Suggestions for Improving Measurement Plans: A BMP application in Spain

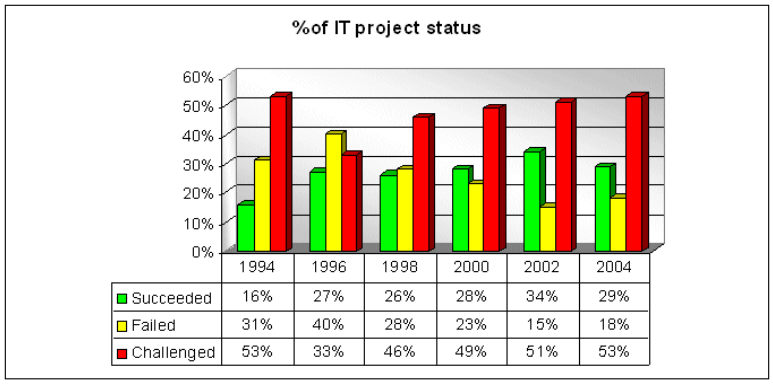
Juan José Cuadrado-Gallego, Luigi Buglione, Daniel Rodríguez

Abstract

Time and Cost are most often in industry the two main (often solely) dimensions of analysis against which a project is monitored and controlled, excluding other possible dimensions such as Quality, Risks, impact on society and Stakeholders' viewpoints in a broader sense. Another issue of interest is the proper amount of measures and indicators to implement in an organization to optimizing the two sides of the cost of quality (COQ - cost of quality - and CONQ - cost of non quality). How can multiple concurrent control mechanisms across several dimensions of analysis be balanced? The approach of Balancing Multiple Perspectives (BMP) has been designed to help project managers choose a set of project indicators from several concurrent viewpoints. After gathering experiences from Canada, Germany and Turkey, this paper presents the results from a new BMP application in Spain, using a list of 14 candidate measures interviewing a double set of respondents from industry. Lessons learned are presented for improving measurement plans.

1. Introduction

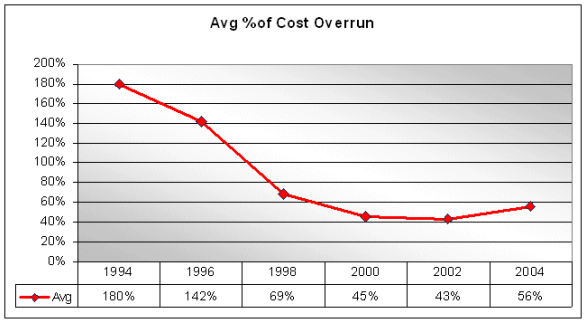
A Software Engineering topic of discussion during the last 15 years has been the identification of main project failure causes; few of these studies list directly the amount of Tracking and Control (T&C) resources, the lack of historical data and the limited ability of internal staff to estimate effort and cost as a major item [1][2]. A well-known and cited study is the Chaos Report by the Standish Group: in 1994 the 52.7% of projects cost over 189% of their original estimates and only 16.2% of software projects were completed on-time and on-budget [3].



	1994	1996	1998	2000	2002	2004
Succeeded	16%	27%	26%	28%	34%	29%
Failed	31%	40%	28%	23%	15%	18%
Challenged	53%	33%	46%	49%	51%	53%

Figure 1: Percentage of IT Project Status (1994-2004) [3]

After ten years - according always to the new Chaos Reports - in 2004 the situation seems to be slightly improved, but fundamentally the percentage of challenged projects remains stable (around 50% of surveyed projects), while the average percentage of costs and time overrun have been decreased – but in any case maintaining too high values - respectively of c.a. one third and one half.



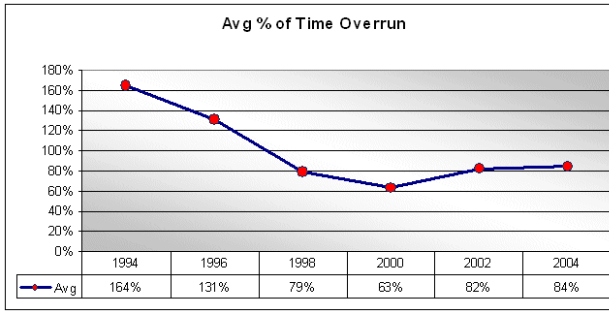


Figure 2: Average % of IT Project Costs & Time Overruns (1994-2004) [3]

