

Family of Evaluation Strategies: A Practical Case for Comparing and Adopting Strengths

Guido Tebes, Denis Peppino, Pablo Becker, Maria Fernanda Papa, Maria Belen Rivera and Luis Olsina

GIDIS_Web, Facultad de Ingeniería, UNLPam, General Pico, LP, Argentina
guido_tebes@hotmail.com, denispeppino92@gmail.com,
{beckerp, pmfer, riveramb, olsinal}@ing.unlpam.edu.ar

Abstract

Companies commonly establish and pursue business goals. A goal states a purpose. In the Software Engineering literature, different evaluation purposes such as to understand, monitor, improve, control, compare and select are mentioned. Considering that these purposes share distinctive and common aspects, in the present work, we include a categorization that helps a better understanding of them. On the other hand, to reach the purpose of a goal, a suitable strategy should be chosen. A strategy describes a particular course of action by means of process and method specifications. We have envisioned different strategies for different evaluation goal purposes. Therefore, in this paper we also present some strategies, which are part of a family of strategies driven by measurement and evaluation activities. Specifically, we document evaluation strategies for the *monitoring*, *improving*, and *comparing and adopting* purposes. In addition, we illustrate the *comparing and adopting strategy* applied to four social network mobile apps.

Keywords: Goal, Evaluation purpose, Compare and adopt, Strategy

1. Introduction

In any mature organization, the continuous quality improvement of its resources, processes, products and services should be ensured. Sometimes, software organizations need to develop new

applications or to improve the existing ones, so it would be desirable to take into account, as a reference, others applications to adopt their strengths. For this end, it is not only necessary to have Measurement and Evaluation (ME) activities which allow to understand the actual or estimated situation of an entity through strengths and weaknesses analysis, but also to have activities that allow to perform changes oriented to the improvement (MEC will stand for ME and Change).

An engineering way to organize the ME/MEC activities and to manage resources is by means of project formulation. Considering the project term definitions given in [6, 16], both agree on that it is necessary to determine a project plan. During planning and resource allocation, the goal and its purpose are taken into account. In addition, the suitable strategy should be selected aimed at achieving the goal. The selected strategy should describe the course of action for the goal and the particular purpose, by means of the specification of activities and methods. Consequently, software organizations that consider ME or MEC projects in a systematic way can ensure process repeatability and consistency of results.

Aimed at illustrating the goal, purpose and strategy concepts, Fig. 1 depicts some relations between them, which will serve as the basis for understanding this work. An organizational goal can be both business and information need goal. The latter supports the former, and consequently has the *analyze* purpose. Additionally, an information need goal could require a ME information need goal, which involve ME activities. Furthermore, typical evaluation purposes for business goals can be to understand, monitor, control, improve, adopt, among others. However, we observe that in the state-of-the-art literature there is some lack of consensus in their descriptions and/or definitions.

On the other side, Fig. 1 shows that a ME/MEC strategy considers an evaluation purpose and one or more quality views. For instance, the *monitoring* strategy considers the *monitor* purpose and one

Citation: G. Tebes, D. Peppino, P. Becker, M. F. Papa, M.B. Rivera and L. Olsina. "Family of Evaluation Strategies: A Practical Case for Comparing and Adopting Strengths", Journal of Computer Science & Technology, vol. 18, no. 1, pp. 48–60, 2018.

DOI: 10.24215/16666038.18.e06

Received: February 6, 2018 **Revised:** April 14, 2018
Accepted: April 18, 2018.

Copyright: This article is distributed under the terms of the Creative Commons License CC-BY-NC.

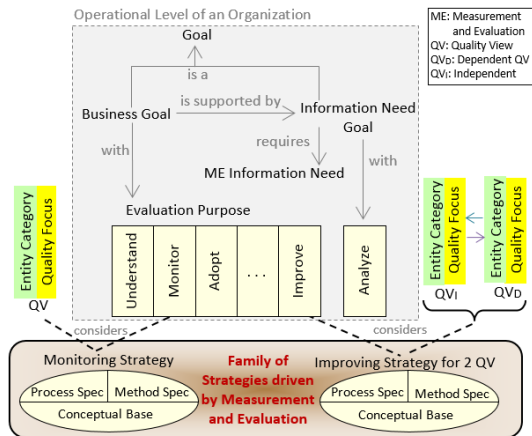


Fig. 1 Schema that relates Business Goal and Information Need Goal with Evaluation Purposes. Also, a Family of Strategies considers different purposes and Quality Views.

quality view. According to [14], a strategy should have at least three pillars such as a domain conceptual base, and the process and method specifications. These allow to have explicitly defined the evaluation domain in order to know what generic and specific activities intervene, and how to perform them through methods. Since a strategy considers the quality view concept, it is worth mentioning that it represents an association between an entity category and a quality focus (see [18] for more details about the quality view concept).

Hence, in a specific ME or MEC project, the more suitable strategy should be selected from a family of strategies, taking into account the goal evaluation purpose, and the amount and types of quality views.

In summary, the contribution of this work is threefold. First, to identify and categorize a set of evaluation purposes for business and information need goals at operational level, which can be satisfied through the systematic use of strategies driven by ME, analysis and, eventually, change. Second, to document one strategy per each category of evaluation purposes. Lastly, to illustrate the applicability of the *comparing and adopting* strategy aimed at determining Usability strengths from LinkedIn, Instagram, Twitter and Facebook social network apps to be adopted in a new entity.

Note that this paper is an extended version of [20]. In the present work, regarding the second contribution, we document two new strategies from the family of strategies, including their scenario descriptions, examples and new figures related to the strategy's process. Considering the third contribution, we include more details about the activities, methods and work products consumed/produced in the practical case. For instance, a metric and indicator methods are specified. In addition, the performed analysis was

extended.

The rest of the article is organized as follows. Section 2 defines the main terms for a better comprehension of this work. Note that some terms and a new figure were added regarding [20]. Section 3 analyses evaluation purposes in the context of the state-of-the-art literature and proposes three categories. Section 4 specifies strategies for the *monitor*, *improve*, and *compare and adopt* purposes, illustrating them with concrete evaluation scenarios. Section 5 illustrates the *comparing and adopting* strategy. Section 6 discusses related work. Finally, Section 7 summarizes conclusions and future work.

2. Definition of some Key Terms

As above mentioned, well-established strategies need three pillars, namely: process specifications, method specifications, and domain vocabularies. Each strategy specifies its own processes and methods that helps accomplishing a specific goal purpose. However, the family of evaluation strategies uses a common vocabulary, which is structured in ontologies as a way to explicitly organize and define the terms, attributes and relationships used in activity and method specifications for non-functional requirements definition, measurement, evaluation and improvement. Looking at the shaded part of Fig. 2, we had specified sub-ontologies (in [2, 19]) for the business goal, project, context, non-functional requirements, measurement and evaluation components. However, the rest of the conceptual components viz. the functional requirements, development and maintenance ones are still unspecified. Having well-structured vocabularies serving as common ground for diverse strategies may promote a more effective operationalization of projects dealing with evaluation, testing, development and maintenance goal purposes.

