Evaluating the suitability of a measurement repository for statistical process control

Conference Paper · January 2010
DOI: 10.1145/1852786.1852822 · Source: DBLP

3 authors, including:

Monalessa Perini Barcellos
Universidade Federal do Espírito Santo
68 PUBLICATIONS **250** CITATIONS
SEE PROFILE

Ricardo de Almeida Falbo
Universidade Federal do Espírito Santo
172 PUBLICATIONS **1,661** CITATIONS
SEE PROFILE

Evaluating the Suitability of a Measurement Repository for Statistical Process Control

Monalessa Perini Barcellos[1,2] Ana Regina Rocha[1] Ricardo de Almeida Falbo[2]

[1] COPPE, Federal University of Rio de Janeiro
Mailbox 68511
21945-970 – Rio de Janeiro, RJ, Brazil
+55 21 2562-8698

[2] Department of Computer Science,
Federal University of Espírito Santo
Ave. Fernando Ferrari, 514, CT VII
29075- 910 – Vitória, ES, Brazil
+55 27 4009-2654

monalessa@inf.ufes.br darocha@cos.ufrj.br falbo@inf.ufes.br

ABSTRACT
Software organizations have increased their interest on software process improvement (SPI). In high maturity levels, SPI involves implementing statistical process control (SPC), which requires measures and data that are suitable for this context. However, this has been pointed in the literature as one of the main obstacles for a successful implementation of SPC in SPI efforts. This paper presents an instrument for evaluating the suitability of measurement repositories in order to support software organizations implementing SPC.

Categories and Subject Descriptors
D.2.8 [**Software Engineering**]: Metrics – *performance measures, process measures, product measures*.

General Terms
Measurement.

Keywords
Measurement, Statistical Process Control, High Maturity.

1. INTRODUCTION
Statistical process control (SPC) was originally developed in the manufacturing area, aiming to support improvement programs [17]. In the context of software organizations, the use of the SPC can be considered recent, so there are still many doubts about it [5, 6].

Real cases of SPC implementation on software organizations have revealed a picture characterized by problems and situations that jeopardize the successful implementation of SPC [4, 11, 19, 21, 22]. The unsuitability of the defined measures and collected data is one of the main problems. It delays the SPC implementation, because firstly the measures should be adapted and only then the SPC techniques could be applied [8, 12, 13, 14, 19, 22].

In high maturity levels, such as levels 4 and 5 of CMMI (Capability Maturity Model Integration) [7], or levels A and B of the MR MPS.BR (Reference Model for Brazilian Software Process Improvement [16] SPC occurs after a measurement program has been institutionalized (a requirement of CMMI level 2 and MPS.BR level F). As a result of the measurement program, measures and data are stored in an organizational measurement repository. It is expected that these measures and data be suitable to be used in the SPC. However, as it was said before, generally this is not the case.

This paper presents an Instrument for Evaluating the Suitability of a Measurement Repository to SPC (IESMR). The paper is organized as follows: Section 2 presents a brief theoretical background about SPC; Section 3 describes the development of the instrument; Section 4 presents the IESMR; Section 5 presents some results from one practical experience using the IESMR, Section 6 discusses related works; and finally Section 7 presents the final considerations of this paper.

2. STATISTICAL PROCESS CONTROL
The growing market demands for better software products and services has increased the interest of software organizations in process improvement [10]. There are several software process improvement (SPI) frameworks, such as CMMI [7] and MR MPS.BR [16] and in almost all of them measurement plays an important role.

In order to use measurement for SPI at high maturity levels, new concepts and practices should be added to the traditional measurement programs. SPI in high maturity levels requires knowing the behavior of critical processes, determining their performance in previous executions, and so predicting their performance in current and future projects, checking if they will be able to achieve the established goals. In this context, SPI is carried out applying statistical process control (SPC) techniques.

SPC uses statistical tools to determine if a process is under statistical control [17]. A process is considered under statistical control if its behavior (described by collected data regarding performance measures defined for the process) is stable. That is, if its variations are within the expected limits, determined from historical data. A stable process has

4th International Symposium on Empirical Software Engineering and Measurement, ESEM '10, September 16-17, 2010, Bolzano-Bolzen, Italy. Copyright 2010 ACM *978-1-4503-0039-01/10/09…$10.00.*

repeatable behavior and it is possible to predict its performance in future executions and, thus, to elaborate realistic plans and to improve the process continuously. On the other hand, a process whose variations are out of the expected limits is considered an unstable process and the causes of these variations (called special causes) must be investigated and treated through improvement actions, which aim to stabilize the process [9]. Stabilizing their critical processes is a characteristic of high maturity organizations or organizations that are looking forward to achieve the highest maturity levels.

3. DEVELOPMENT OF THE INSTRUMENT

Figure 1 shows an overview of the development steps of the Instrument for Evaluating the Suitability of a Measurement Repository to SPC (IESMR).

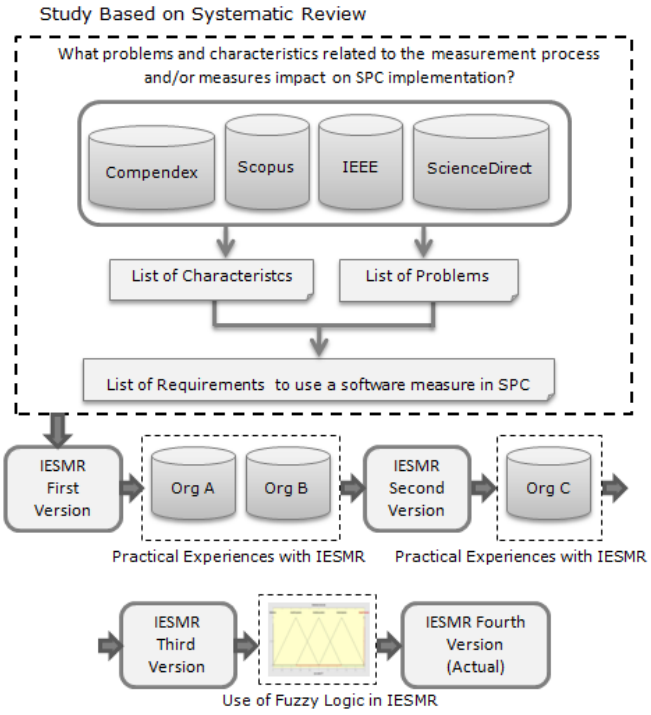

Figure 1. Development of the IESMR.