It is not only necessary to analyze high-level indicators such as the ones presented in Figures 1 and 2, but also a deeper root-cause analysis to explain these trends. Furthermore, a greater attention must (or at least should) be paid to the ways a project could be more profitable and less defect-prone, but often not as much to the project budget allocated for T&C activities. This issue is not only intimately linked with the Project Managers' role skill[1], but also with all those roles involved in project effort and cost estimates. Therefore, relevant questions would be what is the project budget percentage dedicated to those activities and how much does it cost to track and control a software project. From an economic viewpoint, T&C costs can be seen as part of the Cost of Quality (COQ) - including prevention and appraisal costs - as the counterpart to the Cost of Non Quality (CONQ) – including internal and external failure costs[2]. Figure 3 represents the classical view on COQ and CONQ: the break-even point (BOP) will be reached at *t* time, optimizing the overall cost for quality (COQ+CONQ)[3]. Thus, each organization will determine and optimize its BOP in terms of time and money balancing available resources and taking into account the best number and type of measures to be managed for the T&C process choosing among different perspectives. This point, however, is not easy to calculate.

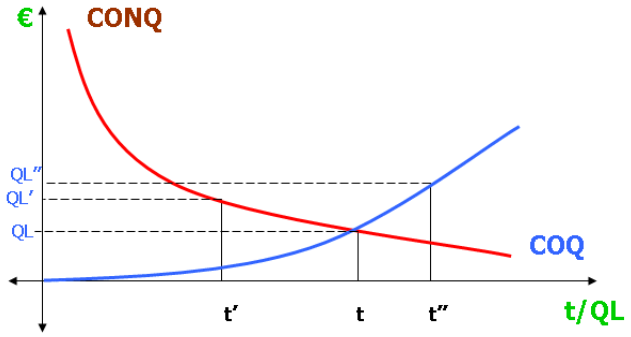


Figure 3: Relationships between COQ-CONQ along time

Also, well-known software process improvement models such as CMMI [4] and ISO 15504 [5][4] require in their specific (basic) practices to take into account the amount of effort and cost that T&C processes require, so that project managers can properly balance their

[1] More and more emphasized by the recent growing demand for Project Management certifications, such as PMI's PMP (http://www.pmi.org/prod/groups/public/documents/info/pdc_pmp.asp), Prince2 (http://www.ogc.gov.uk/methods_prince_2.asp), IPMA (http://www.ipma.ch) and AAPM (http://www.projectmanagementcertification.org/).

[2] For a detailed list of cost items to consider to compute COQ and CONQ, see [6][7].

[3] There is a huge amount of references about the effects while balancing COQ and CONQ in the Total Quality Management (TQM) literature. About a large review of COQ-related issues, see for instance [8].

[4] See also www.isospice.com

available budget across the different project phases and processes. For instance, CMMI in its staged representation requires a Project Historical Database (PHD) at its Maturity Level (ML) 3 in its OPD (Organizational Process Definition) process area, which its purpose is "*to establish and maintain a usable set of organizational process assets […]; the organization process asset library (PAL) is a collection of items […] including […] data*"; furthermore, looking a level below, ML2 processes, PP (Project Planning) in Specific Practice (SP) 1.4, sub-practice #3 states that it is needed to "*estimate effort and cost using models and/or historical data*". There is a couple of issues to highlight:

- In the recurrent hypothesis that a ML2 organization does not have or have planned to create its own PHD, a common practice is to run an extended "*external*" benchmarking process, even if MA (Measurement & Analysis) – another ML2 process – requires for collecting data as an fundamental enabler for the decision-making process.
- "and/or" means that usually ML2 organizations (but also at higher MLs) adopt in a non-critical manner estimation models such as COCOMO [9] or SLIM [10]. Even if their usage would have been performed under a "*calibration process*", as described for instance in [9][5], current literature does not provides clear data about how much does it cost to calibrate these models[6] to evaluate if they allow project managers to initially save time and cost before implementing their own database and estimation models.

In both cases, one direct consequence is to reduce the probability to improve our estimation ability and therefore the overall profitability of the organization. Thus, the problem is not solved, but only shifted, because no matter the (corrective) action taken, we do not control and understand which is the proper level of costs to allocate in our budget for planning, monitoring & control projects in an organization [11]. Demarco [12] stated 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*". In one of the few studies carried out in the '90s proposing actual figures, Jones [13] reported the costs of measurement in projects to be approximately between 3% and 6% for internal projects measurement and between 2.5% and 4.5% for the external ones.

