Toward the Improvement of a Measurement and Evaluation Strategy from a Comparative Study

Fernanda Papa

GIDIS_Web, Engineering School at UNLPam, General Pico, LP, Argentina
pmfer@ing.unlpam.edu.ar

Abstract. For evaluating web applications (WebApps) and other entities in a systematic way and further fostering robust data analysis among projects, measurement and evaluation (M&E) strategies are a valuable asset in any organization. However, the evaluation of the quality of the capabilities for an integrated M&E strategy –seen as a resource– has often been neglected. We regard a M&E strategy is integrated if three coexisting capabilities are supported, namely: i) a conceptual framework, ii) a well-established process specification, and iii) an explicit methodological support. Under this premise, we conducted a case study where GQM*Strategies (*Goal-Question-Metric*), and GOCAME (*Goal-Oriented Context-Aware Measurement and Evaluation*) strategies were evaluated. The results allowed us to understand the strengths and weaknesses of both strategies. From this understanding we have planned improvement actions and implemented some changes in GOCAME as well. This paper ultimately analyzes the achieved gains after recommended changes were performed.

Keywords: Measurement and evaluation strategy, Resource, Improvement.

1 Introduction

Regarding the Web Engineering discipline and its support process areas, quality measurement, evaluation and analysis for WebApps have been in the research forefront. Most of the published works and reported experiences are devoted to the evaluation of WebApps and its use –i.e. WebApp as a product, system or system in use–, and to lesser extent to the evaluation of used methods and tools. In a broad sense, evaluation is intended to assess a given information need and purpose e.g. “understand”, “improve”, etc. for different categories of entities such as product, system, system in use, among others such as service, process, resource or even a project. Oftentimes the evaluation is made specifying nonfunctional requirements by means of quality models. In a narrower sense, for a given category there are many sub-categories of entities of interest for evaluation. For example, for the resource category, we can identify more specific entity sub-categories such as “tool”, “strategy”, “software team”, etc.; and, in turn, for a “strategy” we can identify a “development strategy”, “testing strategy”, “M&E strategy”, among others. Furthermore, an entity is a concrete object that belongs to an entity category; for instance, in [11] we have recently evaluated two concrete entities, namely: GQM*Strategies [2, 3] and GOCAME [10] M&E strategies. In this way, the “M&E strategy” entity sub-category was considered as a resource for a software/web

line of production, which can be employed in quality assurance activities. Based on our review of existing literature, the quality evaluation for an integrated M&E strategy has often been neglected.

With the aim to systematically carry out M&E projects and programs, organizations should establish clearly a set of activities and methods to specify, collect, store, and use trustworthy measures and indicators' values. Moreover, in order to make the analysis and decision-making process more robust, it is necessary to ensure that metrics and indicators are repeatable and comparable among the organization's projects.

In [10], GOCAME –an integrated M&E strategy that allows developing programs with these characteristics– was proposed. This strategy includes the following capabilities: i) a *M&E conceptual framework*, which is modular, flexible and terminologically consistent. A well-established conceptual framework should be built on a robust terminological base as for example a glossary, taxonomy or ontology; an ontology explicitly and formally specifies the main agreed concepts, properties, relationships, and constraints for a given domain, as well as their grouping into components. This capability ensures terminological uniformity and consistency among the other capabilities; ii) a *M&E process*, which describes what to do, by specifying the main activities to be planned and executed, their inputs and outputs, roles, interdependencies, among other aspects. A well-established M&E process not only facilitates the understanding and communication among stakeholders but also ensures repeatability and reproducibility in the implementation of activities; and iii) *methods and tools* that enable to perform and automate the activities' descriptions. Methods are allocated in a flexible way to perform the specified activities, which in turn can be instantiated by tools.

Under the premise that a M&E strategy is integrated if the three above mentioned capabilities are to a great extent achieved simultaneously, we recently conducted a case study [11] where the two quoted strategies (GQM+Strategies and GOCAME) were evaluated considering the *Capability Quality* focus. From the analysis of results we obtained a detailed list of strengths and weaknesses for both strategies. Then we elaborated recommendations and a plan with improvement actions. Based on the improvement plan, we performed some changes for GOCAME –especially for those weaker attributes– and we conducted the re-evaluation. Hence, a comparison between both evaluations –before and after changes– gave us quantitative evidence about the level of gain met for GOCAME.

The contributions of this research were documented in [11], namely: i) understanding the quality of integrated M&E strategies, ii) designing the nonfunctional requirements focusing on the capability quality of this resource; iii) developing a case study to analyze and provide conclusions/recommendations based on identified strengths and weaknesses. We overview in the present work these contributions for a better comprehension, and we go a step further by: iv) establishing and implementing actions aimed at improving GOCAME; and v) re-evaluating it based on some implemented recommendations in order to gauge the improvement gain in quantitative form.

The remainder of this paper is organized as follows. Section 2 discusses the research motivation and related work; also, it provides an overview of the two selected strategies. Section 3 summarizes design and implementation issues for the comparative study. Section 4 shows, considering the three capabilities, the analysis and recommendations to improve GOCAME and also discusses the actual impact of implemented change actions. Finally, Section 5 draws the main conclusions and future work.