Initially we carried out a study based on systematic review of the literature from which we obtained two lists of finds: (i) list of factors related to the measurement process and/or measures that positively influence on SPC implementation (these factors were called *characteristics* in the study); and (ii) list of factors that negatively influence on SPC implementation (these factors were called *problems* in the study). From these lists, we identified a set of requirements for using a software measure in SPC. Due to the space limitations, the study based on systematic review is not detailed in this paper.

The set of requirements identified was used as basis for creating the first version of the IESMR. This first version was a checklist for evaluating a measure and its collected data to determine if the measure was suitable to SPC. This version of the IESMR was used to evaluate the measurement repositories of two software organizations. The main goal of these evaluations was to verify if the identified requirements were correct. For this, two questions were asked during the evaluations:

(a) Can a measure that fulfills the IESMR requirements actually be used in SPC?

(b) Is a measure that does not fulfills the IESMR requirements really unsuitable for SPC?

To answer these questions, first, the measures and the collected data were evaluated using the IESMR Then, the data collected for the evaluated measures were plotted in control charts. As result, we observed that the measures considered suitable to SPC, according to the IESMR, could be correctly plotted in control charts and they actually provided useful information about the performance of the processes. In contrast, the measures considered unsuitable to SPC by IESMR could not be plotted in control charts, or when they could be plotted, they did not describe the process performance and did not provide relevant information related to organizational goals.

Despite the fact that the initial results have shown that the set of requirements identified was correct, we noticed that the IESMR should be restructured. This was necessary because, to carry out the evaluation of the organizational measurement repositories, besides evaluating the measures and the collected data, it was also necessary to evaluate the Measurement Plan and the measurement repository structure.

So, we evolved the IESMR to four checklists: (i) checklist to evaluate the Measurement Plan, (ii) checklist to evaluate the measurement repository structure, (iii) checklist to evaluate the defined measures, and (iv) checklist to evaluate the collected data for the measures. Besides this change, in this version we described procedures for evaluating each requirement and the possible corrective actions to be taken when a requirement is not fulfilled.

The IESMR second version was used to evaluate the measurement repository of a third organization. This experience revealed the need for some minor adjustments in IESMR, basically related to the wording of the IESMR, aiming to improve its understanding.

Finally, whereas the evaluation of a measurement repository is somewhat subjective, we included in the IESMR some principles of Fuzzy Logic to determine the suitability degree of a measurement repository to SPC. This change resulted in the IESMR current version, which is presented in the next section.

The IESMR is part of a strategy defined to support organizations to obtain and maintain measurement repositories suitable for SPC, as well as to perform measurements appropriately for it [1]. One of the components of this strategy is the Software Measurement Ontology [1, 2]. This ontology provides the conceptualization involved in the software measurement domain, including traditional and high maturity aspects. It establishes a common vocabulary to software measurement and provides useful knowledge about this domain. The IESMR was built based on the Software Measurement Ontology.

4. THE INSTRUMENT

According to the proposed Instrument for Evaluating the Suitability of a Measurement Repository to SPC (IESMR), the evaluation of a measurement repository consists of the evaluation of four items: (i) the Measurement Plan, (ii) the

measurement repository structure, (iii) the defined measures, and (iv) the data collected for these measures.

It is important to notice that, according to the approach for software process improvement in high maturity levels, only processes that are critical to the organizational goals must be submitted to SPC. So, it is desirable that the organization identifies these processes before carrying out the evaluation, in order to avoid evaluating measures that are not related to these processes.

Figure 2 shows an overview of the IESMR.

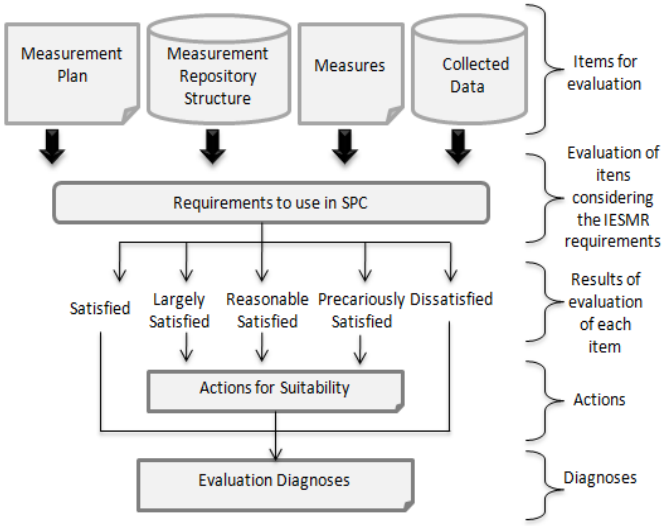

Figure 2. Overview of the IESMR

Each item considered by the instrument is evaluated by a checklist composed by a set of requirements. The evaluation of an item against each requirement can produce one of the following results:

(i) Satisfied: the item satisfies totally the requirement and any corrective action is necessary.

(ii) Largely Satisfied, Reasonably Satisfied or Precariously Satisfied: the item does not completely satisfy the requirement, but it is possible to perform actions to adapt it in order to satisfy the requirement and, consequently, to allow the use of the evaluated item in the SPC. The level of satisfaction (*largely, reasonably or precariously*) is related to the required effort to perform the actions that will make the item meets the requirement. The more effort is necessary, the less the satisfaction level will be.

(iii) Dissatisfied: the item does not satisfy the requirement and there are no possible actions to adapt it for being used in the SPC. Thus the item should be discarded.