For space reasons just some terms presented in Fig. 1, which are fully used for process and methods specifications in Sections 4 and 5, are defined below (note that all the terms, attributes and relations are defined in [2, 19]):

- *Goal* (synonym: Organizational Goal): the statement of the aim to be achieved by an organization which considers the propositional content of a purpose in a given timeframe.
- *Business Goal*: it is the main goal that the organization intends to achieve.
- *Information Need Goal*: it is a goal intended to get insight for a given business goal.
- *ME Information Need Goal*: it is an information need goal driven by ME activities.
- *Purpose*: the rationale for achieving a specific goal. Purposes can be classified into evaluation,

development, testing, and maintenance/change purpose categories. In turn, evaluation purposes (i.e., purposes that fall into the evaluation category) can be subcategorized as we discuss in

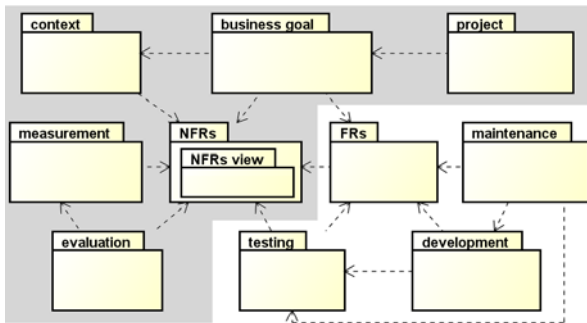


Fig. 2 Conceptual components (sub-ontologies) for the Family of Strategies. Note: FRs stands for functional requirements and NFRs for non-functional requirements.

Section 3.

- *Strategy*: principles, patterns, and particular domain concepts and framework that can be specified by a set of core processes, in addition to a set of appropriated methods and tools, as core resources, for helping to achieve the project's goal purpose.
- *Measurement*: a task that uses a metric definition in order to produce a measure's value. Note: a metric is a method that specifies the measurement or calculation procedure and its scale.
- *Evaluation*: a task that uses an indicator definition in order to produce an indicator's value. Note: an indicator is a method that specifies the calculation procedure and its scale in addition to the model and decision criteria in order to provide an evaluation of a calculable concept with respect to a defined information need.

3. Evaluation Purposes and Categories

Usually, in Software Engineering, projects are intended to and named as evaluation project, testing project, development project, among other types. A given project type can operationalize diverse goal purposes. For example, evaluation projects operationalize purposes like understand, improve, monitor, among others. While development projects operationalize purposes such as create, add, delete, among others. Therefore, we also can distinguish categories of purposes as evaluation, testing, development, and maintenance/change categories. In turn, purposes for the evaluation category can be subcategorized.

In this Section, we address the first stated contribution, viz. evaluation purposes and their categorization.

Analyzing the measurement and evaluation literature, noteworthy, still there is no full consensus about definitions for purposes, particularly, for those purposes that involve evaluation and change activities. Many works consider that *characterize* and *understand* purposes are the same one, which involve to understand or take a snapshot of the current situation of an entity for establishing baselines for future assessment [3, 8, 10, 13]. However, the *characterize* purpose could be achieved previously to the *understand* purpose, since it would allow the characterization of an entity from the relevant nonfunctional requirements identification standpoint. Once identified the relevant characteristics, the understanding of the current entity state can be performed by means of ME activities. On the other hand, the *predict* purpose usually is intended to identify and understand cause/effect relationships between characteristics/attributes of entities [8] aimed at establishing predictive models that could be useful in MEC activities.

Besides, the *monitor* purpose implies tracking the status or performance of attributes of an entity [3]. That is, it consists of a continuous understanding of an entity over time for analyzing its evolution and possible trends. Note that monitoring includes the *understand* purpose applied in a repetitive way, at a given frequency for a defined interval of time. Sometimes, the *control* purpose is related with *monitor* since it addresses the identification of deviations that influence the status or performance of processes and products for reducing risks [3]. Some authors [3, 10] consider the *control* and *monitor* purposes as a single purpose, due to they involve a continual evaluation and occasionally changes for improving the expected performance.

On the other hand, Preece and Rombach [17] categorized evaluation purposes as passive or active:

"passive purposes are aimed at better understanding or visualizing existing software items without influencing them, whereas active ones are aimed at actually influencing them in some way".

We can infer that purposes for the active category always imply changes on the target entity and/or its context, aimed at improving it. Conversely, evaluation purposes for the passive category imply no changes. However, we observe that both categories do not fit well the intention of all evaluation purposes.

For example, the *compare and adopt* purpose, which involves the determination of strengths and weaknesses of a set of preselected representative entities, and the adoption of strengths in a new entity or in one that already exist, could be both active and passive. If it is about adopting detected strengths in an existent entity, this purpose implies changes and falls into the active category; otherwise, if the strengths are adopted in a new entity to be developed, it falls into the passive category.

Furthermore, exists the *select an alternative* purpose among preselected competitive entities aimed at adopting and installing it. As a result, for purposes that have entity preselection – independently if purposes are active or passive- a new category can be envisioned.

Consequently, we envision three categories for evaluation purposes, namely: *Characterize and Understand*, *Control and Change*, and *Adopt an Alternative*.

The *Characterize and Understand* category includes purposes such as to *understand*, *monitor* and *predict*. *Understand* implies getting information about the current state/performance of an entity, so change actions are not performed on the entity. Note that *understand* may also foster to find additional entity's characteristics to be included in future evaluations. In this sense, the achievement of the *understand* purpose can help to *characterize* as well. However, in our approach, the measurement and evaluation activities are not intended directly to *characterize*. Regarding the *monitor* purpose, as above mentioned, it consists of a continuous understanding over time. The *predict* purpose helps to understand possible cause-effect relations among characteristics/attributes of entities. This category matches the passive category mentioned by [17].

The *Control and Change* category includes purposes such as to *improve*, and *monitor and control*. These purposes are not just intended for understanding the current state or performance of an entity, but also for improving it by introducing changes. In addition, these purposes can be achieved taking into account one or two quality views [18]. The *improve* purpose is related with the identification of problems, vulnerabilities and others opportunities for improving the entity quality. So, it always implies changes. In turn, the *monitor and control* purpose implies performing a critical analysis in order to maintain under control the detected performance problems by doing corrective actions, if needed. *Control and Change* category is similar to the active category considered in [17].

The *Adopt an Alternative* category includes purposes such as to *select an alternative* and to *compare and adopt*. These purposes share one activity, which is devoted to preselect competitive or representative entities. The *select an alternative* purpose implies to understand the current state of each preselected competitive entity and therefore select the one with the best performance. Note that the selected entity does not undergo changes, but rather is adopted for installation and/or use. On the other hand, the *compare and adopt* purpose is based on determining strengths and weaknesses of a set of representative entities in order to adopt strengths in an existing entity or in one to be developed. Thus, this purpose may imply changes on the existing entity.