Again, two out of ten problems leading to failure in the implementation of software measurement programs are reported by Rubin to be the intensive use of a single measure or, conversely, the use of too many [14]. 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 (e.g., 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* [15]. It includes a questionnaire with a list of 14 candidate measures from 4 sections (respondents profiles and viewpoints, measures, causal relationships, cost of the T&C process), with the objectives of representing the "*as is*" situation and determining what the "*to be*" situation will be,

[5] See Chapter 4.
[6] See a recent thread on the IFPUG bulletin board: http://www.ifpug.org/discus/messages/1778/8469.html

including cost figures to be possibly considered in future project budgets [16]. Since the maturity level when using and applying measurements can vary a lot among countries (i.e. educational programs in Software Engineering, ICT market demands, cultural resistance to measurement, etc.), it was decided to extend the experimentation to other countries to observe other possible attitudes and perceptions. BMP was proposed during 2005 and first results were gathered and presented in the first semester 2006 (H1/2006) by two sets of Canadian and German respondents, respectively MSc, BSc and PhD students and ICT professionals [17]. A second BMP experience was conducted in Turkey during 2006 by other two sets of respondents, Turkish ICT professionals and MSc or PhD degree graduates [18].

In this paper, we present a Spanish BMP application in a sample from Information and Communication Technology (ICT) Industries. The aim of this study is twofold. On the one hand, to stimulate the discussion in the technical community about which is the proper level of costs for properly supporting the measurement process to achieve established goals and, on the other hand, to take care of possible elements for corrective/improvement actions, mainly working on the cause-effect links in the company's process strategic map.

The paper is organized as follows: Section 2 presents the BMP, its objectives and the related procedure. Section 3 presents results from the survey and the assumptions under which it 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. A first indirect evidence can be verified through Gantt charts or Project Plans, where T&C is often planned as activity and not as process. Using the Plan- Do-Check-Act (PDCA) schema, costs for the "*Plan*" phase are usually not considered, not including a series of micro tasks about coordination activities during the project lifetime and subsequent controls in the "*Act*" phase before arriving to the "*Check*" one. From an the SPICE perspective, the clause less accomplished is Clause 8, the one about Measurement and Improvement[7].

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 [15] is that increasing performance does not need to be limited to *reducing cost,* but it 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. It is obvious that increasing the number of controls increases 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.

[7] See [6] about the Italian situation.

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, 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.

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 [19], Baldrige [20] and BSC [21];

(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 to build a measurement plan for the project.

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* [21]. 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*" [22].

3. Application of the BMP Survey in Spain

To corroborate and extend the lessons learned on the applicability of the BMP approach, a new trial was conducted in Spain, again taking into account two sets of questionnaires collected from Industry professionals from two companies. The description of the questionnaire used, the results and the related analyses are presented in this section. In addition, we seek insights from this trial on how to integrate such a procedure into project management activities.

The questionnaire used is available on the SEMQ website[8]. It is composed of four main sections: (1) Respondent profiles and viewpoints; (2) Measures; (3) Causal Relationships and (4) Cost of the T&C process. A detailed list of the measures selected for the BMP questionnaire is presented in Table 1. The purpose was to obtain useful information about the current and desired measurement programs, both from technical and economic viewpoints.

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

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

3.1. Subjects of the Questionnaire

The sample of this study consists of 15 Spanish professionals who had been involved in Software Engineering for years, and data was gathered between Q4/06 and Q1/07. In this paper, this sample will be referred as S1. The BMP questionnaire was provided to the respondents by the authors, who briefly outlined for them its main objectives and provided them with instructions for completing the questionnaire.

3.2. Questionnaire Results and Discussion

In the following subsections, the results are presented for each of the respondents (R1, R2, etc.) against the measures listed in Table 1[9].

3.2.1. Question 1 – Respondent profiles and viewpoints

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

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	R 7	R8	R9	R10	R11	R12	R13	R14	R15	#	%	Avg (yrs)
	Project Mgr								30	5							2	8	17,5
	Team Lead.		4				1						1				3	12	2,0
	Quality Ass.											3		1			2	8	2,0
	Developer			3	1					4					0,5	1	5	20	1,9
	Tester	1	3			1	5	2	0	11	3		4		3		9	36	3,7
	Other										3	1			0,5	3	4	16	1,9

In terms of respondent demographics, the S1 respondents (n=15) were mostly Testers and Team Leaders (27%), followed by Project Managers, Developers and Quality Assistants (13%), while there was just a Systems Engineer. In terms of years of experience, Project Managers had an average 17.5 years of experience, Testers 3.7 years, Team Leaders and Quality Assistants 2 years while Developers and the Systems Engineering 1.9 years.

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

1b	**In the project(s) you worked on, you contributed in the capacity of (*stress your current role*):**																			
	Viewpoint	**R1**	**R2**	**R3**	**R4**	**R5**	**R 6**	**R 7**	**R8**	**R9**	**R10**	**R11**	**R12**	**R13**	**R14**	**R15**	**#**	**%**	**Rank**	**Avg**
	Time	x	x	x	x	x	x	x	x	x	x		x	x	x	x	14	67	1	
	Cost		x	x	x	x	x	x	x	x		x	x	x		x	12	57	2	
	Quality	x		x	x	x		x	x	x	x			x		x	10	48	3	
	Risk				x	x											2	10	4	

[9] 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%.

