software projects are most often based, in industry, on just two dimensions of analysis: time and cost. Most often, these activities exclude other dimensions (such as quality, risks, impact on society, the stakeholders' viewpoint in a broader sense) taken into account in Performance Management models such as EFQM or the Malcolm Baldrige model. How can these multiple concurrent control mechanisms across several dimensions of analysis be balanced? Balancing Multiple Perspectives (BMP) is a procedure designed to help project managers choose a set of project indicators from several concurrent viewpoints.

This paper also presents the initial results from a BMP application, using a list of 14 candidate measures, with the objectives of representing the "as is" situation and determining what the "to be" situation will be, including cost figures to be possibly considered in future project budgets. Based on the results presented, which have been gathered both from an industrial and an academic sample, make it possible to look at several potential viewpoints and provide suggestions for improving measurement plans.

1. Introduction

There has been considerable interest in recent years in how projects should be tracked and controlled, including a greater demand for Project Management Professional (PMP) certification. There has also been greater attention paid to the ways in which a project could be more profitable and less defect-prone, but often not as much to the project budget allocated for tracking and control (T&C) activities. What is the project budget percentage dedicated to those activities? Is it 1% or 10%, for example, and how much does it cost to track and control a software project?

A few well documented studies were carried out in the '90s on this issue. Capers Jones reported the costs of measurement in projects (one of the few studies proposing actual figures) to be between approximately 3% and 6% for internal project measurement and between approximately 2.5% and 4.5% for an external one [6].

Tom Demarco, whose well-known motto about measurement has for years been, "*You can't control what you can't measure,*" said in 1995 that "*metrics cost a ton of money. It costs a lot to collect them badly and a lot more to collect them well […] At its best […] metrics can inform and guide developers and help organizations to improve. At its worst, it can do actual harm. And there is an entire range between the two extremes*" [2].

Two out of ten problems leading to failure in the implementation of software measurement programs are reported by Howard Rubin to be: the intensive use of a single measure or, conversely, the use of too many [9]. What, then, is the issue surrounding measurement costs? Is it to reduce or cancel a portion of a measurement program in order to meet budgeted targets from an economic/financial viewpoint, or – more appropriately – to balance how the T&C process budget should be spent across several dimensions of analysis (for instance: quality, risk, ethics, user satisfaction, and so on)?

Management tools such as the Balanced Scorecard (BSC) are based on multiple concurrent perspectives. In this paper, a procedure called Balancing Multiple Perspectives (BMP) is proposed to tackle this measurement issue, and could be used as a tool to reinforce the choice of measures and indicators to support the design of *strategic maps*.

The paper is organized as follows: Section 2 presents the BMP, its objectives and the related procedure. Section 3 discusses the initial results and the assumptions under which a test was conducted, while in section 4 the results are analyzed and discussed. Finally, section 5 reports our conclusions and some suggestions for future work.

2. BMP: Balancing Multiple Perspectives

The average percentage of a project budget dedicated to the T&C process is generally underestimated. The ultimate corporate objective is (obviously) profitability – as also stated in the BSC approach. When fiscal quarterly results are strained, the counteraction is to reduce costs on projects and in cost-based activities, including what pertains to the "control" (and therefore measurement) sphere.

2.1. Objectives

A key concept in the BMP approach [1] is that increasing performance does not need to be limited to *reducing cost,* but can also be achieved by *optimizing through balancing* the actual forces and energies at play within a project. While time and cost are the main analysis perspectives of interest to managers, other concurrent perspectives could be profitably be taken into account as well. Of course, adding to current controls would increase the budget percentage allocated to T&C activities. Therefore, while maintaining both constraints (broadening the perspectives of analysis under the same project budget percentage for T&C activities), an interesting solution would be to balance the number of measurement controls across more than two perspectives.

A basic mechanism behind BMP is to make more explicit trade-offs across several dimensions of analysis. For instance, if the priority is to pay more attention to time-to-market aspects, than quality could suffer (in terms of product defect rate). Similarly, if the priority is to produce defect-free software products, a more adequate testing phase might be required, by increasing project costs while reducing the prospective project mark-up on the one hand, and, on the other, reducing the potential rework following the release through a lower defect rate, and so on. Fig.1 presents a possible template for a generic four-dimension case.

2.2. The procedure

The BMP procedure for controlling multiple concurrent dimensions consists of four steps (which could be performed jointly by a project manager and his quality assurance assistant):

(1) Determine the dimensions of interest in the project: at least three dimensions – or four or five, as in EFQM, Baldrige, BSC;

(2) Determine the list of the most representative measures associated with each dimension;

(3) For each of the measures selected, identify which other control variables might be impacted negatively (e.g. counterproductive impacts: for instance, higher quality will often mean a greater initial cost or longer project duration; the same applies to cost and risk);

(4) Determine the best combination of indicators and the causal relations between them in order to build a measurement plan for the project.

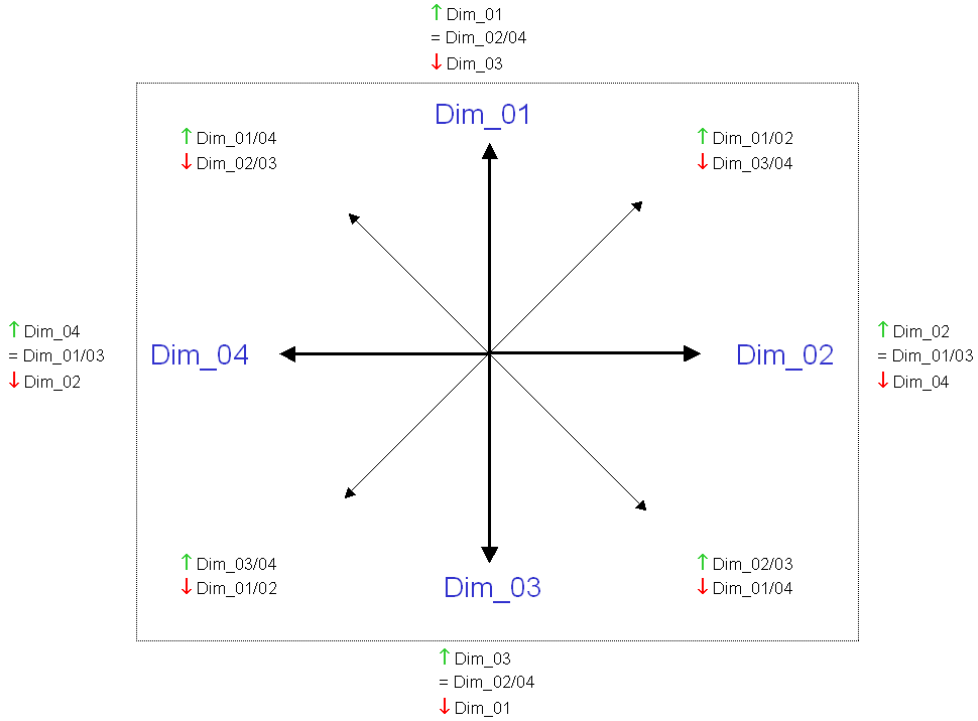
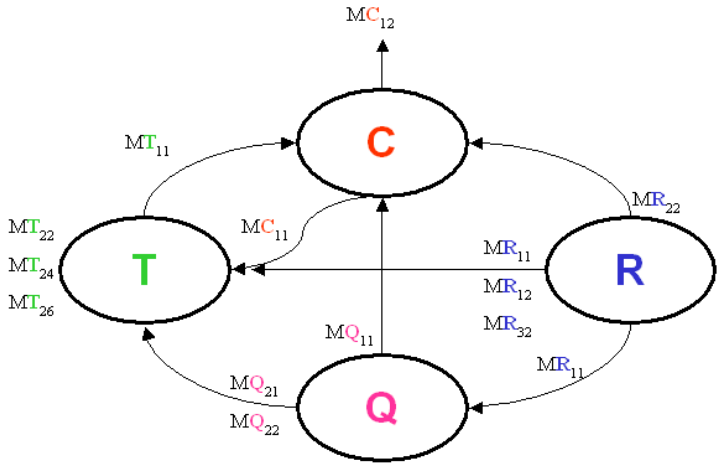