4. Some Evaluation Strategies from the Family of Strategies

Given an evaluation goal, the selection of the suitable strategy from a family of strategies is a critical aspect. The strategy defines a specific course of action to be followed for achieving the goal purpose. In order to illustrate one evaluation strategy per each category analyzed in the previous Section, we use three evaluation scenarios. That is, one *monitoring* strategy for the *monitor* purpose (which is included into the *Characterize and Understand* category); one strategy for the *improve* purpose (which is included into the *Control and Change* category); and, lastly, one strategy for the *compare and adopt* purpose (which is included into the *Adopt an Alternative* category).

Each scenario includes the business goal purpose, the amount/type of involved views, the scenario description and the specification of the generic process of the strategy. In addition, one example of each scenario is given.

4.1. Evaluation Scenario for the *Monitor* Purpose

Business goal purpose: Monitor.

Type of view: Quality. Amount of views: One.

Scenario description: The business goal purpose at operational level is to monitor the situation of an entity, in a given context, for a set of characteristics and attributes related to a quality focus, through the systematic use of a monitoring strategy driven by measurement, evaluation and analysis activities. The measurement activity is performed by quantifying attributes by means of the selected metrics. The evaluation activity is performed by interpreting characteristics and attributes by means of indicators. The measurements and evaluations are carried out periodically (with more or less frequency) in order to understand -during the established time- the situation or state of the evaluated entity, to determine comparisons and trends in the performance level of the indicators, and/or to establish baselines.

Evaluation scenario example:

- Business goal statement: Monitor the Quality in Use of the JGUIAr during a semester with a monthly frequency. (Note that JGUIAr stands for *Java Graphic User Interface Architect*. It is a tool for designing Java AWT GUIs developed by Dr. H. Molina in the context of an Object-oriented Programming course in the Engineering School at UNLPam. Its version 1.3 is available at <https://sites.google.com/site/jguiarsoftware/>).
- Quality View: System-in-Use Quality View.
- Entity Category: System in Use.

- Quality Focus: Quality in Use.
- Concrete Entity: JGUIAr in use.
- Characteristic: Actual Usability (Sub-characteristics: Effectiveness, Efficiency and Learnability in Use).

Strategy to be applied: GOCAMEM (*Goal-Oriented Context-Aware Measurement, Evaluation and Monitoring*).

Generic process specification for GOCAMEM: see Fig. 3.

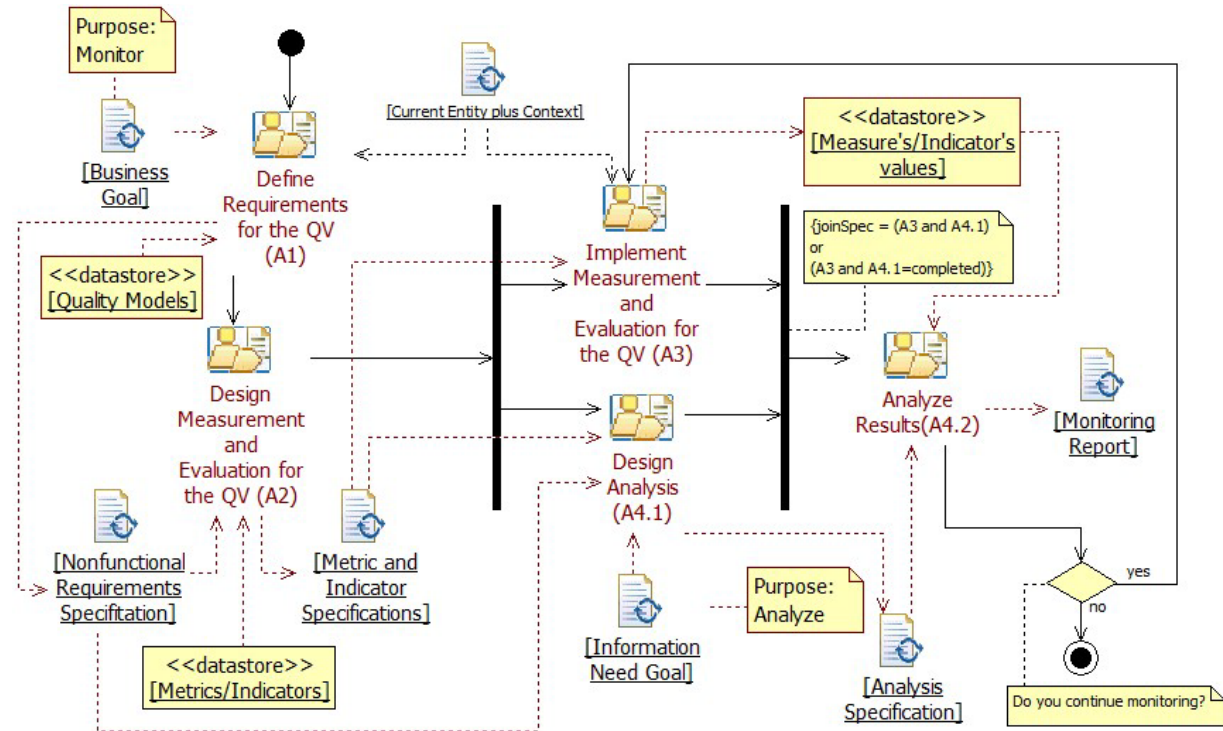


Fig. 3 Generic process specification for the GOCAMEM strategy for one Quality View (QV).

The GOCAMEM process begins with the definition of non-functional requirements (A1 activity). This activity is devoted to define the characteristics and attributes to be evaluated. In A2, the ME are designed by selecting the most suitable metrics and indicators from a repository.

Then, in A3, based on the specification of metrics and indicators, ME are implemented to obtain the measure's/indicator's values. Concurrently with A3 activity, A4.1 can be performed. In the latter, the analysis is designed and the document "Analysis Specification" is obtained, serving as input to A4.2. The A4.2 activity ("Analyze Results") permits obtaining conclusions and recommendations about strengths and weaknesses of the evaluated entity, providing an analysis of the current status of the entity during the monitoring period. A4.2 also allows to perform a comparative analysis on the evolution and trends for the level of performance of elementary and global indicators, as well as to establish baselines in order to adjust decision criteria.

Note that in Fig. 3 the cycle between A3 and A4.2 indicates that the measurement, evaluation and analysis are performed over time in order to understand the situation of the entity in the established monitoring period.

4.2. Evaluation Scenario for the *Improve Purpose*

Business goal purpose: Improve.

Type of view: Quality. Amount of views: One.

Scenario description: The business goal purpose at operational level is to understand and improve the current state of an entity, in a given context, for a set of characteristics and attributes related to a quality focus, through the systematic use of an improvement strategy driven by measurement, evaluation, analysis and change activities. The measurement activity is performed by quantifying attributes by means of the selected metrics. The evaluation activity is performed by interpreting characteristics and attributes by means of indicators. The analysis is based on determining weaknesses or vulnerabilities of the evaluated entity in a given moment, which produces a conclusion and recommendation report about opportunities for improvement. The improvement can be achieved by means of changes in the entity and/or in its context. Once the changes were performed, the new entity version (and/or context) is re-evaluated for analyzing the actual impact and gain of the improvement.