When the result of a requirement is *Largely Satisfied*, *Reasonably Satisfied* or *Precariously Satisfied*, *Actions for Suitability* are suggested. These actions are guidelines provided for correcting the item so that it could be used in SPC.

The results of an evaluation of a measurement repository are registered in a document called Evaluation Diagnosis. It includes, besides the detailed evaluation of each item, suggestions of actions for possible adaptation and the degree of suitability of the measurement repository to SPC, informed as a percentage.

4.1 IESMR Checklists

Tables 1 to 4 present the IESMR checklists. The checklists to evaluate the Measurement Plan (Table 1) and the measurement repository structure (Table 2) are applied only once during an evaluation of a measurement repository. Checklists to evaluate the measures (Table 3) and the collected data (Table 4) must be applied once for each measure evaluated. The possible answers for a requirement in a checklist are: S = Satisfied; LS = Largely Satisfied; RS = Reasonable Satisfied; PS = Precariously Satisfied; D = Dissatisfied; NE = It could not be evaluated.

Table 1. Checklist to evaluate the Measurement Plan.

Requirements	Evaluation					
1. The Organizational Measurement Plan is aligned to the organization goals.	S	LS	RS	PS	D	NE
1.1 The organization business goals that are relevant to measurement are registered in the Measurement Plan.	S	LS	RS	PS	D	NE
1.2 The measurement goals are registered in the Measurement Plan and they are correctly associated to the organization business goals.	S	LS	RS	PS	D	NE
1.3 The information needs for monitoring the measurement goals are identified.	S	LS	RS	PS	D	NE
1.4 The measures able to supply the useful information needs to monitor the measurement goals are identified and properly associated.	S	LS	RS	PS	D	NE

Table 2. Checklist to evaluate the Measurement Repository Structure.

Requirements	Evaluation					
1. The measurement repository is well structured and allows measures to be integrated to the processes and activities of the organization.	S	LS	RS	PS	D	NE
1.1 The measurement repository structure allows the defined measures to be related to processes and activities of the organization in which the measurement must be carried out.	S	LS	RS	PS	D	NE
1.2 The measurement repository is unique or it is composed by several correctly integrated sources.	S	LS	RS	PS	D	NE
2. The projects are characterized satisfactorily.	S	LS	RS	PS	D	NE
3. A mechanism for identifying similarity between projects is established.	S	LS	RS	PS	D	NE
4. It is possible to identify the version of the processes executed in the projects.	S	LS	RS	PS	D	NE
5. It is possible to store and to retrieve the context information of the collected measures.	S	LS	RS	PS	D	NE
For each collected measure, it is possible to store and to retrieve:						
5.1 Measurement moment (process and activity in which the measurement was carried out)	S	LS	RS	PS	D	NE
5.2 Measurement conditions (relevant data about the execution of the process or project at the moment of the measure collection).	S	LS	RS	PS	D	NE
5.3 Performer of the measurement.	S	LS	RS	PS	D	NE
5.4 Project in which the measure was collected.	S	LS	RS	PS	D	NE
5.5 Characteristics of the project in which the measure was collected.	S	LS	RS	PS	D	NE

Table 3. Checklist to evaluate Measures.

Item: Measures **Measure Evaluated:_____**

Requirements	Evaluation					
1. The operational definition of the measure is correct and satisfactory.	S	LS	RS	PS	D	NE
The operational definition of the measure correctly includes:						
1.1 Measure Definition	S	LS	RS	PS	D	NE
1.2 Measured Entity	S	LS	RS	PS	D	NE
1.3 Measured Property	S	LS	RS	PS	D	NE
1.4 Measure Unity	S	LS	RS	PS	D	NE
1.5 Scale Type	S	LS	RS	PS	D	NE
1.6 Scale Values	S	LS	RS	PS	D	NE
1.7 Expected Interval for the Data	S	LS	RS	PS	D	NE
1.8 Formula(s) (if applicable)	S	LS	RS	PS	D	NE
1.9 Precise Description of the Measurement Procedure	S	LS	RS	PS	D	NE
1.10 Responsible for the Measurement	S	LS	RS	PS	D	NE
1.11 Measurement Moment	S	LS	RS	PS	D	NE
1.12 Measurement Periodicity	S	LS	RS	PS	D	NE
1.13Precise Description of the Measurement Analysis Procedure (if essential)	S	LS	RS	PS	D	NE
1.14 Responsible for the Analysis (if essential)	S	LS	RS	PS	D	NE
1.15 Analysis Moment (if essential)	S	LS	RS	PS	D	NE
1.16 Analysis Periodicity (if essential)	S	LS	RS	PS	D	NE
2. The measure is aligned to the projects goals and/or organization goals.	S	LS	RS	PS	D	NE
The measure is associated to:						
2.1 Organization goals.	S	LS	RS	PS	D	NE
2.2 Project goals.	S	LS	RS	PS	D	NE
3. The measure analysis results are relevant for making decisions.	S				D	NE
4. The measure analysis results are useful to the process improvement.	S				D	NE
5. The measure is related to the performance of a process.	S				D	NE
6. The measure is related to a critical process.	S				D	NE
7. The measure is associated to an activity or process that produces measurable items.	S				D	NE
8. The related measures are defined.	S	LS	RS	PS	D	NE
9. The related measures are valid.	S	LS	RS	PS	D	NE
10. The measure has appropriate granularity level.	S	LS	RS	PS	D	NE
11. It is possible to normalize the measure (if applicable).	S	LS	RS	PS	D	NE
12. The measure is correctly normalized (if applicable).	S	LS	RS	PS	D	NE
13. The data grouping criteria for measure analysis are defined.	S	LS	RS	PS	D	NE
14. The measure does not consider aggregated data.	S	LS	RS	PS	D	NE

Table 4. Checklist to evaluate the Data Collected to the Measures.