Figure 1: A generic four-dimensional BMP

2.3. Causal Relationship

It is not sufficient to perform steps (1) and (2) for designing a measurement plan within an organization, because, in such a context, this produces only a list of measures (often project goal-based, and derived and classified by dimension of analysis; e.g. time, cost, quality, risk, ethics, user satisfaction). The added value from this list can be leveraged if relationships among those goals (measured and tracked against their measures over time) are established in the planning phase of this measurement plan, realizing what Kaplan and Norton called the *strategic map* [7].

Hoffman recently asserted that "*one problem comes from a lack of relationship between the metrics and what we want to measure […]. And a second problem is the overpowering side effects from the measurement programs*" [5]. An example of linkages is presented in Fig.2, in the context of a BMP analysis with four perspectives (time, cost, quality, risk), extracting a set of core measures from the PSM Guide [3].



Time	• MT_{11} – Milestone Dates • MT_{22} – Problem Report Status • MT_{24} – Change Request Status • MT_{26} – Test Status
Cost	• MC_{11} – Earned Value • MC_{12} – Cost
Quality	• MQ_{11} – Defects • MQ_{21} – Defect Containment • MQ_{22} – Rework
Risk	• MR_{11} – Staff Experience • MR_{12} – Staff Turnover • MR_{22} – Functional Change Workload • MR_{32} – Resource Utilization

Figure 2: Indicators by perspective and causal impacts

3. Initial Results from a BMP application

To investigate the applicability of the BMP approach, an initial trial was conducted at the École de Technologie Supérieure (Montréal, Canada) and at the University Otto-Von-Guericke (Madgeburg, Germany). The questionnaire we used, as well as the results and the related analyses, are presented in this section. In addition, we were seeking insights from this trial on how to integrate such a procedure into project management activities.

3.1. BMP questionnaire

The questionnaire we used is presented in Appendix A (also available on the SEMQ website[1]). It is made up of four main sections:

1. Respondent profiles and viewpoints
2. Measures
3. Causal Relationships
4. Cost of the T&C process

3.2. Presentation of the samples

Two samples were obtained for this analysis. The first sample was made up of 6 graduate students registered in a Software Measurement course.in the M.Sc.A. program at the École de Technologie Supérieure (ETS) in Q1/2005, some of whom were also working full-time in industry. The second sample was made up of 10 German professionals who had been involved in Software Engineering for years, and data gathered in Q1/2006. In this paper, these two samples will be referred to respectively as S1 and S2.

The BMP questionnaire was given to the respondents by the authors, who briefly outlined for them its main objectives and provided them with instructions for completing the questionnaire.

3.3. List of possible indicators

A detailed list of the measures selected for the BMP questionnaire is presented in Table 1, the purpose of which was to obtain useful information about the current and desired measurement program, both from a technical and an economics viewpoint.

Table 1 – A list of indicators from the BMP questionnaire

QUESTION	#	DESCRIPTION
1a	1	Respondent profiles by project role (# and %)
	2	Experience profiles for current project role (# and %)
1b	3	# of analysis viewpoints (OLD)
1c	4	# of analysis viewpoints (NEW)
2	1	# of selected measures (OLD)
	2	# of selected measures (NEW)
	3	# of affected viewpoints (NEW)
	4	Average (avg) number of measures by viewpoint (# and %)
	5	Ranking of selected measures by: abs value, respondent project role, analysis viewpoint
3a	1	List of causal relationships among measures
	2	Ranking of relationships by: abs value, respondent project role, analysis viewpoint
4a	1	% of respondents knowing the cost of M&C (monitoring and control) activities
4b	1	Max, Min, Avg and Med for the returned values (%) – OLD
4c	1	Max, Min, Avg and Med for the returned values (%) – NEW

[1] http://www.geocities.com/lbu_measure/qestlime/bmp.htm

3.4. Presentation of questionnaire results and related analysis

In the following subsections, the results are presented for each of the respondents (R1, R2, etc.) for the two samples against the measures listed in Table 1[2].

3.4.1. Question 1 – Respondent profiles and viewpoints

Table 2a – Respondent profiles by project role and experience profiles for current project role – Sample S1

1a	In the project(s) you worked on, you contributed in the capacity of (*stress your current role*):							1a.1		1a.2
	Role	R1	R2	R3	R4	R5	R6	#	%	M(Yr)
	Project Manager	4	0	0	0	0	0	1	17%	4.0
	Team Leader	0	0	0	0	0	0	0	0%	---
	Quality Assistant	0	0	0	0	0	0	0	0%	---
	Developer	0	3	0	4	6	4	4	67%	4.3
	Tester	0	0	5	0	0	0	1	17%	5.0
	Other	0	0	0	0	0	0	0	0%	---

In terms of respondent demographics, the S1 respondents were mostly developers (67%), followed by project managers, testers and other roles (17%), while there was no team leader or quality assistant. In terms of years of experience, developers had an average 4.3 years of experience, project managers 4 years and testers 5.0 years.

Table 2b – Respondent profiles by project role and experience profiles for current project role – Sample S2

1a	In the project(s) you worked on, you contributed in the capacity of (*stress your current role*):											1a.1		1a.2
	Role	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	#	%	M(Yr)
	Project Manager	15	2	0	12	5	0	0	0	5	8	6	35%	7.8
	Team Leader	0	0	0	5	0	0	0	0	0	0	1	6%	5.0
	Quality Assistant	4	7	0	0	0	5	0	0	0	0	3	18%	5.3
	Developer	10	1	0	0	0	0	5	8	0	0	4	24%	6.0
	Tester	8	0	0	0	0	0	0	0	0	0	1	6%	8.0
	Other	0	0	5	0	0	6	0	0	0	0	2	12%	5.5

S2 respondents are more experienced people, and they are mostly project managers (35% with an average close to 8 years of experience), followed by developers (24% with 6 years of experience) and Quality Assistants (18% with an average 5.3 years of experience). The sample also included other roles (tester, team leader, others).