Evaluation scenario example:

- Business goal statement: Improve 15 percentage points the JGUIAr External Quality in 3 months.
- Quality View: System Quality View.
- Entity Category: System.
- Quality Focus: External Quality.
- Concrete Entity: JGUIAr application.
- Characteristics: Usability, Information Quality and Functional Quality.

Strategy to be applied: GOCAMEC (*Goal-Oriented Context-Aware Measurement, Evaluation and Change*).

after changes on the entity, a re-evaluation should be performed. If the planned improvement gain on quality after the implemented changes was not achieved (i.e., the operational business goal was not achieved considering the information need), then new change and re-evaluation cycles should be performed.

4.3. Evaluation Scenario for the *Compare and Adopt Purpose*

Business goal purpose: Compare and Adopt.

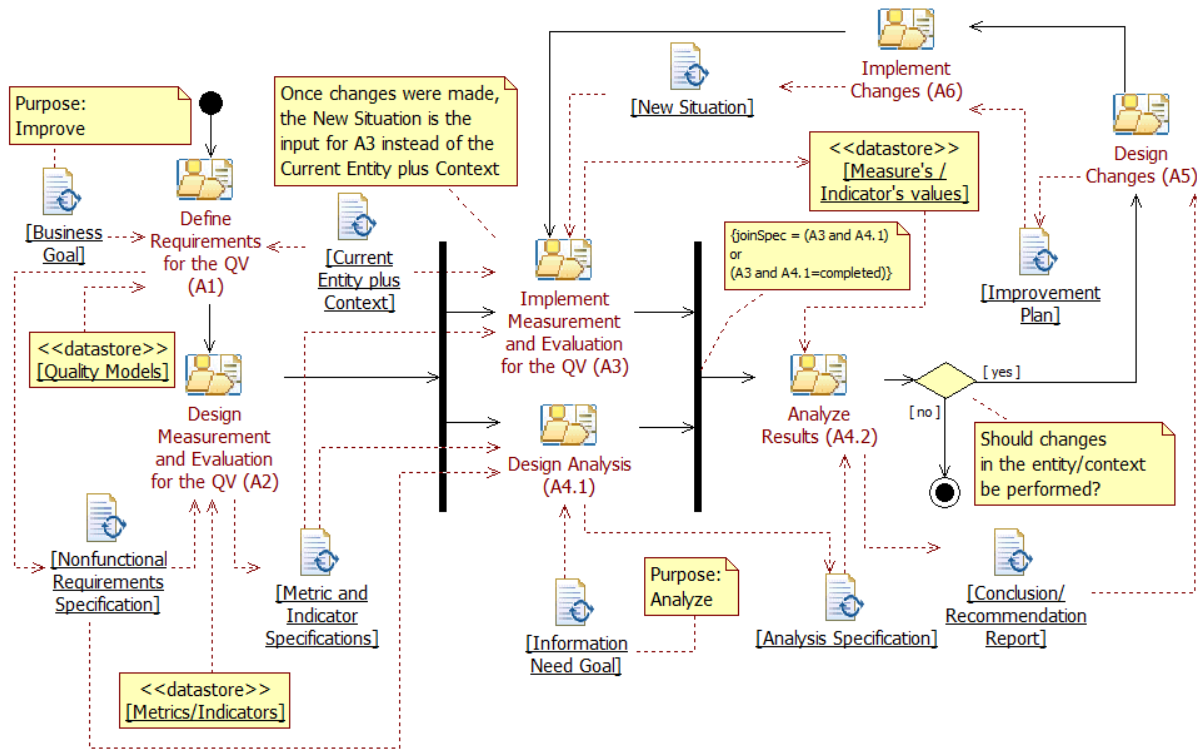


Fig. 4 Generic process specification for the GOCAMEC strategy for one Quality View (QV).

Generic process specification for GOCAMEC: see Fig. 4.

From A1 to A4.2 activity, the process in Fig. 4 is very similar to the process in Fig. 3. However, in GOCAMEC the analysis (A4.1) includes, among others aspects, the established criteria to decide when an entity needs changes to improve the performance of attributes. Then, A4.2 uses the measures, indicator's values and the analysis specification as input, and produces the recommendation report as output. This report indicates change actions to be made in order to improve the vulnerabilities (or inefficiencies) detected in the entity/context. Consequently, for designing and implementing changes, A5 and A6 must be done respectively. Changes are designed and implemented according to the improvement recommendations given in the A4.2 activity.

Aimed at knowing the quality impact produced

Type of view: Quality. Amount of views: One.

Scenario description: The business goal purpose at operational level is to compare characteristics and attributes from a set of representative entities, in a given context, with the aim of adopting recommended strengths through the systematic use of a comparing strategy driven by ME, analysis and eventually, change. The measurement activity is performed by quantifying attributes (using metrics) and the evaluation activity is performed by interpreting characteristics and attributes (using indicators). The comparative analysis is based on determining strengths and weaknesses of the preselected entities in a given moment, for recommending and adopting the detected strengths in a new entity or in one that already exists.

Evaluation scenario example:

- Business goal statement: Compare a set of social network applications for smartphones with the aim of adopting the best Usability characteristics

in a new social network application to be developed.

- Quality View: System Quality View.
- Entity Category: System.
- Quality Focus: External Quality.
- Concrete Entities: Facebook, Instagram, Twitter and LinkedIn.
- Characteristics: Usability (Sub-characteristics: Comprehensibility, Learnability, Operability and User error protection).

Strategy to be applied: GOCAMECom (*Goal-Oriented Context-Aware Measurement, Evaluation and Comparison*).

in parallel with A3, which involves implementing the ME per each preselected entity as shown in Fig. 5 (producing measures and indicator's values).

In the sequel, A4.2 uses the measures, indicator's values and the analysis specification as input, and produces the recommendation report as output. This report contains the strengths to be adopted in an existing entity or in a new one to be developed. If the target entity already exists and weaknesses were detected, then changes should be carried out for adopting the strengths of the other compared entities. Thus, for designing and implementing changes, A5 and A6 must be done respectively.