	Other(1)				x						x					2	10	4	
	Other(2)										x					1	5	6	
	1b.3																		2.9

Tables 3 presents the viewpoints taken into account in the current or previous projects, and Table 4 the expectations of which viewpoints should be taken into account. The variables *time* and *cost* are currently the two most common viewpoints, followed by *quality*. Time was chosen by all respondents except R11 (quality assistant), who chose *other* alternative viewpoints. On average, the number of viewpoints chosen was 2.9, and therefore two viewpoints are usually considered for T&C activities for that dataset of respondents.

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

1c	How many viewpoints were usually managed for monitoring & controlling such project(s)?																		
	Viewpoint	R1	R2	R3	R4	R5	R 6	R 7	R8	R9	R10	R11	R12	R13	R14	R15	#	%	Rank	Avg
	Time											x					1	5	5	
	Cost														x		1	5	5	
	Quality		x				x						x		x		4	19	2	
	Risk	x		x				x	x	x			x	x	x	x	9	43	1	
	Other(1)					x			x								2	10	3	
	Other(2)					x			x								2	10	3	
	1c.3																			1.1

The second part of the question is what should be (or would be) added in terms of controls. Half of 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 other perspectives (*resources*, in particular *people; safety*). With the exception of R11, respondents mentioned *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.

3.2.2. Questions 2 – Measures

The next group of questions concerned measures currently used/selected and those desired for future projects. We decided to propose a sufficiently standardized set of measures, i.e., the list of 67 measures/indicators suggested by the PSM (Practical Software & Systems Measurement) Guide, Version 4.0b [23]. This set of measures covers at least the four viewpoints suggested in the introductory paper on BMP (*time*, *cost*, *quality* and *risk*).

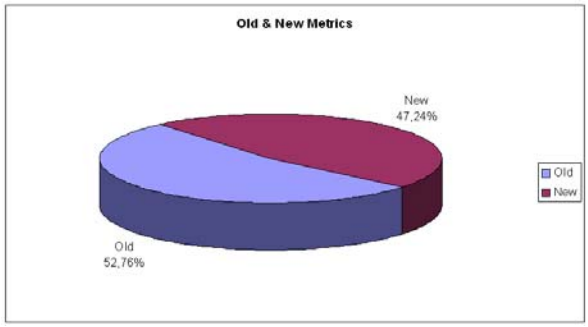


Figure 4 – % of selected measures (Old/New)

The overall number of currently selected measures was equal to 100% (67 out of 67), while 60 out of 67 measures would be introduced for a better control in future projects (90% of the total PSM proposed set). Next table presents the detailed figures by project role.

Table 5 – Number of selected measures by project role

Project Role	#	# OLD	# NEW	Avg # (OLD)	Avg # (NEW)	Comments
Developer	2	41	19	20	9	c.a. 2:1 ratio between old/new measures
Project Manager	2	30	32	15	16	
Tester	4	57	37	14	9	c.a. 3:2 ratio between old/new measures
Team Leader	4	67	10	17	4	c.a. 7:1 ratio between old/new measures
Quality Assistant	2	25	25	12	12	
Systems Engineer	1	21	5	21	5	c.a. 4:1 ratio between old/new measures

Table 5 proposes the distribution of selected measures (old and new) by viewpoint. There is a general tendency to be more focused on "old" measures, with a "conservative" approach not really devoted to apply new controls. In particular, the most defensive roles seem to be the Team Leaders group and the Systems Engineer, followed by Developers.

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

2.3	Affected viewpoints and average number of measures by viewpoint							
		T	C	Q	R	O1	O2	
	Gen	269	245	279	136	52	37	**Abs**
		44,83	40,83	46,50	22,67	8,67	6,17	**Avg**
		26,42%	**24,07%**	**27,41%**	**13,36%**	**5,11%**	**3,63%**	**%**
	Old	212	196	213	107	26	10	**Abs**
		35,33	32,67	35,50	17,83	4,33	1,67	**Avg**
		27,75%	**25,65%**	**27,88%**	**14,01%**	**3,40%**	**1,31%**	**%**
	New	57	49	66	29	26	27	**Abs**
		9,50	8,17	11,00	4,83	4,33	4,50	**Avg**
		22,44%	**19,29%**	**25,98%**	**11,42%**	**10,24%**	**10,63%**	**%**

Table 6 proposes the distribution of selected measures (old and new) by viewpoint. As observed from question (1), the most frequently chosen viewpoints overall are *quality* (27%), *time* (26%) and *cost* (24%), followed by *risk* (13%), and other secondary perspectives (9%). The same trend was observed also analyzing separately "old" and "new" measures. This is also shown graphically in Figure 5.