Item: Collected Data **Measure Evaluated:_____**

Requirements	Evaluation					
1. The collected data for measure are known and have accessible location.	S	LS	RS	PS	D	NE
2. There is sufficient volume of collected data.	S				D	NE
3. There is no lost data for the measure or the quantity of lost data does not compromise the analysis.	S	LS	RS	PS	D	NE
4. The collected data are precise.	S	LS	RS	PS	D	NE
5. The collected data are consistent.	S	LS	RS	PS	D	NE
Characteristics of the collected data:	S	LS	RS	PS	D	NE
5.1 The data were collected at the same moment of the execution of the process along the projects.	S	LS	RS	PS	D	NE
5.2 The data were collected under the same conditions.	S	LS	RS	PS	D	NE
5.3 The data compose relatively homogeneous groups.	S	LS	RS	PS	D	NE
6. The data that describe the collection context of the measure are stored.	S	LS	RS	PS	D	NE
The following items are stored:						
6.1 Measurement moment (process and activity in which the measurement was carried out).	S	LS	RS	PS	D	NE
6.2 Measurement conditions (relevant data about the execution of the process or project at the moment of the collection of the measure).	S	LS	RS	PS	D	NE
6.3 Performer of the measurement.	S	LS	RS	PS	D	NE
6.4 Project in which the measure was collected.	S	LS	RS	PS	D	NE
6.5 Characteristics of the project in which the measure was collected.	S	LS	RS	PS	D	NE

4.2 Evaluating the Fulfillment of the IESMR Requirements

As shown in the previous tables, most of the requirements of the IESMR can produce one of the following results: *Satisfied*, *Largely Satisfied*, *Reasonably Satisfied*, *Precariously Satisfied* or *Dissatisfied.* Only a few ones (requirements 3 to 7 in Table 3, and requirement 2 in Table 4) have as possible results *Satisfied* or *Dissatisfied*. In these cases, there is no possibility of partial fulfillment of the corresponding requirement, since there are no possible actions for correcting it. For example, the requirement "The measure is related to the performance of a process" (requirement 5 of Table 3) is satisfied or not. If a measure does not describe the performance of a process, there is nothing that can be done to use it in the SPC.

In order to guide the evaluation, for each requirement, we provided a description of what a possible answer means. For instance, concerning the requirement "The projects are characterized satisfactorily" (requirement 2 of Table 2), the following descriptions are provided:

- *Satisfied*: The project characterization is explicit. That is, there is a characterization schema formally defined and implemented in the measurement repository structure, taking basis on relevant criteria that enable the organization to identify the profiles of projects that it develops. The subsets composed by projects with the same profile (that is, projects of which criteria of characterization have the same values) are homogeneous.
- *Largely Satisfied*: The project characterization is explicit, but it requires some additional criteria. These criteria can be identified from data of projects stored in the measurement repository, interviews with projects members and projects documents.
- *Reasonably Satisfied*: The project characterization is explicit, but it requires many additional criteria. These criteria can be identified from data of projects stored in the measurement repository, interviews with projects members and projects documents.
- *Precariously Satisfied:* The project characterization is implicit. That is, there is no formal characterization of the projects, but it is possible to identify a characterization by analyzing data of projects stored in the measurement repository, carrying out interviews with projects members and projects documents.
- *Dissatisfied:* There is no explicit characterization, or it is inadequate, and it is not possible to identify criteria to determine a satisfactory characterization analyzing the data of projects stored in measurement repository, carrying out interviews with projects members or analyzing projects documents.

Note: Some requirements are decomposed into sub-requirements (e.g. requirement 6 of Table 4). The result of their evaluation is obtained from an aggregation of the results of the evaluation of their sub-requirements, as discussed in subsection 4.4.

4.3 Actions for Suitability to the IESMR Requirements

If the result of the evaluation of a requirement is *Largely Satisfied, Reasonably Satisfied or Precariously Satisfied*, actions for suitability are suggested, aiming to support the organization to change the evaluated item to fulfill the requirement. So, for each requirement, we identified potential problems and actions for suitability. For instance, to the requirement 2 of Table 2, discussed in the previous subsection, the following potential problems and actions for suitability were identified:

- **Problem I**: *The projects have an implicit characterization in the measurement repository.*

  Actions for Suitability: (a) Define explicitly the projects characterization by analyzing data of projects that are stored in the measurement repository. For this, it is necessary to identify the data that describe characteristics of the executed projects, such as size, constraints, team features, technologies, development paradigm, application domain, project type etc. (b) Restructure the measurement repository making the identified characterization criteria explicit in classes and properties. (c) Record the characterization data of the projects in the restructured measurement repository appropriately.

- **Problem II**: *The projects do not have characterization (implicit or explicit) in the measurement repository.*

  Actions for Suitability: Set up a characterization based on the analysis of documents and interviews with people related to the projects. For example, project managers can provide relevant information about characteristics of the executed projects, such as technologies, development paradigm, project type, considered constraints and so on. This action must be followed by the actions (b) and (c) cited above.

- **Problem III:** *The explicit characterization of the projects requires additional criteria.*

  Actions for Suitability: Refine the project characterization, identifying new criteria. This can be realized carrying out the actions cited on problems I and II.

4.4 Degree of Suitability of a Measurement Repository to SPC

The evaluation results from each requirement in a checklist determine the suitability of the corresponding item to SPC. For example, the results from requirements 1 to 5 in Table 2 are used to determine the suitability of the measurement repository structure to SPC. Based on the suitability of the four items considered in the IESMR (Measurement Plan, measurement repository structure, measures and collected data), the suitability of the measurement repository to SPC is determined. Organizations can use the degree of suitability of their measurement repositories as basis for deciding if it is better to fix a repository or to develop a new one.

However, a new problem arose: How could be the individual results from the checklist requirements used for computing the suitability of a measurement repository as a whole? To solve this problem, we used principles of Fuzzy Logic and Fuzzy Set Theory in order to, from the subjective individual results from the checklists requirements, obtain a single value that represents the degree of suitability of the measurement repository to SPC. For this we followed the steps of fuzzy reasoning: fuzzification, fuzzy inference and defuzzification [15].

Initially we carried out the fuzzification of the input and output values. That is, the linguistic variables related to inputs and outputs were identified, their linguistic terms were determined and the corresponding membership functions were defined.