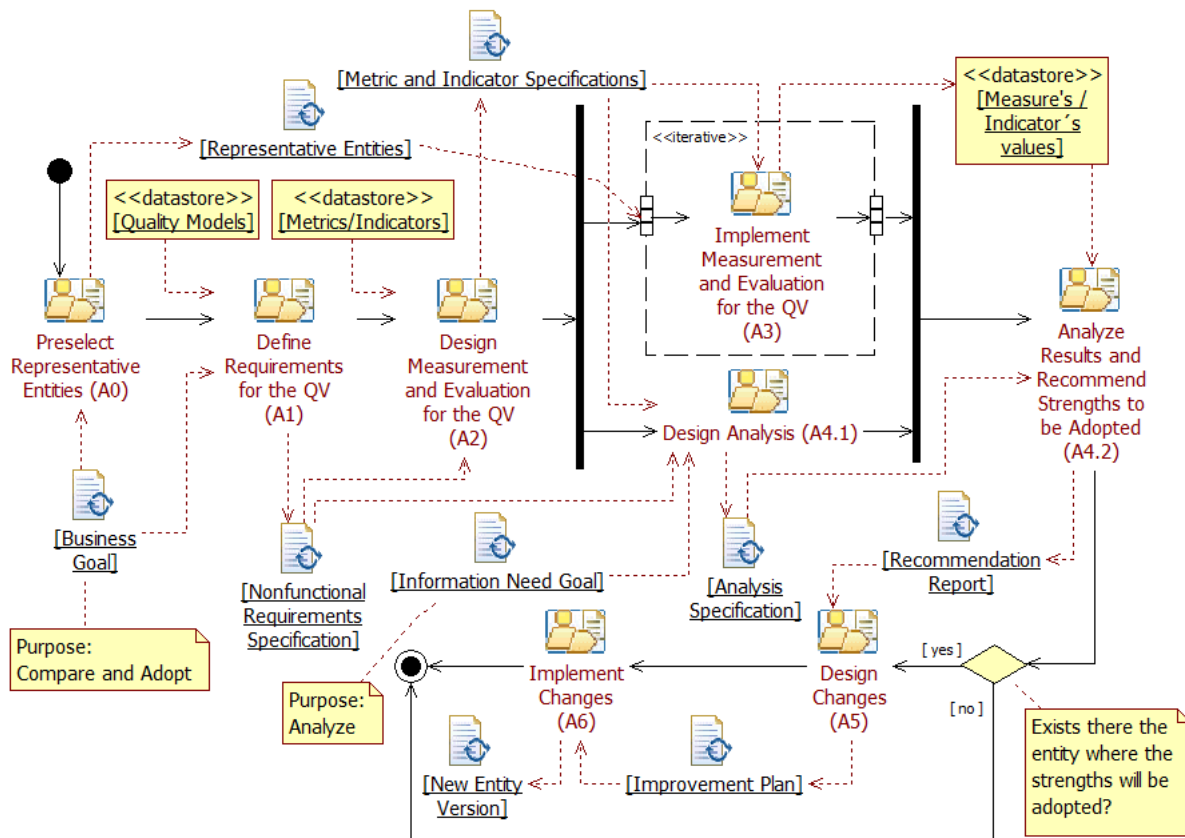


Fig. 5 Generic process specification for the GOCAMECom strategy for one Quality View (QV).

Generic process specification for GOCAMECom: see Fig. 5.

The GOCAMECom process begins with the preselection of representative entities to be compared (A0 activity). The preselection can be based on methods such as expert judgment, case studies, among others. Note that if the strengths would be adopted in an existing entity, this should be one out of the preselected entities. Then, A1 is devoted to define the characteristics and attributes to be evaluated. In the A2 activity the ME are designed by selecting the most suitable metrics and indicators. In A4.1 the analysis is designed, which includes, among others aspects, the establishment of criteria for recommending and adopting. A4.1 can be done

Otherwise, for a new entity, A5 and A6 should not be performed. Instead, a development process should begin.

4.4. Abridged Remarks about Strategies, Purposes and Categories

To summarize, we would like to point out two aspects. Firstly, there are evaluation purposes which are aimed at better understanding the current state or performance of an entity, without influencing it (i.e., without introducing changes in the target entity), another evaluation purposes are aimed to some extent at influencing the target entity (or its context), and another purposes are aimed at selecting an entity

or the better characteristics from a set of entities.

For the Characterize and Understand category, we developed strategies for *understanding* (not illustrated in this paper) and *monitoring* (subsection 4.1), in which change activities are not included (thus, the FRs concepts do not intervene –recall Fig. 2). For the Control and Change category, we developed strategies for *improving* (subsection 4.2), and *monitoring and controlling* (not illustrated in this paper), in which change activities are included to enhance the target entity quality or to keep under control some parameters. Therefore, both NFRs and FRs concepts are used in their activities and methods. Lastly, for the Adopt an Alternative category, we developed strategies for *selecting an alternative*, and *comparing and adopting*. The strategy for selecting one alternative from a set of competitive entities (not illustrated in this paper) does not imply a change in the target entity, but rather its adoption and installation in an organization. While the strategy for comparing competitive entities and adopting the best capabilities or features to a target entity may imply changes as described in subsection 4.3. Both strategies imply to preselect a set of entities.

Secondly, we can observe a high level of activity reuse by looking at the process specifications of illustrated strategies. Thus, all process specifications of the family of strategies share the A1-A4 activities. While the A1-A3 activities are the same at task level for all strategies, the A4.1 and A4.2 activities can vary slightly at task level depending on the specific evaluation purpose. Furthermore, for those evaluation purposes that embrace changes, the A5-A6 activities can be reused totally. Note that while activity specifications can be the same for designing and implementing changes, different methods can be applied such as programming, refactoring, restructuring, or re-parameterization. On the other hand, the A0 activity can be reused for those strategies in which preselection is needed.

Despite the fact that many activities are reused in the process specifications of all evaluation strategies, the dynamic of activities may slightly differ each other (compare figures 3, 4 and 5, in which processes are specified from the behavioral perspective). Ultimately, the sound specification of strategies for different evaluation purposes helps to know what to do and how to do in a systematic and disciplined way for achieving business goal purposes.

5. Running Example: Application of the Comparing and Adopting Strategy

In the scenario example of subsection 4.3, the statement of the business goal, namely, “*compare a*

set of social network applications for smartphones with the aim of adopting the best Usability characteristics in a new social network application to be developed” contains the *compare and adopt* purpose. This belongs to the *Adopt an Alternative* category. Also, the amount of involved quality views is one: The *System Quality View*. So, GOCAMECom is the strategy to be instantiated for achieving this goal purpose. It is worthy to remark that the activities of Fig. 5 are renamed considering the specific quality view because the process specification in Fig. 5 is generic. For example, A1 is now renamed “*Define Requirements for the System Quality View*”, A2 is renamed as “*Design Measurement and Evaluation for the System Quality View*”, and so forth for the remainder activities. Let us describe the GOCAMECom’s activities.

(A0) Preselect Representative Social Network Mobile Apps: The selection was done at random from the 15 most popular social network apps listed at www.ebizmba.com/articles/social-networking-websites. As a result, four representative entities for Android were selected, namely: Instagram (v.10.0.0), Facebook (v. 106.0.0.26.28), Twitter (v. 4.0.100) and LinkedIn (v. 6.28.0). Note that the version of each selected app was the most recent at the time of this study (in December, 2016).

(A1) Define Requirements for the System Quality View: This activity produces the “*Nonfunctional Requirements Specification*” document, which includes the “*Business Goal and Information Need Specification*”, the “*Context Properties*”, and the “*Requirements Tree for External Quality*” (see Table 1, 1st column). *Usability* is the root characteristic of the requirements tree. For this characteristic, 17 attributes were selected –see their definitions in [12]. These are related to some of the following sub-characteristics: *Comprehensibility*, *Learnability*, *Operability* and *User Error Protection*.

(A2) Design Measurement and Evaluation for the System Quality View: In this activity, metrics are selected to quantify all the attributes of the requirements tree shown in Table 1. Fig. 6 specifies the *Foreign Language Support Level* metric, which quantifies the *Foreign Language Support* attribute (coded 1.1.1.3 in Table 1).