Table 3a – Number of analysis viewpoints (current or past project) – Sample S1

1b	How many viewpoints were usually managed for monitoring & control*ling* such project(s)?										
	Viewpoint	R1	R2	R3	R4	R5	R6	#	%	Rank	Mx
	Time	x	x		X	x	x	5	83%	1	
	Cost	x	x		X	x		4	67%	2	
	Quality	x			X	x	x	4	67%	2	
	Risk	x						1	17%	4	
	Other(1)	x						1	17%	4	
	Other(2)							0	0%	6	
	1b.3	5	2	0	3	3	2				2,5

Tables 3a and 3b present the viewpoints taken into account in the current or previous projects, and Tables 4a and 4b the expectations of which viewpoints should be taken into account.

In S1, *time* and *cost* are currently the two most common viewpoints. Quality was chosen as the second-most preferred viewpoint, followed by *user satisfaction*. Time was chosen by

[2] Note that only integers were used in percentages: therefore it is possible, due to rounding, that the sum of a series of values appears not exactly equal to 100%.

all respondents except R3 (tester), who did not choose any viewpoint. Developers chose between two and three viewpoints, while the project manager chose five. On average, the number of viewpoints chosen was 2.5, and therefore two viewpoints are usually considered for T&C activities for that dataset of respondents.

Table 3b – Number of analysis viewpoints (current or past project) – Sample S2

1b	How many viewpoints were usually managed for monitoring & control*ling* such project(s)?														
	Viewpoint	**R1**	**R2**	**R3**	**R4**	**R5**	**R6**	**R7**	**R8**	**R9**	**R10**	**#**	**%**	**Rank**	**Mx**
	Time	x	x	x	x	x		x	x	x	x	9	90%	1	
	Cost	x	x	x	x	x		x	x	x	x	9	90%	1	
	Quality	x	x	x		x	x	x		x	x	8	80%	3	
	Risk		x	x	x		x	x		x	x	7	70%	4	
	Other(1)										x	1	10%	5	
	Other(2)											0	0%	6	
	1b.3	**3**	**4**	**4**	**3**	**3**	**2**	**4**	**2**	**4**	**5**				**3.4**

Also, in S2, *time* and *cost* are the two most common and most often chosen viewpoints (90%), followed by *quality* (80%) and then *risk* (70%); only one respondent chose a further perspective[3]. A comparison of the two tables leads to the realization that there is substantial homogeneity in the currently adopted viewpoints for managing a project.

Table 4a – Number of analysis viewpoints (next project) – Sample S1

1c	Which in your opinion, whether or not previously taken into account, would need to be considered for future projects?										
	Viewpoint	**R1**	**R2**	**R3**	**R4**	**R5**	**R6**	**#**	**%**	**Rank**	**Mx**
	Time							0	0%	4	
	Cost			x			x	2	33%	2	
	Quality		x	x				2	33%	2	
	Risk		x		x		x	3	50%	1	
	Other(1)							0	0%	4	
	Other(2)							0	0%	4	
	1c.3	0	2	2	1	0	2				**1.17**

The second part of the question is what should be (or would be) added in terms of controls. Half the respondents, no matter what their project role, felt that it was more urgent to consider the *risk* viewpoint in a structured way, followed by *quality* and then *cost*.

In S1, taking into account the R3 answer, the respondents would mention *cost* and *quality* as the two most important analysis viewpoints from his/her perspective, with no reference to *time* analysis. Moreover, respondents expressed the need for one further viewpoint, on average, for analyzing future projects.

Table 4b – Number of analysis viewpoints (current or past project) – Sample S2

1b	How many viewpoints were usually managed for monitoring & controlling such project(s)?														
	Viewpoint	**R1**	**R2**	**R3**	**R4**	**R5**	**R6**	**R7**	**R8**	**R9**	**R10**	**#**	**%**	**Rank**	**Mx**
	Time											0	0%	3	
	Cost											0	0%	3	
	Quality				x				x			2	20%	1	
	Risk	x				x						2	20%	1	
	Other(1)											0	0%	3	
	Other(2)											0	0%	3	
	1b.3	**1**	**0**	**0**	**1**	**1**	**0**	**0**	**1**	**0**	**0**				**0.4**

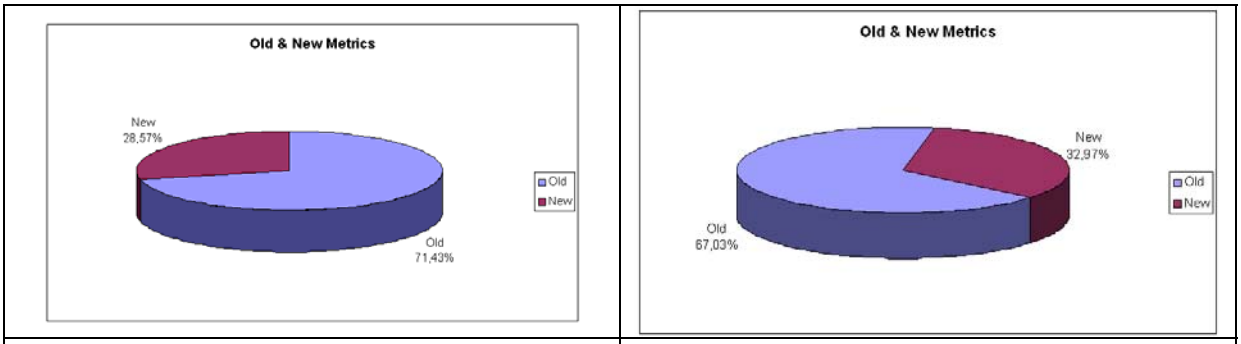
In S2, *quality* and *risk* were both considered by two respondents (20%) as possible perspectives to be considered for future projects. The distance is therefore shorter than in S1, since those two perspectives had already been implemented and become part of industrial

[3] On this point, it was noted that what was meant by "perspective" and "dimension of analysis" in the context of the current study or that of multi-perspective models was not well understood: a couple of respondents suggested something like "Change Requests" or "COTS Application" or "Productivity" as a fifth perspective, each of these pertaining to, and being assignable to, one of the four expressed viewpoints (*time*, *cost*, *quality* and *risk*). This provided us with useful indications about improvements to the BMP questionnaire and its premises and goals.

project management practices (70%, as in Table 4a); moreover, they were not only chosen by project managers, as they were in S1.

3.4.2. Questions 2 – Measures

The next group of questions concerned the measures currently used/selected and those desired in a future project. We decided to propose a sufficiently standardized set of measures, that is, the list of 67 measures/indicators suggested by the PSM (Practical Software & Systems Measurement) Guide, version 4.0b [3]. This set of measures covers at least the four viewpoints suggested in the introductory paper on BMP (*time*, *cost*, *quality* and *risk*).