2 Motivation and Related Work

The motivation for this work was driven by the purpose of continuously improving the capability quality of GOCAME looking at the strengths of other well-established and integrated M&E strategy. With this in mind, we selected two concrete entities to be assessed: GOCAME and GQM⁺Strategies. The selection criteria used were basically: i) the M&E strategies are documented in the literature of public domain, i.e. there exists documentation –in English language– in digital libraries with recognized visibility such as IEEE Xplore, SpringerLink, ACM digital library, Scopus, etc.; ii) they have some impact in the academia or industry, and; iii) they have to great extent integrated the three capabilities simultaneously, i.e. the conceptual base/framework, process, and method/tool. With simultaneous compliance of these capabilities is meant that no input value to the conjunctive aggregation function (whose output is the value of the capability quality indicator) must be zero, as we see later on.

The systematic literature review showed there are many proposals published in the M&E area having some of the three capabilities, but they do not consider the integration of them simultaneously as a whole. Therefore, in the first selection process the following strategies were pre-selected: GOCAME, GQM [1], GQM⁺Strategies, FMESP (*Framework for the Modeling and Evaluation of Software Processes*) [8], and CQA-Meth (*Continuous Quality Assessment*) approach [12].

GOCAME is the target entity to be evaluated since it is an integrated M&E strategy and our ultimate purpose is its improvement as well. The second entity considered was GQM. It is a strategy with some level of integration, with abundant level of documentation available, and widely referenced and used both in industry and academia. However, in the literature reviewed we found that GQM⁺Strategies is an integrated M&E approach recently issued which includes GQM. So GQM⁺Strategies was the second entity to be evaluated.

Another contribution is the CQA-Meth approach, a flexible methodology that allows the quality assessment of any software model. This methodology and its tool are part of the CQA integrated environment that can be used by companies to perform quality assessments of their own or third-party products. CQA-Meth defines the processes necessary to carry out the evaluation of UML models, and facilitate communication between the client (sponsor of the evaluation) and the evaluation team. CQA-Meth was not selected as an entity to be assessed because it does not meet the three criteria simultaneously; in particular, it lacks an explicit conceptual framework from a terminological base. While CQA-Meth comes from the same research group who developed the FMESP approach that does have a conceptual framework with an ontological base, in the literature or references provided by authors in [12] the relationship among the three capabilities is not explicit at all.

In summary, the two entities that were selected as part of the comparative study are GOCAME and GQM⁺Strategies. Sub-sections 2.1 and 2.2 present an overview of both strategies to be assessed taking into account the three capabilities.

On the other hand, there are no other related works that document the evaluation-driven improvement of integrated M&E strategies as a resource.

2.1 GOCAME Overview

GOCAME is a multi-purpose M&E strategy which follows a goal-oriented and context-sensitive approach in defining M&E projects. It is based on the three above mentioned capabilities, which are summarized below.

GOCAME has its terminological base defined as an ontology. The metric and indicator ontology provides a domain model that defines all the concepts, properties and relationships which in turn helps to design the M&E activities. This way, a common understanding of data and metadata is shared among the organization's projects leading to more consistent analysis and results across projects. The C-INCAMI conceptual framework (*Contextual-Information Need, Concept model, Attribute, Metric and Indicator*) [10] emerged from the metric and indicator ontology.

C-INCAMI is structured in six components, namely: i) M&E project definition, allows specifying the management data for *M&E projects*; ii) Nonfunctional requirements specification, allows specifying the *Information Need* for a particular *purpose* and the *user viewpoint* related to an *Entity* and *quality focus*. The focus is represented by a *Concept Model* which includes *Concepts* (characteristics), *sub-concepts* and associated *Attributes*. Attributes are measurable properties of an *entity* under analysis; iii) Context specification, allows the description of the relevant *Context*, i.e. the situation of the entity to be assessed as regards the information need, through *Context properties*, which are attributes; iv) Measurement design and implementation, allows specifying *direct* and *indirect metrics* used in *Measurement* activities which produce *Measures*; v) Evaluation design and implementation, allows specifying the evaluation through *Indicators*, which interpret attributes and calculable concepts for a nonfunctional requirements tree. The *Indicator values* represent the degree of satisfaction achieved for a given information need. Two types of indicators are distinguished: *Elementary indicators* which evaluate lower-level requirements (attributes), and, *Partial/Global indicators*, which evaluate higher-level requirements, i.e. sub-characteristics and characteristics. The indicator *Scale* has *Decision criteria* in terms of acceptability levels; and vi) Analysis and recommendation specification, supports methods for data and information analysis in order to provide recommendations for improvement.

GOCAME has a well-defined M&E process [4], which is composed of six main processes: i) Define Nonfunctional Requirements; ii) Design the Measurement; iii) Design the Evaluation; iv) Implement the Measurement; v) Implement the Evaluation; and vi) Analyze and Recommend. These processes are broken down into activities, and sub-activities that are specified in SPEM language.

Lastly, GOCAME is supported by a methodology viz. WebQEM [9] and its C-INCAMI_Tool [10]. The methodology provides the 'how' to implement the requirements, measurement, evaluation, analysis and recommendation processes. It comprises a set of methods, techniques and tools to carry out the process activities accordingly.

2.2 GQM⁺Strategies Overview

GQM⁺Strategies is an approach built on GQM, which allows planning and implementing goal-oriented measurement programs. Taking into account GQM, GQM⁺Strategies adds a mechanism for explicitly linking software measurement goals to higher-level

goals for the organization, as well as goals and strategies (tactics) to all business levels. This linkage helps to justify the efforts made in the software measurement and, in turn, the data collected contribute to make high-level decisions.