Input values are the results of the evaluation of each IESMR requirement and are represented by the linguistic terms: *Satisfied, Largely Satisfied, Reasonably Satisfied, Precariously Satisfied* and *Dissatisfied*. Output values are the possible levels of suitability of an item to SPC and are represented by the linguistic terms *Suitable, Largely Suitable, Reasonably Suitable, Precariously Suitable* or *Unsuitable*. For each input and output value we determined its membership function. A membership function indicates the degree of pertinence of an element to a fuzzy set. We defined the membership functions as triangular functions represented by

(a, b, c). a and *c* determine the points where the degree of pertinence is zero and *b* indicates the point where the degree of pertinence is maximum. For example, for the linguistic term *Largely Satisfied*, we determined the membership function (2.0, 3.0, 4.0) and for the linguistic term *Largely Suitable* we determined the membership function (50, 75, 100). Due to space limitations, as an example, we present here only the membership functions of one input value and one output value.

After the fuzzification, we defined the rules for the fuzzy inference. Considering the large number of input variables and the variety of possible combinations, we concluded that defining IF-THEN rules was inappropriate to cover all possible combinations of inputs. We decide then to use the $OWA_{AVERAGE}$ operator (Ordered Weighted Average) [23] to carry out the aggregation of the fuzzy input variables into a single fuzzy output value. For this, for each input linguistic term, we associated a numeric value, as follows: 4 - *Satisfied*, 3 - *Largely Satisfied*, 2 - *Reasonably Satisfied*, 1 - *Precariously Satisfied*, 0 - *Dissatisfied*. The aggregate resulting from several input linguistic terms (remember that an input linguistic term is the result of an evaluation of a IESMR requirement) is the average of their numeric values.

To evaluate each item, in addition to the use of $OWA_{AVERAGE}$, we established the following rule *R*: *If the note corresponding to an input linguistic term is 0, then the aggregation is also 0*. This rule means that if any requirement of an item is evaluated as *Dissatisfied*, the item will be considered *Unsuitable* to SPC. This rule was defined because, when a requirement is evaluated as *Dissatisfied*, there are no possible actions for suitability to fit the evaluated item in SPC, so the item must be discarded, even if it satisfies the other requirements. If an item has a requirement evaluated as *Dissatisfied*, it can be rebuilt. However, it cannot be fixed. Therefore, not fulfilling a single requirement makes the item *Unsuitable* to SPC.

Since some requirements are decomposed into sub-requirements (for example, requirement 1 in Table 2 is decomposed into sub-requirements 1.1 and 1.2), we should first determine the aggregated result of the composed requirements, using $OWA_{AVERAGE}$ and the rule *R*, as shown in Figure 3.

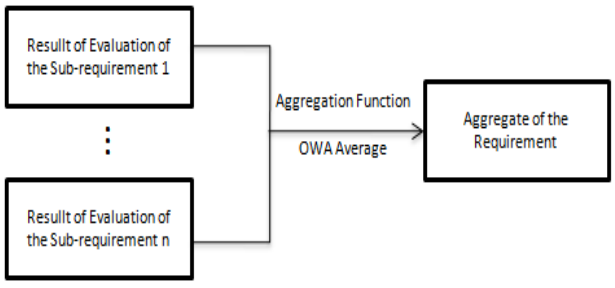

Figure 3. Evaluation of composed requirements.

Once the results of the composed requirements have been determined, the results from the requirements are aggregated (also applying $OWA_{AVERAGE}$ and the rule *R*) to get the aggregated value of each item. Finally, the suitability of the measurement repository as a whole is computed as the average of the aggregated scores of the items. Figure 4 shows the procedure adopted for determining the suitability of a measurement repository.

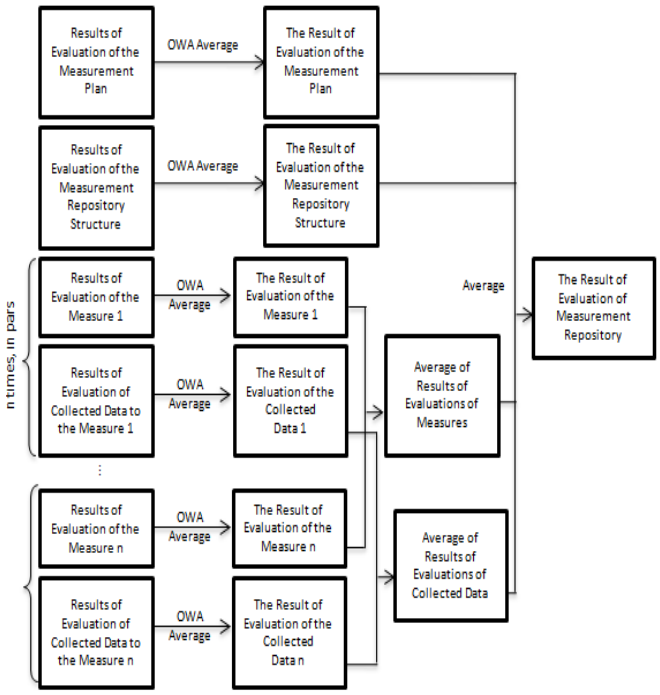

Figure 4. Procedure adopted to obtain the value of the suitability of a measurement repository.

After obtaining a unique fuzzy output (Result of the Evaluation of the Measurement Repository in Figure 4), we applied the Centroid Defuzzification method [18] to determine the corresponding crisp value (that is, the single value) that represents the degree of suitability of the measurement repository. It is worth pointing out that it is also possible to obtain the degree of suitability of each item, individually, performing the defuzzification of the values of their evaluation results.

The implementation of the principles of Fuzzy Logic and Fuzzy Sets to determine the suitability of a measurement repository was carried out with the support of MATLAB R2007a[1].

It is important to emphasize that the solution adopted is an initial fuzzy solution and it is not considered a fuzzy optimal solution. The proposed solution will be refined in future opportunity, considering, for example, weights for the requirements, assigned by experts, and calibrations considering results of practical experiences.