(a) (b)

Figure 3 – % of selected measures (Old/New) – (a) Sample S1; (b) Sample S2

In S1, the overall number of currently selected measures was 20 out of 67 (30%, also considering those measures selected more than once), while there are just 8 new ones that people would introduce to better control future projects (equal to 17% of the remaining ones). The next table presents the detailed figures by project role.

In contrast, in S2, the overall number of currently selected measures was 61 out of 67 (91%), while the new ones numbered 30 out of 67 (45%), showing a possible deeper knowledge of benefits and possibilities to be achieved introducing new measures in future projects.

Table 5a – Number of selected measures by project role – Sample S1

Project Role	#	# OLD	# NEW	Avg # (OLD)	Avg # (NEW)	Comments
Developer	4	10	9	2.5	2.25	No usage of measures in 50% of the respondents' companies
Project Manager	1	14	1	14	1	
Tester	1	0	0	0	0	No usage of measures in his/her company

Table 6a proposes the distribution of selected measures (old and new) by viewpoint. In S1, as observed from question (1), the most frequently chosen viewpoints overall are *cost* (39%) and *time* (31%), followed by *quality* (22%), *risk* (6%) and *user satisfaction* (3%). To illustrate the degree to which *time* and *cost* are well entrenched, it can be noted that current cost-related measures number 13 out of 14, and in general 36 out of 42 measures are yet to be considered by respondents. That would lead to a conservative scenario, where new measures seems not to be so needed.

Table 5b – Number of selected measures by project role – Sample S2

Project Role	#	# OLD	# NEW	Avg # (OLD)	Avg # (NEW)	Comments
Project Manager	5	53	7	5.3	0.7	More than 7:1 (old/new) ratio
Developer	2	12	24	1.2	2.4	1:2 (old/new) ratio
QA/QM	2	31	5	3.1	0.5	c.a. 6:1 (old/new) ratio
Project Office	1	31	0	3.1	0	Really conservative

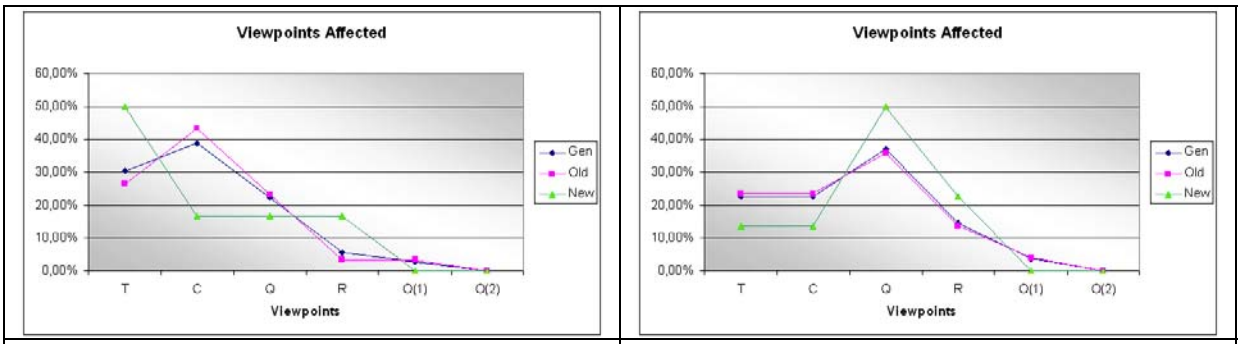
For S2, the more complex method was to clearly subdivide people by project role, because more people have worked in more roles over the years. However, in these cases, it is possible to consider the amount of time spent in a position. From Table 5b, it is possible to note that the only role category with more requests for new measures is the developer role, rather than project managers and quality assurance/management people, or the more conservative project office staff, who have yet to make their choices, or confirm their previous choices, for future projects.

Table 6a – Affected viewpoints and average number of measures by viewpoint – Sample S1

2.3	Affected viewpoints and average number of measures by viewpoint							
		T	**C**	**Q**	**R**	**O1**	**O2**	
	Gen	11	14	8	2	1	0	**Abs**
		1,83	2,33	1,33	0,33	0,17	0,00	**Avg**
		31%	**39%**	**22%**	**6%**	**3%**	**0%**	**%**
	Old	8	13	7	1	1	0	**Abs**
		1,33	2,17	1,17	0,17	0,17	0,00	**Avg**
		27%	**43%**	**23%**	**3%**	**3%**	**0%**	**%**
	New	3	1	1	1	0	0	**Abs**
		0,50	0,17	0,17	0,17	0,00	0,00	**Avg**
		50%	**17%**	**17%**	**17%**	**0%**	**0%**	**%**

Table 6b - Affected viewpoints and average number of measures by viewpoint – Sample S2

2.3	Affected viewpoints and average number of measures by viewpoint							
		T	**C**	**Q**	**R**	**O1**	**O2**	
	Gen	58	58	95	37	9	0	**Abs**
		9,67	9,67	15,83	6,17	1,50	0,00	**Avg**
		23%	**23%**	**37%**	**14%**	**3%**	**0%**	**%**
	Old	55	55	84	32	9	0	**Abs**
		9,17	9,17	14,00	5,33	1,50	0,00	**Avg**
		23%	**23%**	**36%**	**14%**	**4%**	**0%**	**%**
	New	3	3	11	5	0	0	**Abs**
		0,50	0,50	1,83	0,83	0,00	0,00	**Avg**
		14%	**14%**	**50%**	**23%**	**0%**	**0%**	**%**



(a) (b)

Figure 4 – Viewpoints affected (Old/New) – (a) Sample S1; (b) Sample S2

Tables 6a and 6b shows a substantial difference, also emphasized with trend lines in Fig.4. Concerning "old" measures, in S1, most of the attention was on *costs* (43%) followed by *time* (27%) and *quality* (23%), while in S2, the viewpoint chosen most was *quality* (about? 36%), followed by *time* and *costs* (both at 23%). Concerning "new" measures, S1 respondents (mostly developers) consider paying more attention to the *time* viewpoint in the future (50%), while in S2 (mostly project managers), the plan was to pay more attention to *quality* (50%) and *risk* (23%).

Some possible avenues for further investigation could therefore be to answer the following questions:

- Are respondents really aware of T&C mechanisms in their projects? Does their project role affect this answer?

- Could the current project measurement plan be considered satisfactory? If yes, why, in question 1c, was there a demand for the *further analysis* viewpoint (to be associated with new measures)?
- Is it possible that the PSM set of measures is not sufficiently comprehensive to consider the project respondents' requirements in the projects in which they are involved?

Staying with the measures, it is interesting to analyze which were selected more often, in terms of both currently used measures and desired measures. In order to show more significant data gathered, note that in tables from Table 7 up to Table 13 only measures selected at least by multiple respondents or assigned to more than a single perspective.