In order to interpret the metric’s values, an elementary indicator should be selected for each attribute as well. The indicator scale (e.g., in the percentage unit) is interpreted using agreed decision criteria, which help to analyze the level of satisfaction met by each attribute. Particularly, we used in our study three acceptability levels, namely: “*●Satisfactory*” whose values ranges between [90-100%]; “*◆Marginal*” [80-90]; and “*■Unsatisfactory*” [0-80). Fig. 7 specifies the elementary indicator for the *Foreign Language*

Support attribute.

Name: Foreign Language Support Level (FLSL)	
Objective: Determine the level of foreign language support.	
Author: Tebes – Peppino	Version: 1.0
Measurement Procedure:	Type: Objective
Specification: The inspector analyzes the application searching for language configuration to determine the category (0, 1 or 2). Where 0 means that the application does not have any option to change the language. 1 means that the application has 3 or less languages to be configured. And 2 means that the application has more than 3 languages to be configured.	
Numerical Scale:	Representation: Discrete
Value Type: Integer	Scale Type: Interval

Fig. 6 Direct metric specification for the *Foreign Language Support* attribute.

Name: Performance Level of the Foreign Language Support (P_FLS)	
Author: Tebes – Peppino	Version: 1.1
Elementary model:	
Specification: the mapping is:	
P_FLS	$\left\{ \begin{array}{ll} 0 & \text{iff FLSL}=0 \\ 80 & \text{iff FLSL}=1 \\ 100 & \text{iff FLSL}=2 \end{array} \right.$
Decision criterion (3 acceptability levels):	
Name 1: ■ Unsatisfactory; Range: if $0 \leq P_FLS < 80$	
Description: Indicates that it is not advisable to be adopted.	
Name 2: ◆ Marginal; Range: if $80 \leq P_FLS < 90$	
Description: Indicates that it could be adopted.	
Name 3: ● Satisfactory; Range: if $90 \leq P_FLS \leq 100$	
Description: Indicates that it is advisable to be adopted.	
Numerical Scale:	Representation: Continuous
Value Type: Real	Scale Type: Ratio
Unit: Name: Percentage	Acronym: %

Fig. 7 Elementary indicator specification for the *Foreign Language Support* attribute. Note that FLSL stands for the *Foreign Language Support Level* metric, with 0, 1 or 2 allowed values (see Fig. 6).

Additionally, derived indicators are defined by means of an aggregation model, which permits interpreting characteristics and sub-characteristics of a requirements tree. To the present study, the LSP (*Logic Scoring of Preferences*) [7] aggregation method was selected, which supports the specification of weights for elements. Weight

represents the relative importance of elements (characteristics and attributes) in a requirements tree. In this case, weights were agreed by experts (see 2nd column in Table 1), but other techniques can be used.

Note that in the generic process of Fig. 5 both metrics and indicators are retrieved from a repository (*“Metrics/Indicators <<datastore>>”*). The work product produced in A2 is named *“Metric and Indicator Specifications”*.

(A3) Implement Measurement and Evaluation for the System Quality View: This activity produces the measures and indicators’ values. Indicators’ values both for elementary and derived indicators are shown in Table 1, from the 3rd to 6th column for Facebook, Twitter, Instagram and LinkedIn mobile apps respectively. Measured and calculated values are recorded in the *“Measure’s/Indicator’s values <<datastore>>”*. It is important to remark that the measurement was performed by two data collectors following the measurement procedure specified in the corresponding metric.

(A4.1) Design Analysis: Concurrently to A3 the A4.1 activity was carried out. In A4.1, a set of criteria was agreed for deciding whether an entity attribute should be considered as a strength to be adopted or not. Thus, an attribute is considered a strength if its indicator value falls into the *“●Satisfactory”* acceptability level. When several entities share the same strength (attribute), the one with the highest score is recommended to be adopted. In case that they have the same score, any of them could be recommended. For those attributes which are not benchmarked satisfactorily in any selected entity but fall into the *“◆Marginal”* level, their adoption could be considered taking into account that some kind of improvement must be included.

(A4.2) Analyze Results and Recommend Strengths to be Adopted: After A3 and A4.1 were performed, the instantiated A4.2 activity can be enacted, which produces a *“Recommendation Report”* following the criteria documented in the *“Analysis Specification”*. This report includes the strongest-benchmarked attributes to be adopted in the new mobile app GUI to be developed.

Looking at the resulting *Usability* indicator values in Table 1, no app reached the *“●Satisfactory”* level. Nevertheless, within the *“◆Marginal”* level, Facebook got the best score (87.56%) followed by Twitter (85.65%). While Instagram and LinkedIn fell into the *“■Unsatisfactory”* acceptability level, with 70.47% and 69.85% values respectively.

Following the analysis at attribute level, only Twitter obtained the maximum score (100%) for *Contextual Control Icon Ease to be Recognized*

(1.1.1.2.2), *Current Location Feedback Appropriateness* (1.2.1.1) and *Mandatory Entry* (1.3.1.2). Also, it has the highest score for the *Brightness Difference Appropriateness* (1.3.2.1.1) and *Color Difference Appropriateness* (1.3.2.1.2) attributes, which reached 91.40% and 90.44% respectively. These five attributes are strengths in the Twitter app and therefore are recommended to be adopted in the new entity.

On the other hand, Facebook met the highest score for the *Error Message Appropriateness* (1.2.1.3) and *Context-Sensitive Help Appropriateness* (1.2.2.1) attributes, i.e., 93.75% and 90% respectively. Thus, these attributes should be adopted from Facebook.

Permanence of Main Controls (1.3.3.1.1), *Permanence of Contextual Controls* (1.3.3.1.2) and *Error Recovery* (1.4.1.2), more than one entity met the highest score. E.g., both Instagram and Facebook reached the best score (100%) for *Foreign Language Support* attribute. In these cases, it is recommended to adopt the strength of any of these entities.

Finally, regarding *Main Control Icon Ease to be Recognized* (1.1.1.2.1) and *Error Prevention* (1.4.1.1) attributes, it is worth to remark that the highest scores fall into the “◆Marginal” level. Therefore, it may be recommended to adopt them, but considering that improvements must be incorporated to reach a “●Satisfactory” level. For 1.1.1.2.1, it could be adopted from LinkedIn or

Table 1 Requirements Tree for Usability (1st column) with weights or importance (2nd column) and indicator's values (in [%]) for each compared mobile app: Facebook (3rd c.), Twitter (4th c.), Instagram (5th c.) and LinkedIn (6th c.).
The ● symbol means “Satisfactory” acceptability level; ◆ “Marginal”, and ■ “Unsatisfactory”.