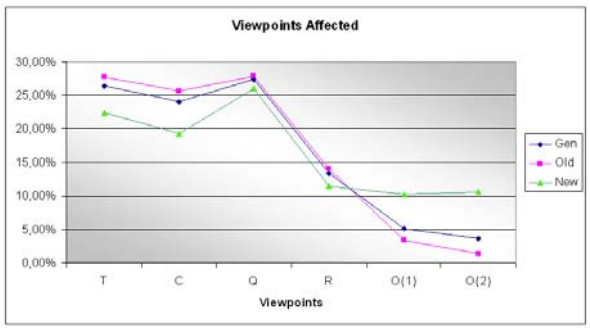


Figure 5 – Viewpoints affected (Old/New)

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.

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
34	Supportability-Mainten	Time to Restore	SysFailures and Restoration	6	7	9	6	1	0	27	2	29
1	Milestone Performance	Milestone Dates	Dev.Milestone Schedule	13	8	5	2	0	0	28	0	28
25	Functional Size-Stabil	Requirements	Requirements Stability	8	3	12	3	1	1	26	2	28
52	Process Effectiveness	Defect Containm	Req's Def. discovered after Req Ph	7	7	6	4	3	1	18	10	28
53	Process Effectiveness	Rework	Dev.Effort by Activ.vs Tot.Rew.Eff	9	7	7	2	2	1	14	14	28
2	Milestone Performance	Milestone Dates	Milestone Progress	13	7	4	3	0	0	27	0	27
15	Personnel	Effort	Staffing Level	5	9	8	4	1	0	16	11	27
20	Envir.-Support Resour.	Resource Utiliz.	Resource Utilization	7	10	6	3	0	1	18	9	27
4	Work Unit Progress	ProblReport Stat	PR Status	7	4	11	3	1	0	26	0	26
16	Personnel	Staff Experience	Staff Experience	7	6	8	4	1	0	19	7	26

Table 7 presents the measures most selected, grouped by frequency of selection in four chunks. The most chosen measure is about the time to restore, but in the second chuck there are four measures about planning, requirements and defectability. It is interesting to note (see after also comments on Question 3) the selection of Staffing level and Resource Utilization in the third chuck, showing a huge attention on resources for a proper planning. The perspectives more frequently associated to these measures are Quality and Cost (not the opposite), showing a possible interesting signal about the priorities respondents provided in their answers. In the fourth chuck this indication has been stressed again with the Staff Experience, accompanied by the analysis of Problem Report (PR) status, in order to properly track the progress of the project related to the restorability from defects. Also here the most associated perspective is Quality before Cost. A general consideration is about the Risk perspective, resulting the less associated in general: this is an interesting signal about the role that risk has in the overall estimation process, often used more from a qualitative than quantitative viewpoint and it is possible to note it also looking back at Question 1b and 1c.

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

# Id.	Category	Measure	Indicator	T	C	Q	R	O(1)	O(2)	Old	New	Tot
25	Functional Size-Stabil	Requirements	Requirements Stability	2	0	2	1	1	0	6	0	6
1	Milestone Performance	Milestone Dates	Dev.Milestone Schedule	2	2	1	0	0	0	5	0	5
2	Milestone Performance	Milestone Dates	Milestone Progress	2	2	1	0	0	0	5	0	5
3	Milestone Performance	Milestone Dates	Maintenance Activities	1	2	1	1	0	0	2	3	5
34	Supportability-Mainten	Time to Restore	SysFailures and Restoration	1	1	2	1	0	0	5	0	5

Table 8 shows the five more selected measures by Project Managers. The stability of requirements is thought to be the most relevant, with a initial and evident impact on the following three milestone dates measures. Risk is slightly associated to those measures (50% of possible answers); measure #3 (maintenance activities) was the solely one with an association to risk as a new measure. It is possible to note that quite all measures are yet actually used, with a low tendency to adopt new measures.

Table 9 – 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
20	Envir.-Support Resour.	Resource Utiliz.	Resource Utilization	1	2	2	1	0	0	4	2	6
19	Financial Performance	Earned Value	Cost Profile w/Actual Costs	1	2	1	1	0	0	5	0	5
34	Supportability-Mainten	Time to Restore	SysFailures and Restoration	2	2	1	0	0	0	5	0	5
53	Process Effectiveness	Rework	Dev.Effort by Activ.vs Tot.Rew.Eff	2	1	2	0	0	0	2	3	5
63	Customer Feedback	Survey Results	Customer Satisfaction Survey	1	1	2	1	0	0	4	1	5

Table 9 shows the selections by Developers. Resource Utilization is the most rated measure, followed by cost, time and customer issues. A couple of elements must be noted: measure #53 is mostly desired for future project and not always yet applied; measure #63 seems to be yet applied also to the development part of the project and "lived" by those people, not only by project managers as the responsible for the project in front of the external stakeholders or the internal Top Management.