Sample S1

Table 7 – Measures selected, ranked and with detail by analysis viewpoint

# Id.	Category	Measure	Indicator	T	C	Q	R	O(1)	O(2)	Old	New	Tot
1	Milestone Performance	Milestone Dates	Dev. Milestone Schedule	2	0	1	0	0	0	2	1	3
9	Work Unit Progress	Component Status	Design Progress w/replan	1	0	1	1	0	0	1	2	3
17	Financial Performance	Earned Value	Cost & Schedule Variance	1	2	0	0	0	0	3	0	3
20	Envir.-Support Resour.	Resource Utiliz.	Resource Utilization	1	2	0	0	0	0	3	0	3
2	Milestone Performance	Milestone Dates	Milestone Progress	2	0	0	0	0	0	1	1	2
11	Work Unit Progress	Action Item Stat.	Action Item Status	1	1	0	0	0	0	2	0	2
16	Personnel	Staff Experience	Staff Experience	0	1	1	0	0	0	2	0	2
18	Financial Performance	Earned Value	Planned Cost Profile	0	2	0	0	0	0	2	0	2
40	Portability	Std. Compliance	Interface Compliance Validation	0	0	1	1	0	0	2	0	2

Table 7 presents the measures selected most often (23) grouped by frequency of selection in three chunks. The first considers 4 measures mainly assigned to the *time* and *cost* viewpoints (respectively 41.67% and 33.33%) followed by *quality*, *risk* and *user satisfaction*; the measures used most are #17 (Cost and Schedule Variance) and #20 (Resource Utilization), while the most desired new measure would be #9 (Design progress w/replan).

The second group is composed of 5 measures mainly assigned to *cost* and *time* (respectively 40% and 30%), followed by *quality*, *risk* and *user satisfaction*; the current measure used less often (and also the most desired new measure) would be #2 (Milestone Progress).

Finally, the third group is composed of 14 measures (not shown in Table 7, according to the premise done before) mainly assigned to *cost* and *time* (respectively 42.86% and 21.43%), followed by *risk*, *quality* and *user satisfaction*; the most desired new measure would be #3 (Maintenance Activities) and 23 (SW Size – Lines of Code).

Now that the overall data have been presented, the next tables propose the same analysis by project role, in order to identify some of the peculiarities or commonalities that exist, no matter what the role. Let us start with developers (n=4).

Table 8 – Measures selected, ranked and with detail by analysis viewpoint – Developers

# Id.	Category	Measure	Indicator	T	C	Q	R	O(1)	O(2)	Old	New	Tot
1	Milestone Performance	Milestone Dates	Dev. Milestone Schedule	1	0	1	0	0	0	1	1	2
16	Personnel	Staff Experience	Staff Experience	0	1	1	0	0	0	2	0	2
20	Envir.-Support Resour.	Resource Utiliz.	Resource Utilization	1	1	0	0	0	0	2	0	2
40	Portability	Std Compliance	Interface Compliance Validation	0	0	1	1	0	0	2	0	2

The first considered 4 measures were mainly assigned to *quality* (37.5%), followed by *time*, *cost* (20%) and then *risk* and *user satisfaction*; the most used measures are #16 (Staff Experience), #20 (Resource Utilization) and #40 (Interface Compliance Validation), while the most desired new measure would be #1 (Development Milestone Schedule). Residual 10

measures (single respondent or perspective selected, not listed therefore in Table 8) were mainly assigned to the *time* and *cost* viewpoints (46.875% each) followed by *quality*, *risk* and *user satisfaction*; the most used measures are #9 (Design Progress w/replan), #10 (Subsystem Acceptance Status), #12 (Incremental Content), #17 (Cost and Schedule Variance) and #20 (Resource Utilization), while the most desired new measures would be #2 (Milestone Progress), #3 (Maintenance Activities) and #23 (SW Size – Lines of Code). This last response is quite relevant, since none of the respondents included a size measure among the measures currently used, and, in the overall ranking, it only appears in the third chunk of measures chosen by frequency of use. Moreover, it reveals (or is perhaps simply a symptom of) the way in which estimations and project tracking are carried out: often without taking into account a dimensional measure for sizing the project, with an intensive use of effort and cost figures as well for doing it, as they would be perfect substitutes for size. It is also interesting to see this kind of request from developers, and that measure #23 has been associated with the "cost analysis"? viewpoint and with another at the same time (i.e. with *time* for using it for tracking purposes, or with *quality* for defectability measures).

Table 9 presents the detail from the project managers (n=1). The 3 most measures have been mainly assigned to *time* and *quality* (both with the 33%), followed by quality and risk (both with 17%) and finally user satisfaction; the most used measures are #11 (Action Item Status) and #17 (Cost and Schedule Variance) while the most desired new measure would be the #9 (Design progress w/replan).

Table 9 – Measures selected, ranked and with detail by analysis viewpoint – Prj Managers

# Id.	Category	Measure	Indicator	T	C	Q	R	O(1)	O(2)	Old	New	Tot
9	Work Unit Progress	Component Status	Design Progress w/replan	0	0	1	1	0	0	0	2	2
11	Work Unit Progress	Action Item Status	Action Item Status	1	1	0	0	0	0	2	0	2
17	Financial Performance	Earned Value	Cost & Schedule Variance	1	1	0	0	0	0	2	0	2

The third group of respondents, the testers, cannot be analyzed in detail because their companies did not make use of measures (Table 5a).

Another *static* observation can be made in terms of the differences in the selection of the measures among groups. The number of measures chosen by both groups (developers and project managers) is 6 out of the 14 each group selected (43%), showing a quite high degree of commonality in choosing control mechanisms, no matter what the analysis perspective (these measures are #1, #2, #9, #17, #18, #28), with a proper distribution on the *time* (#1, #2, #9, #17) and *cost* (#17, #18) dimensions and few on *quality* (#28).

From the project manager's viewpoint, there was instead a greater emphasis on personnel measures (#13, #14, #15), on Problem Report-Change Request (PR-CR) status (#5, #6, #28) for re-estimation purposes, action item status (#11), earned value in terms of cost variance (#19) and customer satisfaction (#63). In contrast, from the developer's viewpoint, more attention seems to be paid to milestone progress (#2), acceptance status of single subsystems (#10), the content put into further revisions of a sub-system (#12), the staff experience (#16) of personnel involved in the project team and – obviously – on typical defectability indicators impacting on the *quality* dimension (#29, #40, #57), including project size (#23).

Sample S2

The same analysis was conducted on the S2 respondents.