Characteristics and Attributes (<i>in italic</i>)	weight	Facebook	Twitter	Instagram	LinkedIn
1. Usability		87.56 ◆	85.65 ◆	70.47 ■	69.85 ■
1.1. Understandability	0.20	91.15 ●	85.12 ◆	89.82 ◆	35.88 ■
1.1.1. Familiarity	1.00	91.15 ●	85.12 ◆	89.82 ◆	35.88 ■
1.1.1.1. <i>Global Organization Scheme Understandability</i>	0.50	100 ●	100 ●	100 ●	0 ■
1.1.1.2. <i>Control Icon Ease to be Recognized</i>	0.40	77.88 ■	87.80 ◆	74.56 ■	89.71 ◆
1.1.1.2.1. <i>Main Control Icon Ease to be Recognized</i>	0.60	85.71 ◆	80 ◆	80 ◆	85.71 ◆
1.1.1.2.2. <i>Contextual Control Icon Ease to be Recognized</i>	0.40	66.67 ■	100 ●	66.67 ■	95.83 ●
1.1.1.3. <i>Foreign Language Support</i>	0.10	100 ●	0 ■	100 ●	0 ■
1.2. Learnability	0.20	88.22 ◆	81.34 ◆	77.10 ■	81.23 ◆
1.2.1. Feedback Suitability	0.50	86.46 ◆	79.38 ■	74.25 ■	87.65 ◆
1.2.1.1. <i>Current Location Feedback Appropriateness</i>	0.35	66.67 ■	100 ●	80 ◆	83.33 ◆
1.2.1.2. <i>Alert Notification Feedback Appropriateness</i>	0.35	100 ●	62.50 ■	100 ●	87.50 ◆
1.2.1.3. <i>Error Message Appropriateness</i>	0.30	93.75 ●	75 ■	37.50 ■	92.86 ●
1.2.2. Helpfulness	0.50	90 ●	83.33 ◆	80 ◆	75 ■
1.2.2.1. <i>Context-Sensitive Help Appropriateness</i>	1.00	90 ●	83.33 ◆	80 ◆	75 ■
1.3. Operability	0.30	80.96 ◆	98.15 ●	71.76 ■	89.19 ◆
1.3.1. Data Entry Ease	0.40	80 ◆	100 ●	60 ◆	80 ◆
1.3.1.1. <i>Defaults</i>	0.40	100 ●	100 ●	100 ●	100 ●
1.3.1.2. <i>Mandatory Entry</i>	0.40	50 ■	100 ●	50 ■	50 ■
1.3.1.3. <i>Widget Entry Availability</i>	0.20	100 ●	100 ●	0 ■	100 ●
1.3.2. Visibility	0.20	77.99 ■	90.92 ●	78.05 ■	86.91 ◆
1.3.2.1. <i>Color Visibility Appropriateness</i>	1.00	77.99 ■	90.92 ●	78.05 ■	86.91 ◆
1.3.2.1.1. <i>Brightness Difference Appropriateness</i>	0.50	78.33 ■	91.40 ●	79.70 ■	87.08 ◆
1.3.2.1.2. <i>Color Difference Appropriateness</i>	0.50	77.65 ■	90.44 ●	76.44 ■	86.74 ◆
1.3.3. Consistency	0.40	83.45 ◆	100 ●	81.13 ◆	100 ●
1.3.3.1. <i>Permanence of Controls</i>	1.00	83.45 ◆	100 ●	81.13 ◆	100 ●
1.3.3.1.1. <i>Permanence of Main Controls</i>	0.80	84 ◆	100 ●	80 ◆	100 ●
1.3.3.1.2. <i>Permanence of Contextual Controls</i>	0.20	81.25 ◆	100 ●	85.71 ◆	100 ●
1.4. User Error Protection	0.30	91.67 ●	77.50 ■	55 ■	75 ■
1.4.1. Error Management	1.00	91.67 ●	77.50 ■	55 ■	75 ■
1.4.1.1. <i>Error Prevention</i>	0.50	83.33 ◆	80 ◆	60 ■	50 ■
1.4.1.2. <i>Error Recovery</i>	0.50	100 ●	75 ■	50 ■	100 ●

As seen in Table 1, for *Global Organization Scheme Understandability* (1.1.1.1), *Foreign Language Support* (1.1.1.3), *Alert Notification Feedback Appropriateness* (1.2.1.2), *Defaults* (1.3.1.1), *Widget Entry Availability* (1.3.1.3),

Facebook, since both obtained 85.71%. Regarding 1.4.1.1 attribute, it is recommended to adopt it from Facebook, since it obtained the highest score, namely 83.33%.

Lastly, taking into account that the detected

strengths will be adopted in a new social network mobile app, the A5 (“*Design Changes*”) and A6 (“*Implement Changes*”) activities were not performed for this study. Recall that change activities (A5 and A6) are just performed when the target entity already exists, as shown in Fig. 5. So, once A4.2 was finished, the process for this case study finished as well.

6. Related Work and Discussion

Every organization struggle to achieve its business goals successfully. To do this, it is advisable to have a systematic approach, which permits the establishment of goals at different organizational levels, as well as organizing the work by means of projects and strategies for helping to reach goal purposes. Therefore, this systematic and integrated approach should include the following characteristics: i) the definition of business goals and information need goals at different organizational levels; ii) the definition of different evaluation purposes for goals; and iii) the conception of a family of evaluation strategies that helps achieving goal purposes.

In the present work, we have considered aspects of the two latter approach’s features. Regarding the evaluation purposes, many of them are documented in several works [3, 4, 8, 10, 16, 17]. However, there is no broad consensus yet in some purpose definitions. Even more, just in [17] a classification for purposes is presented, but it does not allow encompassing the intentionality of some purposes such as *select the suitable alternative* or *compare and adopt*. For example, the *compare and adopt* purpose is intended to determine strengths and weaknesses of a set of pre-selected representative entities in order to adopt strengths in an existing entity or in a new entity that needs to be developed. Note that in [17], the classification of evaluation purposes is based on the active category, which always imply changes in the target entity aimed at improving it. Or on the passive category which implies no change in the target entity. Therefore, in the context of this discussion, in Section 3, we have proposed three categories that embrace the intentionality of all evaluation goal purposes.

Regarding the third (iii) approach’s feature, in the current literature there is few works that deals with the importance of having a family of evaluation (ME/MEC) strategies that helps to achieve different evaluation goal purposes, considering also quality views. Additionally, it is important to remark that strategies are important resources in helping to achieve project’s goal purposes, since the well-specified processes and methods can benefit and foster their understandability, communicability and applicability.

For example, a MEC strategy that integrates

simultaneously the domain conceptual base, the process specification and the method specification is presented in [14]. But a family of strategies for different purposes is not considered.

A well-known approach is GQM+Strategies [1]. It includes a goal-oriented framework for the design and implementation of software measurement projects at different organizational levels. Unlike its predecessor, GQM (*Goal Question Metric*), the business goals that GQM+Strategies defines can be aligned at different organizational levels through the establishment of strategies. In [1], strategies define objectives for reaching goals and require the definition and fulfillment of lower level goals. Therefore, business goals are linked to measurement goals using GQM. Nevertheless, well-established process specifications for different ME/MEC strategies regarding different evaluation purposes and quality (quality-cost) views are missing.