GQM⁺Strategies has its terminological base defined as a glossary [2, 3], reusing totally the GQM terms. In addition, terms are part of two primary components, namely: GQM⁺Strategies Element and GQM Graph. The former component includes a *single goal* and *derived strategies*, as well as all *context information* and *assumptions*, which explain how those goals and strategies are linked. On the other hand, the latter component reflects a single *GQM goal*, the corresponding set of *questions* and *metrics*, and an *interpretation model* that specifies how data items are to be combined and what criteria for determining the goal's success are.

Regarding the explicitness of the process, in [3] authors define basically two processes, which may be performed in parallel: one describing the tasks needed to define goals and related strategies, and the other describing the tasks needed to measure already defined goals and strategies. The first process involves activities such as: i) Elicit General Context and Assumptions; ii) Define Top-Level Goals; iii) Make Strategy Decisions; and iv) Define Goals. This process is iterated through all organizational levels. The other process involves activities such as: i) Define GQM goals for each selected GQM⁺Strategies goal at the appropriate level. The GQM goal template [3], which includes the object, purpose, quality aspect, viewpoint, and context terms, is used to formalize the measurement goal; ii) Specify the GQM graph for evaluating the achievement of the goal, i.e. GQM questions, metrics and interpretation models are defined; and iii) Identify relationships between the interpretation models on this level and the ones for the level above, if existing. Finally, it must be implemented the measurement and interpreted its results.

Lastly, in the reviewed literature no methodology is mentioned for GQM⁺Strategies; however, GQM explicitly defines one, covering several phases such as planning, definition, data collection and interpretation. Features of a tool that is configurable for each organizational measurement program is described in [13].

3 Evaluating Integrated M&E Strategies

In this section we summarize design and implementation issues for evaluating the two quoted strategies. GOCAME strategy was used in turn for conducting the evaluation itself. This evaluation allows understanding and comparing integrated strategies. Hereafter, using the same evaluation requirements and criteria, the re-evaluation of GOCAME allows gauging the improvement gain after recommended changes were made. Note that for didactic reasons activities are presented below in a bit different order that they were actually enacted. In fact, we performed first the design (i.e. non-functional requirements, measurement and evaluation), then the implementation of measurement and evaluation, and finally, the analysis and recommendation activities, as introduced in section 2.1.

Define Nonfunctional Requirements is the first activity. The purpose of the information need –for the first evaluation– is “*understand and compare*” from the “*quality assurance leader*” user viewpoint. The entity category is an “*integrated M&E strategy*” whose super-category is a “*resource*”, recalling that *GOCAME* and

GQM⁺Strategies are the concrete entities to be assessed. The focus of the evaluation is on their “*Capability Quality*”, which is defined as *degree to which a resource is suitable and appropriate for supporting and performing the actions when used under specified conditions*. Since in the related literature there is no ISO or *de facto* standard that specifies the capability quality model, we had to define our own concept model, as shown in the first column in Table 2.

The sub-characteristics associated to *Capability Quality* (coded 1) represent the three required capabilities of an integrated strategy, namely: *1.1 Process Capability Quality*, *1.2 Conceptual-Framework Capability Quality*, and *1.3 Methodology Capability Quality*. The sub-characteristics and attributes in the requirements tree –shown in the first column in Table 2– are all defined. For example, *1.1* represents *the degree to which a process is suitable and appropriate for supporting and performing the actions defined in it*; while the *1.1.1.5 Role-to-Activity Allocation Availability* attribute is defined as *the explicit indication of one or more role assigned to an enunciated activity* and its objective is *to find out the extent to which the activities have one or more roles allocated*.

The result of this activity is a nonfunctional requirements specification artifact, which has 48 definitions including 17 sub-characteristics and 31 attributes.

The second activity is Design the Measurement. For each attribute in the requirements tree a metric was assigned; e.g. the indirect metric *Role-to-Activity Allocation Availability Degree (RAAAD)* quantifies the *1.1.1.5* attribute. The metric objective is *to quantify how many process activities have an allocated role with regard to the total amount of enunciated activities*; also it has the following formula specification:

$$RAAAD = \begin{cases} \text{If TEA} = 0 \rightarrow 0 \\ \text{If TEA} > 0 \rightarrow (\#AAR / TEA) \times 100 \end{cases} \quad (1)$$

Where, *TEA* stands for *Total number of Enunciated Activities*; and *#AAR* for *Number of Activities with Allocated Role*. The metric scale is *numerical* and its unit is *percentage (%)*.

The metric specification document consists of 54 metrics (direct and indirect) to quantify the 31 attributes of the requirements tree.

Table 1. Measured values of direct metrics used for calculating the RAAAD indirect metric

	GOCAME	GQM ⁺ Strategies
<i>Total number of Enunciated Activities (TEA)</i>	47	101
<i>Number of Activities with Allocated Role (#AAR)</i>	0	18

Then, the next activity is Implement the Measurement. Data collection was performed from Sept. to Dec., 2010 based on published and accessible material as commented at the beginning of Section 2. We used the most relevant documents disregarding those that were not coauthored by at least one member of the authors of the original research. Moreover, we gave greater priority to the most current documents when they represented a contribution with regard to previous ones.

To our example, the indirect metric RAAAD result was 0% for GOCAME and 17.82% for GQM⁺Strategies. These values are calculated from measurement values in Table 1 and Eq. 1.