Table 10 – Measures selected, ranked and with detail by analysis viewpoint

# Id.	Category	Measure	Indicator	T	C	Q	R	O(1)	O(2)	Old	New	Tot

2	Milestone Performance	Milestone Dates	Milestone Progress	7	2	3	1	0	0	12	1	13
4	Work Unit Progress	Prob. Report Status	PR Status	4	3	3	3	0	0	13	0	13
9	Work Unit Progress	Component Status	Design Progress w/replan	3	2	4	2	1	0	11	1	12
1	Milestone Performance	Milestone Dates	Dev. Milestone Schedule	6	1	2	2	0	0	11	0	11
13	Personnel	Effort	Effort Allocation w/replan	4	3	0	3	1	0	11	0	11
16	Personnel	Staff Experience	Staff Experience	2	3	3	3	0	0	7	4	11
28	Functional Size-Stabil	Funct. Change WL	CRs by Priority	3	2	1	2	0	0	8	0	8
63	Customer Feedback	Survey Results	Customer Satisfaction Survey	0	0	7	1	0	0	6	2	8
21	Physical Size-Stability	Interfaces	Interface Stability	1	0	4	2	0	0	2	5	7
30	Functional Correctness	Defects	Defect Density	0	1	4	2	0	0	5	2	7
3	Milestone Performance	Milestone Dates	Maintenance Activities	4	2	0	0	0	0	6	0	6
11	Work Unit Progress	Action Item Status	Action Item Status	2	1	1	1	1	0	6	0	6
12	Incremental Capability	Increment Content	Incremental Content	3	0	2	0	1	0	6	0	6
25	Functional Size-Stabil	Requirements	Requirements Stability	1	1	2	1	1	0	5	1	6
29	Functional Correctness	Defects	Severity-1 defect status	0	1	4	1	0	0	5	1	6
31	Functional Correctness	Defects	Defect Density in Code Inspections	0	1	4	1	0	0	6	0	6
46	Process Compliance	Ref. Model Rating	Ref. Model Level – Continuous type	2	2	2	0	0	0	6	0	6
48	Process Compliance	Proc. Audit Find.	Process Audit Findings	1	1	3	1	0	0	6	0	6
10	Work Unit Progress	Component Status	Subsystem acceptance status	1	1	2	0	1	0	4	1	5
17	Financial Performance	Earned Value	Cost & Schedule Variance	2	3	0	0	0	0	5	0	5
18	Financial Performance	Cost	Planned Cost Profile	1	3	0	1	0	0	5	0	5
23	Physical Size-Stability	Lines of Code	SW Size – Lines of Code	2	2	1	0	0	0	5	0	5
49	Process Compliance	Proc. Audit Find	Audit Findings by Reason code	1	1	2	1	0	0	5	0	5
6	Work Unit Progress	Prob. Report Status	PR Status – Open by Priority	1	0	2	1	0	0	3	1	4
22	Physical Size-Stability	Lines of Code	SW Size by Config. Item	2	1	1	0	0	0	4	0	4
24	Physical Size-Stability	Physical Dim.	Electrical Power Budget	0	2	1	1	0	0	4	0	4
38	Efficiency	Timing	Response Time – OL Functions	0	0	3	1	0	0	4	0	4
40	Portability	Std. Compliance	Interface Compliance Validation	0	1	3	0	0	0	4	0	4
56	Technology Suitability	Req. Coverage	Technology Fit – Trends	1	1	1	1	0	0	4	0	4
5	Work Unit Progress	Prob. Report Status	PR Aging – Open PRs	0	0	2	1	0	0	3	0	3
14	Personnel	Effort	Effort Allocation by Dev. Activity	1	1	0	0	1	0	3	0	3
39	Efficiency	Timing	Response Time During Test – OL Fun	0	1	2	0	0	0	3	0	3
41	Usability	Operator errors	PR by Type of Problem data	0	0	2	1	0	0	3	0	3
42	Usability	Operator errors	Operator Error distrib. (by reason?)	0	0	2	0	1	0	3	0	3
50	Process Efficiency	Productivity	SW Productivity – Hist. vs. Propos	0	3	0	0	0	0	2	1	3
54	Process Effectiveness	Rework	Rework Effort by Activity	1	2	0	0	0	0	3	0	3
55	Technology Suitability	Req. Coverage	Critical Tech. Requirements	0	1	1	1	0	0	3	0	3
59	Impact	Tech. Impact	Estimated Yearly Maintenance Cost	0	3	0	0	0	0	3	0	3
19	Financial Performance	Cost	Cost Profile w/Actual Costs	0	2	0	0	0	0	2	0	2
20	Envir.-Support Resour.	Resource Utiliz.	Resource Utilization	0	1	1	0	0	0	2	0	2
33	Functional Correctness	Defects	Defect Density Distribution	0	0	2	0	0	0	2	0	2
35	Supportability-Mainten	Time to Restore	Mean Time to Repair or Fix	0	1	1	0	0	0	1	1	2
36	Supportability-Maint.	Cyclomatic Complex.	SW Complex. – CI "A"	0	0	1	1	0	0	2	0	2
37	Supportability-Maint.	Cyclomatic Complex.	SW Complex. – CI "A" w/complex.>10	0	0	1	1	0	0	2	0	2
52	Process Effectiveness	Defect Containment	Req's Def. discovered after Req Ph	0	0	2	0	0	0	2	0	2
53	Process Effectiveness	Rework	Dev. Effort by Activ. vs. Tot. Rew. Eff.	0	1	1	0	0	0	2	0	2
57	Impact	Tech. Impact	Mean Processing Time	0	0	2	0	0	0	2	0	2
66	Customer Support	Req. for Support	Total Calls per Month by Priority	0	0	2	0	0	0	2	0	2
67	Customer Support	Req. for Support	Mean Response Time by Priority	0	0	2	0	0	0	1	1	2

As seen above, in S2, almost all the measures were selected, proposing eleven chunks. Table 10 proposes some of the main points within the *time* viewpoint in terms of milestones and allocation of human resources, followed by *customer satisfaction* and *reduction of product defectability*, directly and also indirectly affected? by greater interface stability.

Table 11 – Measures selected, ranked and with detail by analysis viewpoint – Prj Managers

# Id.	Category	Measure	Indicator	T	C	Q	R	O(1)	O(2)	Old	New	Tot
1	Milestone Performance	Milestone Dates	Dev. Milestone Schedule	3	2	0	1	0	0	6	0	6
12	Incremental Capability	Incremental Content	Incremental Content	2	2	0	1	0	0	5	0	5
46	Process Compliance	Ref. Model Rating	Ref. Model Level – Continuous type	2	3	0	0	0	0	5	0	5
2	Milestone Performance	Milestone Dates	Milestone Progress	3	0	0	1	0	0	4	0	4
4	Work Unit Progress	Prob. Report Stat	PR Status	3	0	1	0	0	0	3	1	4
11	Work Unit Progress	Action Item Stat.	Action Item Status	2	0	1	0	0	0	3	0	3
18	Financial Performance	Cost	Planned Cost Profile	0	0	2	1	0	0	1	2	3
30	Functional Correctness	Defects	Defect Density	1	1	1	0	0	0	3	0	3
47	Process Compliance	Ref. Model Rating	.Model Level – Staged type	0	0	3	0	0	0	2	1	3
5	Work Unit Progress	Prob. Report Stat	PR Aging – Open PRs	0	0	2	0	0	0	1	1	2
7	Work Unit Progress	Prob. Report Stat	PR Status – Open Priority 1/2 by CI	0	2	0	0	0	0	1	1	2
9	Work Unit Progress	Component Status	Design Progress w/replan	1	1	0	0	0	0	2	0	2
16	Personnel	Staff Experience	Staff Experience	1	0	1	0	0	0	1	1	2
19	Financial Performance	Cost	Cost Profile w/Actual Costs	0	0	2	0	0	0	2	0	2
21	Physical Size-Stability	Interfaces	Interface Stability	1	0	1	0	0	0	2	0	2
23	Physical Size-Stability	Lines of Code	SW Size – Lines of Code	1	1	0	0	0	0	2	0	2