5. Practical Experiences using IESMR

At the end of 2008, the software organization *X* (for the sake of confidentiality we omitted the real name), evaluated CMMI level 2 in the same year, starts to implement the practices required by CMMI level 3, planning to achieve this level in 2009. Although it is not a requirement for level 3, they want to anticipate the adjustment of measures and data to implement the SPC in the next future (to fulfill the CMMI

[1] Available in http://www.mathworks.com.

level 4 requirements). Thus, during the level 3 implementation, we carried out an evaluation of the measurement repository of this organization.

First, we evaluated the Measurement Plan. It was developed relating business goals, measurement goals, information needs and measures, using the Goal Question Metric paradigm [3]. The Measurement Plan was composed by 4 business goals and 22 measurement goals, of which 20 dealt with process monitoring. The plan was stored in electronic spreadsheets. During the evaluation of the Measurement Plan, we found problems in some measurement goals, information needs and measures, as well as in the relationship between them. Thus, applying the actions for suitability present in IESMR, we suggested that they revise the Measurement Plan in order to delete, add or change measurement goals, information needs and measures and to correct the relationships between these elements. Figure 5 shows the checklist fulfilled in the evaluation of the Measurement Plan. In the sequel, Table 5 presents a fragment of the document containing some of the observations made during the evaluation and actions for suitability suggested.

Item: Measurement Plan of the organization X

Requirements	Evaluation					
1. The Organizational Measurement Plan is aligned to the organization goals.	S	LS	RS	PS	D	NE
1.1 The organization business goals that are relevant to measurement are registered in the Measurement Plan.	S	LS	RS	PS	D	NE
1.2 The measurement goals are registered in the Measurement Plan and they are correctly associated to the organization business goals.	S	LS	RS	PS	D	NE
1.3 The information needs for monitoring the measurement goals are identified.	S	LS	RS	PS	D	NE
1.4 The measures able to supply the useful information needs to monitor the measurement goals are identified and properly associated.	S	LS	RS	PS	D	NE

Figure 5. Checklist fulfilled in the evaluation of the Measurement Plan of the organization X.

The next item evaluated was the measurement repository structure. The measurement repository had 25 tables. Its evaluation revealed some problems that we considered critical for implementing SPC, mainly its failure to identify the versions of the processes executed in the projects, the poor projects characterization (projects were classified only by type and paradigm) and the inability to store contextual information for the data collected to the measures. Figure 6 presents a fragment of the checklist fulfilled during the evaluation of the measurement repository structure and some observations recorded. The actions for suitability are not showed in Figure 6, only some considerations made during the evaluation.

Table 5. Fragment of observations made during the evaluation of the Measurement Plan and actions for suitability suggested.

Business Goal	BG1: Increasing the current client level of satisfaction.
Measurement Goal	MG1: Monitoring the Planning Project process.
Question	Q1: What is the precision of the schedule estimates in development software projects?
Measures	ETP (Estimated Time of Project) TRP (Real Time of Project)
Considerations made during the evaluation	a) Looking at the spreadsheet, we noted that the measures ETP and TRP are defined but they are not related to Q1. b) There are no measures for calculating the precision of schedule estimates.
Actions for Suitability Suggested	a) The measures related to the schedule of the development phases are defined but are not related to these objectives and question. So, it is necessary to make this association for including these measures in the list of measures related to Q1-MG1-BG1. b) Creating schedule measures to some macro-activities and activities (look at the critical processes). c) Creating measures to calculate the precision of schedule estimates (by project, phase and in some cases by macro-activity and activity).

Next, we evaluate the measures. First, we evaluated the template for measure operational definitions used by the organization. It included the following fields: name, description, mnemonic, basic value, upper limit, lower limit, equation for calculating, measure unit, measurement procedure, analysis procedure, responsible for the measurement, responsible for the analysis, measured entity, measurement periodicity, measurement frequency and the flags: active, mandatory, atomic and automatic. According to IESMR, we suggested them to add the following information: measured property, measurement moment, analysis moment, analysis periodicity, analysis moment, type of scale and scale values. We also suggested them to exclude the field measurement frequency, which was redundant.

After evaluating the measure operational definition template, we carried out the individual evaluations of the defined measures. The main problems found were: ambiguous, incomplete or inconsistent operational definitions, absence of related measures and inappropriate granularity. In order to solve the identified problems, actions for suitability were suggested for defining new measures, redefining some and correcting others. These suggestions were derived from a guideline which we proposed taking basis on the Software Measurement Ontology [1, 2] to establish measure operational definitions suitable to SPC. This guideline is presented in Table 6.

Figure 6. Fragment of the checklist of evaluation of the Measurement Repository Structure of the organization X.

Item: Measurement Repository Structure of the organization X

Requisitos	Avaliação
1. The measurement repository is well structured and allows measures to be integrated to the processes and activities of the organization.	S **LS** RS PS D NE
	Considerations: The database is unique, but it is necessary to review its structure because there is important information that can not be stored and retrieved. If there are other information systems in the organization which contain the information, restructuring and integration must be carried out.
1.1 The measurement repository structure allows the defined measures to be related to the processes and activities of the organization in which the measurement must be carried out.	S LS **RS** PS D NE
	Considerations: The processes are not stored explicitly. Apparently it is possible to identify activities that compose each process and to adjust the repository to store explicitly the processes and their activities.
1.2 The measurement repository is one or is composed by several correctly integrated sources.	**S** LS RS PS D NE
	Considerations: In this evaluation we considered that the measurement repository is composed only by the evaluated structure, since no other was mentioned.

Table 6. Guideline to elaborate a Measure Operational Definition.

The operational definition of a measure must include a description of all the necessary information for consistently collecting and analyzing the measure. Complete and precise operational definitions contribute to consistent measurements. An incomplete, ambiguous or weakly documented operational definition makes it possible that different people understand the measure in different ways and, consequently, they can collect invalid data, carry out measurements that cannot be compared, or perform incorrect analyses, which makes the measurement inconsistent and inefficient. An operational definition of measure must include:	
Name	Measure name.
Definition	A brief description of the measure.
Mnemonic	Acronym used to identify the measure.
Measured Entity	Entity that the measure measures. Examples: organization, project, process, activity, human resource, hardware resource, software resource, among others.