Table 2. Requirements tree -attributes are in italic (1st column). In the rest of columns, legend (1) shows the indicator values in % for GOCAME, and legend (2) for GQM⁺Strategies

	(1)	(2)
1.Capability Quality (for an integrated M&E strategy)	66.48	45.89
1.1.Process Capability Quality	58.88	54.34
1.1.1.Activities Suitability	46.67	38.37
1.1.1.1.Activities Description Availability	31.91	24.75
1.1.1.2. Activities Description Completeness	15.47	14.40
1.1.1.3.Process Breakdown Structure Granularity	70	70
1.1.1.4. Activities Description Formality	100	61.39
1.1.1.5.Role-to-Activity Allocation Availability	0	17.82
1.1.2.Artifacts Suitability	3	31.58
1.1.2.1.Artifacts Description Availability	0	27.59
1.1.2.2.Artifacts Description Completeness	0	26.96
1.1.2.3.Artifacts Breakdown Structure Granularity	30	70
1.1.3.Process Modeling Suitability	83.56	70.70
1.1.3.1.Functional View Suitability	88	76.42
1.1.3.1.1.Functional View Availability	100	100
1.1.3.1.2.Functional View Completeness	100	61.39
1.1.3.1.3.Functional View Granularity	70	70
1.1.3.2.Informational View Suitability	82.13	74.97
1.1.3.2.1.Informational View Availability	100	100
1.1.3.2.2.Informational View Completeness	90.32	72.41
1.1.3.2.3.Informational View Granularity	30	30
1.1.3.3.Behavioral View Suitability	88	60.42
1.1.3.3.1.Behavioral View Availability	100	100
1.1.3.3.2.Behavioral View Completeness	100	61.39
1.1.3.3.3.Behavioral View Granularity	70	30
1.1.3.4.Organizational View Suitability	0	63.78
1.1.3.4.1.Organizational View Availability	0	100
1.1.3.4.2.Organizational View Completeness	0	44.44
1.1.3.4.3.Organizational View Granularity	0	30
1.1.4.Process Compliance	85.79	71.11
1.1.4.1.Process-to-Concept-Base Terminological Compliance	94.74	88.89
1.1.4.2.M&E Process Standards Compliance	50	0
1.2.Conceptual-Framework Capability Quality	75.09	35.82
1.2.1.Conceptual Framework Suitability	75	25
1.2.1.1.Conceptual Framework Modularity	50	0
1.2.1.2.Conceptual Framework Modeling Formality	100	50
1.2.2.Conceptual Base Suitability	68.53	18.53
1.2.2.1.Conceptual Base Completeness	21.33	1.33
1.2.2.2.Conceptual Base Structure Richness	100	30
1.2.3.Conceptual Framework Compliance	84.31	81.82
1.2.3.1.Framework-to-C-Base Terminological Compliance	84.31	81.82
1.3. Methodology Capability Quality	77.43	57.35
1.3.1.Methodology Suitability	83.19	51.88
1.3.1.1.Methodology Availability	100	100
1.3.1.2.Method-to-Activity Completeness	82.98	29.70
1.3.1.3.Methodology Automated Support Availability	50	0
1.3.2.Methodology Compliance	73.68	61.11
1.3.2.1.Methodology-to-C-Base Terminological Compliance	73.68	61.11

The next activity is Design the Evaluation –both elementary and global. For each attribute in the requirements tree, which represents an elementary nonfunctional requirement, an elementary indicator that interprets it was defined. For example, *Performance of Role-to-Activity Allocation Availability (P_RAAA)* is the indicator name for the 1.1.1.5 attribute, and its elementary model is specified as a direct mapping.

Regarding the global evaluation, we selected the LSP (*Logic Scoring of Preference*) model [6] for calculating the requirements tree partial/global indicators. LSP is a weighted multi-criteria aggregation model, which has operators for modeling simultaneity (C –conjunctive- operators) and replaceability (D –disjunctive- operators) relationships between attributes and (sub-)concepts. Thus, the C+ weak conjunction operator lets modeling the simultaneity criterion among the 1.1, 1.2 and 1.3 sub-concepts, yielding zero if one input was zero. It is worth remarking that this was one of the criteria for pre-selecting integrated strategies as indicated in the introduction of Section 2.

Regarding decision criteria we have used three acceptability levels for interpreting indicator values in percentage scale. A value between [0-50] represents an *unsatisfactory level* and means that *change actions must be taken with high priority*. A value between [50-75] represents a *marginal level* and means that *improvement actions should be taken*. While a value between [75-100] corresponds to a *satisfactory level*.

So the next activity is Implement the Evaluation. In this process we calculated each elementary indicator taking into account the measured value and its elementary model as well. For example, the elementary indicator value to *Role-to-Activity Allocation Availability* attribute is 0% for GOCAME and 17.82% for GQM+Strategies. Finally, once calculated all elementary indicators, by enacting the LSP aggregation model, all partial/global indicators values are yielded. Table 2 shows in the 2nd and 3rd columns the resulting elementary, partial and global indicator values for both strategies.

The following section discusses in detail the results of the last activity named Analyze and Recommend. Once many of given recommendations were implemented in GOCAME during 2011, we carried out its first re-evaluation. For the re-evaluation we used the same M&E requirements with the aim yielded values were repeatable and comparable among studies.

4 Analysis, Recommendation and Improvement

The output of the Analyze and Recommend process is a conclusion and recommendation report comprised of tables, comparison charts, among other mechanisms. This report summarizes for instance strengths and weaknesses, and also recommends change actions to facilitate further improvements.