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

# Id.	Category	Measure	Indicator	T	C	Q	R	O(1)	O(2)	Old	New	Tot
7	Work Unit Progress	ProblReport Stat	PR Status – Open Priority 1/2 by CI	1	1	3	2	0	0	7	0	7
8	Work Unit Progress	ProblReport Stat	PR Status – Open Priority ½ by Type	2	1	3	1	0	0	7	0	7
16	Personnel	Staff Experience	Staff Experience	2	2	3	0	0	0	4	3	7
2	Milestone Performance	Milestone Dates	Milestone Progress	3	3	0	0	0	0	6	0	6
4	Work Unit Progress	ProblReport Stat	PR Status	2	0	3	1	0	0	6	0	6
22	Physical Size-Stability	Lines of Code	SW Size by Config.Item	1	2	3	0	0	0	6	0	6
53	Process Effectiveness	Rework	Dev.Effort by Activ.vs Tot.Rew.Eff	2	3	1	0	0	0	3	3	6

Table 10 shows two main chunks, with the Team Leaders choices: the first one mostly focused on PR status as well as on Staff Experience; the second one again on PR status,

rework, milestone progress and software size (even if using LOC). Also here, rework (measure #53) is half associated both as an old/new measure.

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

# Id.	Category	Measure	Indicator	T	C	Q	R	O(1)	O(2)	Old	New	Tot
52	Process Effectiveness	Defect Containm	Req's Def. discovered after Req Ph	4	2	2	2	1	1	8	4	12
15	Personnel	Effort	Staffing Level	2	3	4	1	1	0	6	5	11
4	Work Unit Progress	ProblReport Stat	PR Status	2	2	4	1	1	0	10	0	10
67	Customer Support	Req. for Support	Mean Response Time by Priority	2	2	3	1	1	1	9	1	10
1	Milestone Performance	Milestone Dates	Dev.Milestone Schedule	3	2	3	1	0	0	9	0	9
16	Personnel	Staff Experience	Staff Experience	3	1	2	2	1	0	6	3	9
25	Functional Size-Stabil	Requirements	Requirements Stability	3	1	3	1	0	1	9	0	9
26	Functional Size-Stabil	Requirements	Req.Stability by Type of Change	2	2	2	1	1	1	9	0	9
29	Functional Correctness	Defects	Severity-1 defects status	2	2	2	2	1	0	9	0	9
34	Supportability-Mainten	Time to Restore	SysFailures and Restoration	2	2	3	2	0	0	7	2	9
53	Process Effectiveness	Rework	Dev.Effort by Activ.vs Tot.Rew.Eff	3	2	2	0	1	1	2	7	9
66	Customer Support	Req. for Support	Total Calls per Month by Priority	2	2	2	1	1	1	8	1	9

Table 11 presents the Testers' choices, with four main chunks. The most selected measure is #52, showing a strong attention to timely evidence defects from the analysis phase in order to produce less and less defects when coding the software solution. But it is interesting the following selection (measure #15), denoting an interest to use the proper staff people for a certain testing activity (quality is the most associated perspective, more than cost), even if it's half-rated as a new, desired measure, not always yet in place. Anyway, looking at all the other measures selected, it is possible to note that requirements are perceived more and more as the crucial element to work on before arriving to the CUT (Code and Unit Test) phase, in order to save time and money and increase the overall product quality for the Customers.

Another common element noted also in this group is that the most desired new measure is #53 about rework, that – from a tester's viewpoint – would mean to have more elements for properly plan testing activities along the whole software lifecycle and not only at the code level, reducing therefore also the number of PR.

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

# Id.	Category	Measure	Indicator	T	C	Q	R	O(1)	O(2)	Old	New	Tot
20	Envir.-Support Resour.	Resource Utiliz.	Resource Utilization	1	1	2	1	0	1	3	3	6
1	Milestone Performance	Milestone Dates	Dev.Milestone Schedule	2	0	1	1	0	0	4	0	4
2	Milestone Performance	Milestone Dates	Milestone Progress	1	0	1	2	0	0	4	0	4
13	Personnel	Effort	Effort Allocation w/replan	1	1	0	1	0	0	2	1	3
14	Personnel	Effort	Effort Allocation by Dev.Activity	1	1	0	1	0	0	2	1	3
16	Personnel	Staff Experience	Staff Experience	1	1	1	0	0	0	2	1	3
25	Functional Size-Stabil	Requirements	Requirements Stability	1	0	2	0	0	0	2	1	3
29	Functional Correctness	Defects	Severity-1 defects status	1	0	1	0	0	1	2	1	3
30	Functional Correctness	Defects	Defect Density	1	0	1	0	0	1	2	1	3
34	Supportability-Mainten	Time to Restore	SysFailures and Restoration	0	1	1	0	1	0	3	0	3
35	Supportability-Mainten	Time to Restore	Mean Time to Repair or Fix	0	1	1	0	1	0	3	0	3
52	Process Effectiveness	Defect Containm	Req's Def. discovered after Req Ph	1	0	1	0	1	0	2	1	3
53	Process Effectiveness	Rework	Dev.Effort by Activ.vs Tot.Rew.Eff	1	0	1	1	0	0	2	1	3
64	Customer Feedback	Perform. Rating	Composite Perfor.Award Scores	0	0	1	1	1	0	2	1	3