Table 6 (cont). Guideline to elaborate a Measure Operational Definition.

Measured Property	Property of the measured entity that is quantified by the measure. Examples: size, costs, defects, effort etc.
Measure Unity	Measure unity in which the measure is expressed. Examples: person/month, function points, dollar etc.
Scale Type	Nature of the values that can be attributed to the measure. Examples: nominal scale, interval scale, ordinal scale, absolute scale and ratio scale.
Scale Values	Values that can be attributed to the measure. Examples: real positive numbers, {high, middle, low} etc.
Expected Interval for Data	Limits of scale values, defined in accordance with historical data or with established purpose. Example: [0, 10].
Measurement Procedure	Description of the procedure that must be carried out to collect values for the measure. The description of the measurement procedure must be clear, complete and unambiguous.
Measure Formula of Calculation	Formula used to calculate the value attributed to the measure, considering its relation with other measures and/or other values. Example: schedule adherence = real time/estimated time.
Responsible for Measurement	Function fulfilled by the human resource responsible by the collection for the measure. It is important that the responsible one for measurement be a direct source of the information collected in measurement. Examples: systems analyst, programmer, project manager, etc.
Measurement Moment	Moment in which measurement must be carried out. The measurement moment must be an activity of the process defined for the projects or an activity of an organizational process. Examples: at the "Accord the Project Requirements Specification" activity, at the "Perform Unity Tests" activity, among others.
Measurement Periodicity	Frequency of collection of data for the measure. Examples: daily, weekly, monthly, once for phase, once for project, once for activity marked as Measurement Moment etc. It is essential to ensure consistency between the measurement periodicity and the measurement moment.
Measurement Analysis Procedure	Description of the procedure that must be carried out to represent and analyze the data collected for the measure, including addition to the procedure itself, the analytical tools to be used (eg, histogram, control chart XmR etc). The description of the measurement analysis procedure must be clear, complete and unambiguous. A measurement analysis procedure can be based on criteria (eg, using a target as a reference) and, if so, the decision criteria considered (including its premises and conclusions) must be clearly established. Measures that are not analyzed individually do not need measurement analysis procedure. Example: if the measure "number of modified requirements" is analyzed only when it is used in the composition of other measures (for example, requirements alteration ratio), there is no need to establish an analysis procedure for it.
Responsible for Measurement Analysis	Function fulfilled by the human resource responsible by the analysis of the data collected for the measure. It is important that the person responsible for measurement analysis is able to apply the analysis procedure and has the organizational knowledge required to the interpretation correct of the data aiming to provide information to support decision making.Examples: project manager, quality manager, etc.
Measurement Analysis Moment	Moment in which the analysis of data collected for the measure must be carried out. The measurement analysis moment must be an activity of the process defined for the projects or an activity of an organizational process. For example, at monitoring activities of the project.
Measurement Analysis Periodicity	Frequency of analysis of the data collected for the measure. Examples: daily, weekly, monthly, once for phase, once for project, once for activity marked as Measurement Analysis Moment etc. It is essential to ensure consistency between the measurement analysis periodicity and the measurement analysis moment.

Finally, for each measure, we evaluated its collected data. In general, the main problems found were the following: (i) the volume of the collected data was insufficient; and (ii) the data needed to describe the context of the collection were not recorded. The first problem is directly related to inappropriate granularity of the defined measures. Measures related to entities that occur once in a project (such as the project itself or a specific phase of the development) were collected only once by project and it is not enough to SPC. It is important to notice that lost data and non collected data can also lead to an insufficient volume. The second problem can make the data useless, because if the context of collection is not known, the data cannot be grouped correctly for analysis.

After receiving the results of the evaluation, the organization *X* performed the suggested actions for suitability. In 2009, the organization was successfully evaluated as a CMMI level 3 organization. After that, the behavior of the critical processes was analyzed using SPC techniques. The results showed that there are still some changes required for stabilizing the processes. Also we notice that some actions for suitability suggested in the evaluation of the measurement repository, in fact, were not performed by the organization.

Currently, the organization *X* is working to achieve a high maturity level. The assessment is planned to occur in this year. Some efforts being done include a review of the actions for suitability suggested in the evaluation of the measurement repository, aiming to identify which actions were not carried out before. In a brief commentary, the quality manager of the organization *X* said that starting the preparation to SPC in previous maturity levels will help them to spend less time to perform the statistical control of its critical processes and thus they can devote more time and effort to other practices required in high maturity levels, such as process componentization. As a consequence, he expects that the time for achieving a high maturity level be smaller than the time normally required if the organization had not previously carried out adjustments in its measurement repository.

6. Related Works

There are several works highlighting the importance of measures and collected data are appropriate for the SPC, in order to be used in this context [8, 12, 13, 19, 20, 22]. However, there are few works treating specifically the evaluation of measures and data for this purpose.

The works of Tarhan and Demirors [19, 20] include a proposal to evaluate measures considering its usefulness in SPC. In fact, the work proposes an approach to select processes to SPC. Among other relevant criteria for choosing the most appropriate processes, the authors point to the existence of appropriate measures associated with the processes. To determine if a measure is useful, they define a set of attributes to evaluate measures and data. However, since the focus of their approach is the selection of processes to SPC, the approach is limited concerning measure evaluation. The authors themselves affirm that the study of the usefulness of measures considering only the attributes defined by them is not enough to select the most appropriate measures in SPC [19].

7. Final Considerations

The growing interest of software organizations in achieving the highest maturity levels has revealed some difficulties of these levels, especially problems with implementing the statistical process control.

The absence of appropriate measures and data for the application of SPC techniques, which allow organizations to understand and to improve the processes behavior, has been an obstacle to achieve the highest maturity levels. In this context, this paper presented an Instrument for Evaluation of the Suitability of a Measurement Repository to SPC (IESMR).