Below we address the analysis mainly for the GOCAME strategy considering those attributes that have to be improved in order to increase their satisfaction levels. Additionally, improvement recommendations arise not only from GOCAME indicators with weaker performance but also from GQM+Strategies indicators with stronger score. Although the global satisfaction level achieved is lower for GQM+Strategies, there are some well-scored elementary indicators that can be taken into account when planning improvements for GOCAME.

Based on Table 2, we can observe for the *Capability Quality* that GOCAME met a

marginal satisfaction level for its global indicator (66.48%, which means that actions for improvement should be taken), while GQM⁺Strategies achieved an unsatisfactory acceptability level (45.89%, which means that must change).

Considering the simultaneity criterion for the three sub-characteristics of the *Capability Quality* focus concept, we observe for GOCAME that the *1.1 Process Capability Quality* falls in the marginal level (58.88); the *1.2 Conceptual-Framework Capability Quality* gets a satisfactory level (75.09); and for *1.3 Methodology Capability Quality* the indicator value is also satisfactory (77.43). As general conclusion, it emerges that GOCAME should strengthen its process. So to improve it we have to analyze and plan which change actions should be prioritized. Next four sub-sections discuss recommendations for each capability and actions taken (if any).

4.1 Process Capability Quality: Analysis, Recommendation and Improvement

The process capability quality sub-characteristics in the unsatisfactory level are namely *1.1.1 Activities Suitability* (46.67%), and *1.1.2 Artifacts Suitability* (3%). Let's start analyzing *1.1.1* and its three attributes which fall in the unsatisfactory level as shown in Table 2.

The attributes *1.1.1.1 Activities Description Availability* (31.91) and *1.1.1.2 Activities Description Completeness* (15.47) are quantified by indirect metrics composed of direct metrics (see Table 3) that deal with the total number of enunciated activities and the number of described activities (either minimally, partially, or completely described ones). We considered an activity is enunciated when it belongs to the process under analysis and has a unique name, or label; also, an activity is completely described when it is an enunciated activity and has explicit and textual specifications of *objective*, *description*, *pre-condition*, *post-condition*, *input* and *output* metadata. The design of these metrics was thoroughly specified in [11]. The metric specification and measured values for its direct metrics (e.g. as shown Table 3 for GOCAME), help us to understand the reasons why the elementary indicator ranks unsatisfactorily, and allow us to make recommendations for planning improvement actions.

Table 3. Measured values of direct metrics used for calculating indirect metrics for 1.1.1.1 and 1.1.1.2 attributes: In 1st column are direct metrics' names, 2nd column shows values gathered in the first evaluations (2010), and 3rd shows measures collected in the re-evaluation (2011).

	2010	2011
<i>Total number of Enunciated Activities (TEA)</i>	47	49
<i>Number of Minimally Described Activities (#MDA)</i>	5	45
<i>Number of Partially Described Activities (#PDA)</i>	10	45
<i>Number of Completely Described Activities (#CDA)</i>	0	31

The R1 recommendation in Table 4 shows two suggested actions to improve both attributes. Once change actions from R1 were planned and performed, the indicator value for *1.1.1.1* attribute will get a maximum level –if it were totally implemented– because its metric is a function of the enunciated activities and minimally described activities i.e. it has *objective* and *description* fields. The same applies to the *1.1.1.2* attribute once all abovementioned template fields for all the activities are filled.

It is worth mentioning that these improvement actions have already been done in

GOCAME. As shown in Table 5, the *1.1.1.1* indicator value increases from 31.91 to 91.84% (representing a positive impact, which in terms of difference is by 59.92), while the *1.1.1.2* indicator value goes from 15.47 to 81.45 (which represents a positive difference of 65.97).

Table 4. Summary of recommendations for the GOCAME process: 1st column represents the recommendation code; the 2nd shows its description; the 3rd indicates the attribute to be improved and the 4th the change priority, i.e. **MT** means action must be taken and **ST** action should be taken. Last column shows if the change was made (R), it is pending (P), and not required (NR).

ID	Process Capability Recommendation	Attribute	Priority	Action
R1	1. Specify a template with the metadata required for a completely described activity. 2. Describe each enunciated activity filling the template.	1.1.1.1 1.1.1.2	MT	R
R2	1. Define used roles/responsibilities in the M&E process. 2. Assign one or more roles to each enunciated activity.	1.1.1.5	MT	R
R3	1. Analyze each process activity if can be divided in atomic activities.	1.1.1.3	ST	P
R4	1. Specify a template with the metadata required for a completely described artifact. 2. Describe each enunciated artifact filling the template.	1.1.2.1 1.1.2.2	MT	P
R5	1. Analyze each enunciated artifact to determine whether it can be divided in atomic sub-artifacts and be manageable from the configuration baseline standpoint.	1.1.2.3	MT	R
R6	1. Model processes from the organizational point of view considering all enunciated activities and roles are included.	1.1.3.4.1 1.1.3.4.2	MT	R
R7	1. Determine if models of the organizational view can be subdivided hierarchically into sub-views.	1.1.3.4.3	MT	R
R8	1. Determine if models of the informational view can be subdivided hierarchically into sub-views.	1.1.3.2.3	MT	R

As result of changes, GOCAME has currently two more enunciated activities (49) than before, being now 45 activities minimally described and 31 completely described activities (as shown in Table 3). The new data collection was made on the document published in [5].