25	Functional Size-Stability	Requirements	Requirements Stability	1	1	0	0	0	0	2	0	2
31	Functional Correctness	Defects	Defect Density in Code Inspections	0	0	1	1	0	0	2	0	2
40	Portability	Std Compliance	Interface Compliance Validation	1	1	0	0	0	0	2	0	2
49	Process Compliance	Proc. Audit Find.	Audit Findings by Reason code	1	1	0	0	0	0	2	0	2
56	Technology Suitability	Req. Coverage	Technology Fit – Trends	0	0	2	0	0	0	1	1	2
59	Impact	Tech. Impact	Estimated Yearly Maintenance Cost	0	0	2	0	0	0	2	0	2
64	Customer Feedback	Perform. Rating	Composite Perform. Award Scores	0	2	0	0	0	0	2	0	2

Taking into account only the project manager's viewpoint, which includes nine chunks, greater attention is given to the first five items, all already applied (old), that concern the *time* viewpoint, with the exception of the Process Compliance rating using continuous-type models (such as CMMI or SPICE), followed by the fourth chunk of measures, which affords more in the *quality* viewpoint.

Table 12 – Measures selected, ranked and with detail by analysis viewpoint – Developers

# Id.	Category	Measure	Indicator	T	C	Q	R	O(1)	O(2)	Old	New	Tot
34	Supportability-Maint.	Time to Restore	Syst. Failures and Restoration	1	1	1	1	0	0	4	0	4
2	Milestone Performance	Milestone Dates	Milestone Progress	1	0	1	0	0	0	2	0	2
4	Work Unit Progress	Prob. Report Status	PR Status	1	0	1	0	0	0	2	0	2
11	Work Unit Progress	Action Item Status	Action Item Status	1	0	1	0	0	0	2	0	2
23	Physical Size-Stability	Lines of Code	SW Size – Lines of Code	1	0	1	0	0	0	2	0	2
35	Supportability-Maint.	Time to Restore	Mean Time to Repair or Fix	0	0	1	0	1	0	2	0	2

Developers pay more attention to the maintainability of the software from all four viewpoints. The second chunk of measures concerns the *time* viewpoint almost entirely, while the third focuses more on *quality*.

Table 13 – Measures selected, ranked and with detail by analysis viewpoint – QA/QM

# Id.	Category	Measure	Indicator	T	C	Q	R	O(1)	O(2)	Old	New	Tot
9	Work Unit Progress	Component Status	Design Progress w/replan	2	2	2	2	0	0	8	0	8
2	Milestone Performance	Milestone Dates	Milestone Progress	2	2	1	1	0	0	6	0	6
4	Work Unit Progress	Prob. Report Status	PR Status	1	1	2	2	0	0	6	0	6
28	Functional Size-Stability	Funct. Change WL	CRs by Priority	2	1	1	2	0	0	6	0	6
46	Process Compliance	Ref. Model Rating	Ref. Model Level – Continuous type	2	2	2	0	0	0	6	0	6
13	Personnel	Effort	Effort Allocation w/replan	2	1	0	2	0	0	5	0	5
1	Milestone Performance	Milestone Dates	Dev. Milestone Schedule	1	1	1	1	0	0	4	0	4
11	Work Unit Progress	Action Item Stat.	Action Item Status	1	1	1	1	0	0	4	0	4
16	Personnel	Staff Experience	Staff Experience	1	1	1	1	0	0	4	0	4
18	Financial Performance	Cost	Planned Cost Profile	1	2	0	1	0	0	4	0	4
25	Functional Size-Stability	Requirements	Requirements Stability	1	1	1	1	0	0	3	1	4
30	Functional Correctness	Defects	Defect Density	0	1	1	2	0	0	3	1	4
48	Process Compliance	Proc. Audit Find.	Process Audit Findings	1	1	1	1	0	0	4	0	4
49	Process Compliance	Proc. Audit Find.	Audit Findings by Reason code	1	1	1	1	0	0	4	0	4
3	Milestone Performance	Milestone Dates	Maintenance Activities	2	1	0	0	0	0	3	0	3
23	Physical Size-Stability	Lines of Code	SW Size – Lines of Code	1	1	1	0	0	0	3	0	3
24	Physical Size-Stability	Physical Dim.	Electrical Power Budget	0	1	1	1	0	0	3	0	3
29	Functional Correctness	Defects	Severity-1 defects status	0	1	1	1	0	0	3	0	3
31	Functional Correctness	Defects	Defect Density in Code Inspections	0	1	1	1	0	0	3	0	3
63	Customer Feedback	Survey Results	Customer Satisfaction Survey	0	0	2	1	0	0	2	1	3
5	Work Unit Progress	Prob. Report Status	PR Aging – Open PRs	0	0	1	1	0	0	2	0	2
6	Work Unit Progress	Prob. Report Status	PR Status – Open by Priority	0	0	1	1	0	0	2	0	2
36	Supportability-Maint.	Cyclomatic Complex.	SW Complex. – CI "A"	0	0	1	1	0	0	2	0	2
37	Supportability-Maint.	Cyclomat. Complex.	SW Complex. – CI "A" w/complex.>10	0	0	1	1	0	0	2	0	2
54	Process Effectiveness	Rework	Rework Effort by Activity	1	1	0	0	0	0	2	0	2
55	Technology Suitability	Req. Coverage	Critical Tech. Requirements	0	0	1	1	0	0	2	0	2

Quality people (both QA and QM) express their *assurance* side more than their *improvement* side, as noted from the first four measures. The first new measure requested to be introduced is Requirements Stability (ranked 11th). The last measure in the ranking (not appearing in the table) is the Mean Response Time by Priority in Providing Support to Customers, preceded by several internal measures.

3.4.3. Questions 3 – Causal Relationships

Sample S1

Question (3) was not answered by any of respondents, and this *consistent silence* was a clear indication of how measures are usually used in a measurement program. One question not asked in the questionnaire (maybe it will be added) is whether or not the respondent company was ISO 9001:2000 certified, or whether or not it had an SPI program (i.e. using CMMI or SPICE models). Also in the ISO standard, which lists a series of *requirements* to be accomplished, is Clause 8 (*Measurement, analysis and improvement*), which introduces the need for continuous improvement through a measurement process. The analysis of data would require at all levels some hypothesis about what are, or should be, the main relationships among the QMS (Quality Management System) processes in order to adopt corrective and/or preventive actions. This is the well-known PDCA (Plan-Do-Check-Act) schema.