Table 12 proposes the selections from the two Quality Assistants interviewed. Resource Utilization was the most rated measure, followed by two milestone performance ones, yet applied in current projects. In the third chuck there are 11 measures about personnel (3), functionalities (3), supportability and maintenance (2), process effectiveness (2) and the customer feedback but with a particular view on the overall performance rating (measure #64), differently from the other stakeholders choosing measures from this group. Few new measures seem to be desired.

Table 13 – Measures selected, ranked and with detail by analysis viewpoint – Sys.Engineer

# Id.	Category	Measure	Indicator	T	C	Q	R	O(1)	O(2)	Old	New	Tot
4	Work Unit Progress	ProblReport Stat	PR Status	1	1	1	1	0	0	4	0	4
5	Work Unit Progress	ProblReport Stat	PR Aging – Open PRs	1	1	1	0	0	0	3	0	3

52	Process Effectiveness	Defect Containm	Req's Def. discovered after Req Ph	0	1	1	1	0	0	3	0	3
55	Technology Suitability	Req. Coverage	Critical Tech. Requirements	1	0	1	1	0	0	3	0	3

Last but not least, the viewpoint by the Systems Engineer interviewed. The greater attention is on PR, both in terms of status (measure #4) and their aging (measure #5), as well as about the frequencies in discovering defects after the requirement phase (measure #52). Finally, it was selected also an interesting non functional element (measure #55), where a Systems Engineer contributes in the Analysis phase. There is no new measure desired.

3.2.3. Questions 3 – Causal Relationships

Question (3) was answered by quite all respondents (13 out of 15), and this consistent answer is a clear indication of how measures are spreadly used in their measurement programs. Two main relationships identified: staff experience & milestone progress (proposed by PM, QA and the Systems Engineer) and defectability & work unit/milestone progress (proposed by developers and testers). Crossing Question 3a (which measures?) and Question 3b (why), it appeared a more visible mid-long term view by the first group (PM, QA and Systems Engineer), probably having in their DNA this kind of approach, linking clearly cause and effects, while the answers provided by the other group for motivating their choices revealed what it could be called a "day by day" planning, where achieving milestones is the main goal to satisfy and report to their managers, through a reduced defectability.

3.2.4. Questions 4 – Cost of T&C process

Concerning the current cost of the T&C process (question (4a)), only four respondents (2 project managers, 1 team leader and a tester) out of 15 stated that he/she had a rough idea about the cost. This answer was expected to be answered by most 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. About the "how much" (question (4b)), project managers were more prudent and realistic, providing a 15% indication, that's the minimum value shown, while the tester provided the higher value.

Table 14 – T&C costs: Sample S1

	Past-Current	Next	Difference
Max	30.00%	30.00%	**0.00%**
Median	17.50%	20.00%	**+2.50%**
Avg	20.00%	17.69%	**-2.31%**
Min	15.00%	5.00%	**-10.00%**

Concerning future projects (question (4c)), quite all respondents provided an expectation of budget allocation for T&C activities (13 out of 15): between 30% and 5% of the project budget, with an average of 18% (the respondent proposing a higher value was a team leader, the two project managers were also more prudent, confirming the 15% yet shown as the actual projects' T&C cost). It is worth noting that the median is higher than the average value, showing a shared willing to slightly increase the budget for the T&C process, no matter the project role actually covered.

4. Conclusions and Future Work

One of the problems when discussing Tracking and Control (T&C) in software projects is the amount of budget allocated 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*, *safefy* 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 (Balancing Multiple Perspectives), introducing a set of possible measures for data gathering and analysis based on the BMP questionnaire, which was tested by means of a samples composed of 15 experienced Spanish ICT professionals and working in large companies or as consultants. The initial results stressed that, in terms of desired perspectives, *risk* would be the first perspective to be implemented, followed by *quality*. Concerning measures, project managers would be more open to introducing new measures on projects, while team leaders pay more attention to not increasing costs and are quite conservative, as well as developers and the systems engineer interviewed. Again, the distribution of measures by viewpoint currently focuses more on the *quality* perspective (followed by *time* and *cost*), that is the same ranking also in terms of desired distribution. It is possible to observe that the measures more often selected from the proposed list have been assigned to the *quality* perspective, in particular "Systems Failure and Restoration" (measure #34).