On the other hand, the indirect metric that quantifies the attribute *1.1.1.5 Role-to-Activity Allocation Availability* uses the total number of enunciated activities and the number of activities with assigned roles as per Eq. 1. Considering that the indicator value was 0%, therefore the R2 recommendation emerged (see Table 4). In this case, GQM*Strategies helped us accomplishing this recommendation, since it has a set of enunciated roles –though seldom with their responsibilities defined. In order to perform R2, 13 roles and their responsibilities were defined in GOCAME. These roles were assigned to higher-level activities. As result of changes GOCAME has now 16 activities with allocated roles, however, the level of satisfaction met is still unsatisfactory, i.e. it upgraded by 32.56% as shown in Table 5.

Additionally, closer to the upper threshold of the marginal acceptability level is the indicator value for the *1.1.1.3 Process Breakdown Structure Granularity* attribute, which scored 70%. So we could work on it to improve it. The 100% of its satisfaction

can be achieved when the whole process has fine-grained activity decomposition. Thus, in Table 4, R3 is the recommendation to reach that score. In GOCAME two activities were added, however, these do not represent an increase in the process granularity. So this improvement change was not tackled yet.

Table 5. Impact of changes for the process capability in GOCAME: 1st column shows process capability attributes (enhanced ones are shaded); 2nd and 3rd columns represent indicator values (in %) before and after improvements were made; last column indicates whether the change was positive (↑), negative (↓), or without variation (↔) and relative change values.

	2010	2011	Diff.
1.Capability Quality (for M&E strategy)	66.48	71.44	↑ 4.96
1.1.Process Capability Quality	58.88	73.26	↑ 14.38
1.1.1.Activities Suitability	46.67	80.29	↑ 33.62
1.1.1.1.Activities Description Availability	31.91	91.84	↑ 59.92
1.1.1.2. Activities Description Completeness	15.47	81.45	↑ 65.97
1.1.1.3.Process Breakdown Structure Granularity	70	70	↔ 0
1.1.1.4. Activities Description Formality	100	95.92	↓ -4.08
1.1.1.5.Role-to-Activity Allocation Availability	0	32.65	↑ 32.65
1.1.2.Artifacts Suitability	3	7	↑ 4
1.1.2.1.Artifacts Description Availability	0	0	↔ 0
1.1.2.2.Artifacts Description Completeness	0	0	↔ 0
1.1.2.3.Artifacts Breakdown Structure Granularity	30	70	↑ 40
1.1.3.Process Modeling Suitability	83.56	89.26	↑ 5.70
1.1.3.1.Functional View Suitability	88	86.78	↓ -1.22
1.1.3.1.1.Functional View Availability	100	100	↔ 0
1.1.3.1.2.Functional View Completeness	100	95.92	↓ -4.08
1.1.3.1.3.Functional View Granularity	70	70	↔ 0
1.1.3.2.Informational View Suitability	82.13	91.65	↑ 9.52
1.1.3.2.1.Informational View Availability	100	100	↔ 0
1.1.3.2.2.Informational View Completeness	90.32	94.12	↑ 3.80
1.1.3.2.3.Informational View Granularity	30	70	↑ 40
1.1.3.3.Behavioral View Suitability	88	86.78	↓ -1.22
1.1.3.3.1.Behavioral View Availability	100	100	↔ 0
1.1.3.3.2.Behavioral View Completeness	100	95.92	↓ -4.08
1.1.3.3.3.Behavioral View Granularity	70	70	↔ 0
1.1.3.4.Organizational View Suitability	0	94	↑ 94
1.1.3.4.1.Organizational View Availability	0	100	↑ 100
1.1.3.4.2.Organizational View Completeness	0	100	↑ 100
1.1.3.4.3.Organizational View Granularity	0	70	↑ 70
1.1.4.Process Compliance	85.79	85.79	↔ 0
1.1.4.1.Process-to-Concept-Base Terminological Compliance	94.74	94.74	↔ 0
1.1.4.2.M&E Process Standards Compliance	50	50	↔ 0

As abovementioned, the *1.1.2 Artifacts Suitability* sub-characteristic met an unsatisfactory level (3%) as well. So, improvement actions for its attributes are analyzed and recommended. Particularly, for *1.1.2.1 Artifacts Description Availability* and *1.1.2.2 Artifacts Description Completeness* attributes, their indicator values were 0%. They are quantified by metrics which deal with enunciated artifacts and described artifacts. An artifact is completely described when it is enunciated and has explicitly specified the *objective*, *description*, and the *activity name* which create/modify it. The R4 recommendation (Table 4) gives the hint for planning the change, but the improvement ac-

tions at this moment are still pending of completion. However, by following R5 we have improved the *1.1.2.3 Artifacts Breakdown Structure Granularity* attribute raising the value from 30 to 70 (Table 5). In this case, we also considered as reference the GQM*Strategies elementary indicator which ranked 70%, as shown in Table 2.

Finally, the *1.1.3 Process Modeling Suitability* sub-characteristic ranked satisfactorily (83.56%). Table 2 also shows that the functional (*1.1.3.1*), informational (*1.1.3.2*) and behavioral process (*1.1.3.3*) views suitability are in the satisfactory level. However, the *Organizational View Suitability* (*1.1.3.4*) sub-characteristic is totally missing (0%). Table 4 shows the R6 recommendation whose improvement action embraces to *1.1.3.4.1 Organizational View Availability* and *1.1.3.4.2 Organizational View Completeness* attributes; also R7 is the recommendation for the *1.1.3.4.3 Organizational View Granularity* attribute. Note that for planning the improvement for these three attributes, the GQM*Strategies strengths were also considered. In [5], all enunciated roles assigned to main GOCAME activities, and a hierarchy of roles are documented. Consequently, the positive impact of improvement was very important for the *1.1.3.4 Organizational View Suitability* sub-characteristic going from 0 to 94% as shown in Table 5.