Another related work is the *Goal-Driven Measurement* approach [15], which describes a general process for the definition of measurement goals aimed at helping to fulfill organizational goals. The approach process begins indicating that the organization should establish business goals at any organizational level. From these goals, questions or issues related to what stakeholders want to know or learn emerge. These issues allow identifying quantitative information through the decomposition of the business goal into related sub-goals. With the list of sub-goals and issues, entities and attributes are identified, following the GQM model and templates as well as its terminology. But authors, neither explicitly define concepts such as business goal, organizational level, information need goal, evaluation purpose, strategy, among others, nor specify different ME/MEC strategies regarding evaluation goal purposes and quality views.

Additionally, Goethert and Fisher [9] describe the GQ(IM) (*Goal Question Indicator Measurement*) approach, which merges the most prominent aspects of [15] with the *Balanced Scorecard* [11] paradigm for the decomposition of strategic goals into sub-goals. GQ(IM) is used to systematically establish organizational goals for each quadrant that Balanced Scorecard defines. Also, it helps identifying and defining measures and indicators. However, GQ(IM) does not deal with the use of different integrated strategies to fulfill business goal purposes from ME information need goals.

Lastly, a recent strategy for monitoring cloud services named *Cloud MoS@RT* has been specified in [5], which relies on a conceptual base. Also, it documents the process and activities using SPEM in addition to different methods and models. *Cloud MoS@RT* only deals with the monitoring purpose, so authors do not discuss a family of strategies for different evaluation goal purposes.

7. Conclusion and Future Work

For the first paper's contribution, we have analyzed in Section 3 a categorization for a set of evaluation purposes in which business and information need goals can be satisfied through the systematic use of a family of strategies driven by measurement, evaluation and eventually change activities.

We argue that the proposed categories may foster a clearer understanding and communication of the purposes' intention to different stakeholders. It can be highlighted that purposes in the *Characterize and Understand* category serve as a basis for the remainder category purposes. For example, it is necessary to understand the current state of an entity before implementing changes for improvement. Likewise, it is necessary to understand the current state of representative/competitive entities before making thoughtful decisions.

For the second paper's contribution, in Section 4, we have presented three strategies of the family of strategies. It is important to remark that many evaluation scenarios' examples can be envisioned for the same goal purpose (e.g., improve) considering the amount of quality views (as documented in [18], in which intervenes two quality views).

Regarding the third contribution, in Section 5, the GOCAMECom strategy for the comparison of four social network apps, which helps evaluators to determine the strengths to be adopted in a new smartphone app to be developed has been illustrated.

Note that a limitation of the current work is the lack of a broader validation that corroborates the different ME/MEC strategies fit well to the corresponding purpose and category. Hence, as an ongoing work, we are designing surveys in order to confirm their appropriateness with domain experts.

Lastly, considering that some sub-ontologies for the family of strategies remained unspecified (recall Fig. 2), we are developing the ontology for FRs. Once this ontology be agreed, the testing, development and maintenance sub-ontologies will be built and related to it.

Acknowledgements

This work and line of research are supported by Science and Technology Agency of Argentina, in the PICT 2014-1224 project at UNLPam.

Competing interests

The authors have declared that no competing interests exist.

References

- [1] Basili V., 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), pp. 57–65, (2010)
- [2] Becker P., Papa F., Olsina L.: Process Ontology Specification for Enhancing the Process Compliance of a Measurement and Evaluation Strategy, *CLEI Electronic Journal*, 18:(1), pp. 1-26, (2015)
- [3] Briand L., Differding Ch., Rombach D.: Practical Guidelines for Measurement-based Process Improvement, *Software Process Improvement and Practice Journal*, 2:(4), pp. 253-280, (1996)
- [4] Briand L., Morasca S., Basili V.: An Operational Process for Goal-driven Definition of Measures, *IEEE Transactions on Software Engineering*, 28:(12), pp. 1106-1125, (2002)
- [5] Cedillo P., Gonzalez-Huerta J., Insfrán E., Abrahao, S.: Towards Monitoring Cloud Services Using Models@run time. In *Workshop on Models@run.time, MODELS*, Valencia, Spain, pp. 31-40, (2014)
- [6] CMMI: Capability Maturity Model Integration, for Dev. V1.3. CMU/SEI:TR-033, (2010)
- [7] Dujmovic J.: A Method for Evaluation and Selection of Complex Hardware and Software Systems. 22nd Int'l Conference for the Resource Management and Performance Evaluation of Enterprise CS. CMG 96 Proceedings, pp. 368-378, (1996)
- [8] Fenton N., Pfleeger S.: *Software Metric: a Rigorous and Practical Approach*, 2nd Ed., PWS Publishing Company, (1996)
- [9] Goethert W., Fisher M.: Deriving Enterprise-Based Measures Using the Balanced Scorecard and Goal-Driven Measurement Techniques, *Software Engineering Measurement and Analysis Initiative*, CMU/SEI-2003-TN-024, (2003)
- [10] INCOSE Systems Engineering Measurement Primer: A Basic Introduction to Measurement Concepts and Use for Systems Eng., INCOSE-TP-2010-005-02, v2, (2010)
- [11] Kaplan R., Norton D.: *The Balanced Scorecard: Translating Strategy into Action*, Harvard Business Press, (1996)
- [12] Olsina L., Santos L., Lew P.: Evaluating Mobileapp Usability: A Holistic Quality Approach, In: *ICWE 2014*, S. Casteleyn, G. Rossi, and M. Winckler (Eds.): LNCS 8541, pp. 111-129, (2014)
- [13] OMG: *Business Motivation Mode (BMM)*, V1.3, (2015)
- [14] Papa M. F.: Toward the Improvement of a Measurement and Evaluation Strategy from a Comparative Study, In: *Current Trends in Web*

- Engineering, ICWE Int'l Workshops, M. Grossniklauss and M. Wimmer (Eds.), LNCS 7703, pp. 189-203, (2012)
- [15] Park R., Goethert W., Florac W.: Goal-Driven Software Measurement. A Guidebook, TR. CMU/SEI-96-HB-002, USA, (1996)
- [16] PMBOK: A Guide to the Project Management Body of Knowledge, 5th Edition, (2013)
- [17] Preece J., Rombach D.: A Taxonomy for Combining Software Engineering and Human-Computer Interaction Measurement Approaches: Towards a Common Framework. Int'l Journal of Human-Computer Studies, 41:(4), pp. 553-583, (1994)
- [18] Rivera B., Becker P., Olsina L.: Quality Views and Strategy Patterns for Evaluating and Improving Quality: Usability and User Experience Case Studies. In: Journal of Web Engineering, Rinton Press, USA, 15:(5&6), pp.433-464, (2016)
- [19] Rivera B., Becker P., Papa F., Olsina L.: A Holistic Quality Evaluation, Selection, and Improvement Approach driven by Multilevel Goals and Strategies. CLEI Electronic Journal, 19:(3), (2016)
- [20] Tebes G., Peppino D., Becker P., Papa M.F., Rivera M.B., Olsina L.: Strategy for Comparing and Adopting Strengths. In: Proceed. of XIV Workshop of SE - XXIII Congreso Argentino de Ciencias de la Computación (CACIC 2017), La Plata, Bs. As., Argentina, pp. 792-801, (2017)