Sample S2

Question (3) was not answered by 4 out of 10 ICT professionals. The other 6 respondents pointed to measures across the whole SLC, process compliance indicators, as a result of a series of causal relationships moving towards excellence, correlation between customer feedback vs. technical measures and maintenance costs vs. baseline changes.

3.4.4. Questions 4 – Cost of T&C process

Sample S1

Concerning the current cost of the T&C process (question (4a)), only one respondent (a developer) out of 6 stated that he/she had a rough idea about that figure. This answer was expected to be answered by quite all respondents: this can be a signal that this kind of project costs are not properly tracked during the project lifetime, but considered part of the more general "project management" cost item.

Concerning the future (question (4c)), respondents provided an expectation of budget allocation for T&C activities, 5 out of 6 of them providing their answers: between 20% and 5% of the project budget, with an average of 10% (the respondent proposing a higher value was, once again, a developer).

Sample S2

Concerning the current cost of the T&C process (question (4a)), 7 out of 10 respondents had an idea of the approximate cost; the three individuals with no idea at all were the two developers and one of the two QA, confirming that T&C is a process mostly managed by project managers, supported by a few other project roles (in our case, a quality manager – not a quality assistant – with a project office component) within a typical ICT organization.

Concerning the future (question (4c)), respondents provided an expectation of budget allocation for T&C activities. Eight out of 10 people provided their answers: between 35% and 5% of the project budget, with an average of 17% (the respondent proposing a higher value was, once again, and for this sample as well, a developer).

4. Conclusions & Prospects

One of the problems that arises when discussing T&C in software projects is the amount of budget to allocate in absolute terms, with little room for evaluating whether or not there is a proper balance in terms of perspectives for these controls. Usually, the two perspectives most often involved are time and cost, while others, such as *quality*, *risk*, *ethics* and so on, are

occasionally taken into account, and possibly assigned the responsibility for any additional costs for new controls to implement on projects. But the key to optimizing T&C activities, making projects more profitable, is not to eliminate controls, but to balance them, by attempting to cover and balance more viewpoints than simply time and cost.

This paper presented an application of the criteria for proper use of **BMP** (**B**alancing **M**ultiple **P**erspectives), introducing a set of possible measures for data gathering and analysis based on the BMP questionnaire, which was tested by means of two samples: S1, composed of 6 Canadians in a M.Sc.A. program in Software Engineering with some industrial experience; and S2, composed of 10 German ICT professionals, with greater experience in Software Engineering, and working in large companies or as consultants.

The initial results stressed that, in terms of desired perspectives, *risk* would be the first new perspective implemented, followed by *quality*.

Concerning measures, developers would be more open to introducing new measures on projects, while project managers pay more attention to not increasing costs, and the testers interviewed did not use any measures at all in their companies. Again, the distribution of measures by viewpoint currently focuses more on the *cost* perspective (followed by *time* and *quality*), while in terms of desired distribution, the ranking would start with *time*, followed by a "pair triple" (*cost*, *quality* and *risk*). It is possible to observe that the measures more often selected from the proposed list have been assigned to the *time* perspective, in particular Earned Value.

Another indication came from question (3): nobody in S1 responded on the subject of a causal relationship established among measures in their measurement programs, while almost all those in S2 provided an opinion, revealing several views on this issue. In any case, combining the answers to questions (2) (in terms of % of old/new measures, close to 70:30 for both samples) and (3), a possible improvement to measurement plans could be to add a section, after listing and describing the measures to be used in the project, designing a sort of *strategic map*, such as the one proposed in Fig. 2. One of the reasons for this is to prevent a project manager from continuing to choose the same project measures without discussing them again, and verifying at all times that they are profitably applicable to the new project before it begins.

Finally, concerning the cost of the T&C process, the perception of how much is currently spent on it is probably higher than the reality (25% of the project budget), with an expectation for the future of a reduction (an average of 10% for sample S1 respondents and higher – 17% – for S2 respondents).

The comparison of the results from the two samples provided useful information and confirmed the initial hypothesis on management measures, by experience and by project role. The particular experience of the S2 respondents in Software Engineering and Measurement probably overemphasized the perceived relevance of the *quality* viewpoint, but a homogeneity of project roles was noted, where developers have only visibility on their duties, with no particular information on project costs. Another point stressed by the S2 respondents, with their greater experience, is the larger number of measures selected and used and the corresponding estimated % of T&C costs to be budgeted for future projects, because of their ROI (see, in particular, the focus on Process Compliance measures).

Future work on BMP developments will involve further investigation through the application of the BMP questionnaire, and, after gathering an appropriate amount of data, a study of how to use the BMP as a tool to facilitate definition of the BSC strategy map in terms of the counter-effects of choosing indicators for each perspective, and of mapping them to the possible dimensions of analysis (e.g. *time*, *cost*, *quality*, *risk*, etc.) to achieve double-check balancing.

5. References

[1] BUGLIONE L. & ABRAN A., *Multidimensional Project Management T&C - Related Measurement Issues*, Proceedings of SMEF 2005, Software Measurement European Forum, 16-18 March 2005, Rome (Italy), pp. 205-214

[2] DEMARCO T., *Why Does Software Cost So Much?: And Other Puzzles of the Information Age*, Dorset House, 1995, ISBN 093263334X

[3] DEPT. OF DEFENSE & US ARMY, *PSM - Practical Software & Systems Measurement. A Foundation for Objective Project Management*, Version 4.0c, March 2003, URL: http://www.psmsc.org

[4] EFQM, *The EFQM Excellence Model*, European Foundation for Quality Management, 1999, URL: http://www.efqm.org/publications/EFQM_Excellence_Model_2003.htm

[5] HOFFMAN D., *The Darker Side of Metrics*, Pacific Northwest Software Quality Conference, 2000, URL: http://softwarequalitymethods.com/SQM/Summaries/DarkerSideMetrics.html

[6] JONES C., *Applied Software Measurement: assuring productivity and quality*, 2/e, McGraw-Hill, 1996, ISBN 0070328269

[7] KAPLAN R.S. & NORTON D.P., *Strategy Maps: Converting Intangible Assets into Tangible Outcomes*, Harvard Business School Press, 2004, ISBN 1591391342

[8] ROEDLER G., *Using PSM at all Levels in an Organization*, PSM Technical Working Group (TWG) 2003 Meeting, March 2003, Herndon, VA (USA), URL: http://www.psmsc.com/Downloads/TWGMarch03/Roedler_Using%20PSM_AllOrgLevels_TWG2003.pdf

[9] RUBIN H.A., *The Top 10 Mistakes in IT Measurement*, IT Metrics Strategies, Vol. II, No. 11, November 1996, URL: http://www.cutter.com/benchmark/1996toc.html