Lastly, Table 4 shows recommendations for the *Process Capability Quality* characteristic; however, some recommendation actions are still in progress. Note that also there were negative impacts, which it will be addressed in sub-section 4.4.

4.2 Conceptual-Framework Capability Quality: Analysis and Recommendation

For the *1.2 Conceptual-Framework Capability Quality* sub-characteristic, its indicator value scored satisfactorily (75.09% in Table 2). Looking in turn at its sub-concepts only *1.2.2 Conceptual Base Suitability* fell in a marginal acceptability level (68.53%), while the others (*1.2.1* and *1.2.3*) reached the satisfactory level. Therefore, the most important recommendation for improvement arises from the analysis of the *1.2.2.1 Conceptual Base Completeness* attribute, which had low performance (21.33%). The recommendation statement emerges from the analysis of the metric specification, i.e., the indirect metric's calculation method that quantifies *1.2.2.1* is specified as a ratio of the level of correspondence between terms defined in the conceptual base and standards, and the total number of terms defined in standards. It is important to remark that the terms added to the terminological base –regardless whether be structured as glossary or ontology- should also be included in the conceptual framework, in order to not affect the *1.2.3.1 Conceptual Framework-to-Conceptual Base Terminological Compliance* indicator performance.

Additionally, in order to increase the score of the *1.2 capability quality* evaluators can recommend improving the *1.2.1.1 Conceptual Framework Modularity* attribute, which scored 50%. This attribute is quantified by a direct metric that measures the degree to which the conceptual framework is divided into different modules or components. It is expected a conceptual framework includes components to manage projects, nonfunctional requirements, context, measurement, evaluation, and analysis/recommendations. Also several of them can be divided into design and implementation as well. GOCAME has considered this concern taking into account 6 mentioned modules. Nevertheless, an opportunity for improvement is that some modules can be

split into design and implementation sub-modules.

Note that recommendations for this capability have not been implemented yet, so that the satisfaction level has not changed. Recall that first we have prioritized improvement actions for the *Process Capability Quality* (1.1), which was the weakest capability (58.88%) out of three.

4.3 Methodology Capability Quality: Analysis and Recommendation

For the *1.3 Methodology Capability Quality* sub-characteristic its indicator value scored also satisfactorily, i.e. 77.43% (Table 2). In addition, we can observe that the *1.3.1 Methodology Suitability* sub-characteristic met 83.19% even though the satisfaction level for its *1.3.1.3 Methodology Automated Support Availability* attribute scored marginally (50%). So, for improving this attribute the recommendation should considering two aspects, namely: 1) For the identified parts of the methodology without tool support, analyze for each part if automated support is feasible; 2) If feasible, implement the tool.

Additionally, in order to increase the score of the *1.3* capability quality evaluators can recommend improving the *1.3.2.1 Methodology-to-Conceptual Base Terminological Compliance* attribute, which scored in the upper limit of the marginal acceptability level, i.e. 73.68%. Hence, the following recommendation should be taken into account: Methodology descriptions should be documented adhering as much as possible to terms and definitions of the GOCAME terminological base.

Note that recommendations for this capability have not been implemented yet, so that the satisfaction level has not changed. In the next sub-section, we discuss the impact of performed changes after re-evaluation, which are summarized in Table 5.

4.4 Discussing Positive and Negative Impacts of Performed Changes

As discussed in sub-section 4.1, so far many improvement recommendations belonging to *Process Capability Quality* have been implemented either partially or totally. Therefore, change actions caused the process satisfaction level to increase by 14.38 points (considering the difference between the 2011 and 2010 results), while the global satisfaction level (i.e. the GOCAME *Capability Quality*) has increased by 4.96 points (see Table 5).

As expected, the performed change actions had usually a positive impact but in some attributes caused an undesired effect or negative impact. This was the case for attributes such as *1.1.1.4 Activities Description Formality*, *1.1.3.1.2* and *1.1.3.3.2* (which refer to the completeness of two process modeling views). The negative impact however was negligible due to our adding of two new activities for the M&E process (see TEA values in Table 3), which were not re-considered in the modeling of functional and behavioral views.

On the other hand, indicator values for *1.1.1.5 Role-to-Activity Allocation Availability* and *1.1.3.4.3 Organizational View Granularity* attributes have increased by 32.65 and 70 points respectively. However, the performed improvement was not enough to achieve the satisfactory acceptability level. More re-work is still necessary.

Regarding *Process Modeling Suitability* sub-characteristics the highest improvement gain was in *1.1.3.4* which increased 94 points. Besides, a small improvement gain

of 9.52 points for *1.1.3.2 Informational View Suitability* has been met. This was caused mainly by improving the *1.1.3.2.3 Informational View Granularity* attribute.

Finally, it is worth mentioning that improvement changes made so far led to increases in the GOCAME *Capability Quality* indicator value but it has still not met the satisfactory acceptability level. To reach the [75-100] level, it remains to implement many of the improvement actions proposed by the original (2010) study.

However, performing recommended changes sometimes it is not an easy and fast job. For example, we have developed during 2002-2003 an ontology for metrics and indicators, which is the terminological base used in the C-INCAMI framework. This framework has 6 components –as commented in sub-section 2.1- in which this ontology was used totally in 4 modules. However, the Analysis and recommendation specification component has no ontological support until now, even though the process is well defined. Thus, for planning an improvement action for the *1.2.2.1 Conceptual Base Completeness* attribute, an ontology for the analysis and recommendation domain should be engineered. As the reader can surmise, this could take considerable effort and calendar time.