Another indication came from question 3 (causal relationships among measures): quite all people provide an answer and it is interesting to note the link stressed between staff experience and milestone progress, before considering defectability & work unit/milestone progresses, revealing the need to considering people as a starting element in the causal chain among SLC processes. Indirectly, this attention was noted also in the ranking of perspectives, where quality was unusually ranked as #1.

Finally, concerning the cost of the T&C process (question 4), few people know how much really costs this process: the perception of how much is currently spent is probably higher than the reality (an average of 20% of the project budget), with an expectation for the future of a slight reduction (an average of 18%).

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.

Acknowledgements

We would like to thank Atos Origin Italy and the Spanish Ministry of Science and Technology for supporting this research (Project CICYT TIN2004-06689-C03).

References

[1] GILB T., *Project Failure: Some Causes and Cures*, Feb 29 2004, URL: http://www.webster.edu/ftleonardwood/COMP5940/Student_Files/Project_Failure/ProjectFailure.pdf

[2] NAO, *Delivering successful IT-enabled business change*, Report, UK National Audit Office, November 2006, URL: www.nao.org.uk

[3] STANDISH GROUP, *The Chaos Chronicles version 3.0,* The Standish Group, 2004

[4] CMMI PRODUCT TEAM, *CMMI for Development, Version 1.2*, CMMI-DEV v1.2, CMU/SEI-2006-TR-008, Technical Report, Software Engineering Institute, August 2006, URL: http://www.sei.cmu.edu/publications/documents/06.reports/06tr008.html

[5] ISO/IEC JTC1/SC7/WG10, *TR 15504-5, Software Process Assessment - Part 5: An Assessment Model and indicators guidance*, v.3.03, International Organization for Standardization, Genève, 1998

[6] THIONE L., *La Qualità delle Imprese Italiane. Stato Attuale, Problemi e Prospettive*, Monografia, Sincert, Dicembre 2005, URL: http://www.sincert.it/docs/405RelQimprese1205.pdf

[7] JAGANNATHAN S.R., BHATTACHARYA S. & MATAWIE K., *Value Based Quality Engineering*, TickIT International, No.1, 2005, pp.3-9

[8] SCHIFFAUEROVA, A. & THOMSON, V., *A review of research on cost of quality models and best practices*, International Journal of Quality and Reliability Management, Vol.23, No.4, 2006, pp. 647-669, Emerald Group Publishing Ltd, URL: http://www.mcgill.ca/files/mmm/CoQModels-BestPractices.pdf

[9] B.W. BOEHM, C. ABTS, A.W. BROWN, ET AL., *Software Cost Estimation with COCOMO II*, Prentice Hall, 2000, ISBN 0130266922

[10] PUTNAM, L.H. & MYERS W., *Five core metrics : the intelligence behind successful software management*, Dorset House Publishing, 2003, ISBN 0-932633-55-2

[11] JONES C., *Software Project Management Practices: Failure Versus Success*, *Crosstalk*, Vol.17, no.19, Oct. 2004, pp.5-9. URL: http://www.stsc.hill.af.mil/crosstalk/2004/10/0410Jones.pdf

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

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

[14] 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

[15] 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, URL: http://www.dpo.it/smef2005/filez/proceedings.pdf

[16] BUGLIONE L.& ABRAN A., *Improving Measurement Plans from multiple dimensions: Exercising with Balancing Multiple Dimensions - BMP*, 1st Workshop on "Methods for Learning Metrics", METRICS 2005, 11th IEEE International Software Metrics Symposium, 19-22 September 2005, Como (Italy), URL: http://metrics2005.di.uniba.it/learining-metrics-workshop/Buglione.pdf

[17] DUMKE R., ABRAN A. & BUGLIONE L., *Suggestions for Improving Measurement Plans: First results from a BMP application*, Proceedings of SMEF2006, 3rd Software Measurement European Forum, 10-12 May 2006, Rome (Italy), pp. 209-224, URL: http://www.dpo.it/smef2006/papers/b11.pdf

[18] BUGLIONE L., GENCEL C. & EFE P., *Suggestions for Improving Measurement Plans: A BMP Application in Turkey*, Proceedings of IWSM/METRIKON 2006, Potsdam (Germany), November 2-4, 2006, Shaker Verlag, ISBN 3-8322-5611-3, pp. 203-227

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

[20] NIST, *Baldrige National Quality Program: Criteria for Performance Excellence*, National Institute of Standards and Technology, 2007, URL: http://www.quality.nist.gov/

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

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

[23] 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