IEMSR was developed to support software organizations that desire to carry out SPC and that have measurement repositories with data collected in executed projects. IEMSR supports evaluating if the measurement repository of an organization is suitable to SPC and, if not, to carry out the adjustments, when it is possible.

So far, IESMR was used to evaluate the measurement repositories of three organizations. The results obtained from these experiences allowed us to conclude that the evaluation and adjustment of the measurement repository before performing the SPC help organizations in preparing themselves for starting the use of statistical techniques. It avoids that them to expend effort in implementing the SPC and, when they notice the unsuitability of the measures, they need to interrupt or abort the SPC implementation. Moreover, this approach ensures that the information provided by SPC is really useful, since the measures and data used are appropriate.

Currently, IESMR does not have a specific software to support its application. It is composed mainly by spreadsheets. To support its application a prototype of a tool was developed and the next step is implementing a tool to support the use of IESMR. New practical experiences are planned too.

8. ACKNOWLEDGMENTS

Our acknowledgments to the organizations that participated in the practical experiences of IESMR, and to Professor Geraldo Xexéo for his collaboration on defining the fuzzy solution.

9. REFERENCES

[1] Barcellos, M. P., Rocha, A. R. and Falbo, R. A. 2009. An Ontology-based Approach for Software Measurement and Suitability Measures Basis Evaluation to Apply Statistical Software Process Controle in High Maturity Organizations. In Proceedings of the PhD Colloquium of the 28th International Conference on Conceptual Modeling – ER 2009, Gramado, Brazil.

[2] Barcellos, M. P., Falbo, R. A. and DAL MORO, R. 2010. A Well-Founded Software Measurement Ontology", In Proceedings of the 6th International Conference on Formal Ontology in Information Systems (FOIS 2010), Toronto, Canada.

[3] Basili, V. R., Rombach, H. D. and Caldiera, G. 1994. Goal Question Metric Paradigm. Encyclopedia of Software Engineering, 2 Volume Set, John Wiley & Sons, Inc.

[4] Boffoli, N., Bruno, G., Caivano, D. and Mastelloni, G. 2008. Statistical Process Control for Software: a Systematic Approach. In Proceedings of the 2008 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, Kaiserslautern, Germany, 327-329.

[5] Card, D., Domzalski, K., Davies, G. 2008. Making Statistics Part of Decision Making in an Engineering Organization. IEEE Software, 25(3), 37-47.

[6] Card, D. N. 2004. Statistical Techniques for Software Engineering Practice. In Proceedings of the 26th International Conference on Software Engineering, Scotland, UK, 722-723.

[7] Chrissis, M. B., Konrad, M. and Shrum, S. 2006. CMMI (Second Edition): Guidelines for Process Integration and Product Improvement. Addison Wesley.

[8] Curtis, B., Reifer, D., Seshagiri, G.V., Hirmanpour, I. and Keeni, G. 2008. The Case for Quantitative Process Management. IEEE Software, 25(3), 24-28.

[9] Florac, W. A. and Carleton, A. D. 1999. Measuring the Software Process: Statistical Process Control for Software Process Improvement. Addison Wesley.

[10] Fuggetta, A. 2000. Software Process: A Roadmap. In Proceedings of the Conference on the Future of Software Engineering, 25-34.

[11] García, F., Piattini, M., Ruiz, F., Canfora, G. and Visaggio, C. A. 2004. FMESP: Framework for the Modeling and Evaluation of Software Process. In Proceedings of the 2004 Workshop on Quantitative Techniques for Software Agile Process, Newport Beach, California, 5-13.

[12] Gou, L., Wang, Q., Yuan, J., Yang, Y., Li, M. and Jiang, N. 2009. Quantitative Defects Management in Interative Development with BiDefect. Software Process Improvement and Practice, 14(4), 227-241.

[13] Kitchenhan, B., Kutay, C., Jeffery, R. and Connaughton, C. 2007. Misleading Metrics and Unsound Analyses. IEEE Software, 24(2), 73 – 78.

[14] Kitchenhan, B. Kutay, C., Jeffery, R. and Connaughton. C. 2006. Lessons Learnet from the Analysis of Large-scale Corporate Databases. In Proceedings of the 28th International Conference on Software Engineering, Shanghai, China, 439-444.

[15] Klir, G. J., Folger, T. A. 1998. Fuzzy Sets, Uncertainty and Information. Prentice Hall, New Jersey.

[16] Montoni, M., Rocha, A,R.C and Weber, K.C. 2009. MPS.BR: A Successful Program for Software Process Improvement in Brazil. Software Process Improvement and Practice, 14, 289-300.

[17] Lantzy, M. A. 1992. Application of Statistical Process Control to the Software Process. In Proceedings of the 9th Washington Ada Symposium on Empowering Software Users and Developers, ACM Press, 113-123.

[18] Sivanandam, S. N,. Deepa, S. N. and Sumathi, S. 2007. Introduction to Fuzzy Logic using MATLAB. Springer-Verlag Berlin Heidelberg, German.

[19] Tarhan, A. and Demirors, O. 2006. Investigating Suitability of Software Process and Metrics for Statistical Process Control. Lectures Notes in Computer Science, 4257, pp. 88-99.

[20] Tarhan, A. and Demirors, O. 2008. Assessment of Software Process and Metrics to Support Quantitative Understanding Lecture Notes in Computer Science, 4895, 102-113.

[21] Wang, Q., Gou, L., Jiang, N., Che, M., Zhang, R., Yang, Y. and Li, M. 2007. An Empirical Study on Establishing Quantitative Management Model for Testing Process. Lecture Notes in Computer Science, 4470, 233-245.

[22] Wheeler, D. J. and Poling, R. S. 1998. Building Continual Improvement: A Guide for Business. SPC Press.

[23] Yager, R. R. 1988. On Ordered Weighted Averagingh Aaggregation Operators in Multicriteria Decision Making. IEEE Transactions on Systems, 18(1), 183-190.