5 Conclusion and Future Work

To summarize, we would like to highlight the particular contributions of this research commented in the Introduction Section. This paper elaborates on the progress achieved using as foundation the comparative study made in 2010 on the *Capability Quality* evaluation of two concrete M&E strategies [11]. By using the same nonfunctional requirements, measurement and evaluation design, and identified strengths and weaknesses, we have gone a step further by: i) planning and implementing change actions aimed at improving GOCAME; and ii) re-evaluating GOCAME based on implemented recommendations in order to gauge the improvement gain in quantitative form.

It is important to remark that from the very beginning of this research our ultimate objective was the improvement of GOCAME, which is a strategy that can be used as an evaluation resource in different stages of a web engineering production line.

Regarding the former contribution, we have planned and implemented improvement recommendations not only considering GOCAME indicators with weaker performance but also GQM⁺Strategies indicators with stronger performance. We have prioritized change actions mainly for the GOCAME *Process Capability Quality*, which had lower performance.

Regarding the latter contribution, we have re-evaluated GOCAME –using the same requirements and M&E design– after the actual change actions were carried out. Particularly, in sub-section 4.4, we have discussed not only expected positive impacts of changes but also their –negligible– undesired effects. So far, implemented improvement changes in *Process Capability Quality* enhanced its partial indicator from 58.88 to 73.26%. Even if this represents a moderate improvement in the process quality, the calculated *Capability Quality* global indicator increased from 66.48 to 71.44%. This is because many of the recommended changes in the other two capabilities are in progress, so they did factor in upgrading the global indicator value. We have also discussed that performing recommended changes sometimes it is not an easy and fast task.

Lastly, a future line of research is broadening the GOCAME scope in order to give

support to M&E information needs (goals) at different organizational levels, following to some extent to GQM Strategies which does support them.

Acknowledgments. Thanks to the support given from the Science and Technology Agency, Argentina, in the PAE-PICT 2188 project, and to the 09-F047 project at UNLPam, Argentina.

References

1. Basili, V.R., Caldiera, G., Rombach, H.D.: The goal question metric approach. In: Encyclopedia of Software Engineering, vol. 1, pp. 528–532 (1994)
2. Basili, V.R., Heidrich, J., Lindvall, M., Münch, J., Regardie, M., Rombach, H.D., Seaman, C.B., Trendowicz, A.: GQM strategies@: A comprehensive methodology for aligning business strategies with software, pp. 253–266 (2007)
3. Basili, V.R., Lindvall, M., Regardie, M., Seaman, C., Heidrich, J., Jurgen, M., Rombach, D., Trendowicz, A.: Linking Software Development and Business Strategy through Measurement. IEEE Computer 43(4), 57–65 (2010)
4. Becker, P., Molina, H., Olsina, L.: Measurement and Evaluation as quality driver. Journal ISI (Ingénierie des Systèmes d'Information), Special Issue: Quality of Information Systems 15(6), 33–62 (2010)
5. Becker, P., Lew, P., Olsina, L.: Specifying Process Views for a Measurement, Evaluation and Improvement Strategy. Advances in Software Engineering Journal 2012, 28 (2012), Open Access at <http://www.hindawi.com/journals/ase/contents/>, doi:10.1155/2012/949746
6. Dujmovic, J.J.: A Method for Evaluation and Selection of Complex Hardware and Software Systems. In: 22nd Int'l Conference for the Resource Management and Performance Evaluation of Enterprise CS. CMG 96 Proceedings, vol. 1, pp. 368–378 (1996)
7. Lew, P., Olsina, L., Becker, P., Zhang, L.: An Integrated Strategy to Understand and Manage Quality in Use for Web Applications. Requirements Engineering Journal 16(3), 1–32 (2011), Loucopoulos P., Mylopoulos, J. (Eds.), doi:10.1007/s00766-011-0128-x
8. Mora, B., García, F., Ruiz, F., Piattini, M., Boronat, A., Gómez, A., Carsí, J.Á., Ramos, I.: Software generic measurement framework based on MDA. IEEE Latin America Transactions 6(4), 363–370 (2008)
9. Olsina, L., Rossi, G.: Measuring Web Application Quality with WebQEM. IEEE Multimedia Magazine 9(4), 20–29 (2002)
10. Olsina, L., Papa, F., Molina, H.: How to Measure and Evaluate Web Applications in a Consistent Way. In: Rossi, G., Pastor, O., Schwabe, D., Olsina, L. (eds.) Springer Book: Web Engineering: Modeling and Implementing Web Applications, ch. 13, pp. 385–420 (2008)
11. Olsina, L., Papa, M.F., Becker, P.: Assessing Integrated Measurement and Evaluation Strategies: A Case Study. In: IEEE Xplore, 7th Central Eastern European Software Engineering Conference, Moscow, Russia, pp. 1–10 (2011) ISSN: 978-1-4673-0844-1/11
12. Rodríguez, M., Genero, M., Torre, D., Blasco, B., Piattini, M.: A Methodology for Continuous Quality Assessment of Software Artifacts. In: 10th International Conference on Quality Software (QSIC 2010), pp. 254–261 (2010)
13. Solingen, R.V., Berghout, E.: The Goal/Question/Metric Method: a practical guide for quality improvement of software development. McGraw-Hill Co. (1999)