Measuring Software Process: A Systematic Mapping Study

AYMAN MEIDAN, JULIÁN A. GARCÍA-GARCÍA, ISABEL RAMOS,
and MARÍA JOSÉ ESCALONA, University of Seville

Context: Measurement is essential to reach predictable performance and high capability processes. It provides support for better understanding, evaluation, management, and control of the development process and project, as well as the resulting product. It also enables organizations to improve and predict its process's performance, which places organizations in better positions to make appropriate decisions. Objective: This study aims to understand the measurement of the software development process, to identify studies, create a classification scheme based on the identified studies, and then to map such studies into the scheme to answer the research questions. Method: Systematic mapping is the selected research methodology for this study. Results: A total of 462 studies are included and classified into four topics with respect to their focus and into three groups based on the publishing date. Five abstractions and 64 attributes were identified, 25 methods/models and 17 contexts were distinguished. Conclusion: capability and performance were the most measured process attributes, while effort and performance were the most measured project attributes. Goal Question Metric and Capability Maturity Model Integration were the main methods and models used in the studies, whereas agile/lean development and small/medium-size enterprise were the most frequently identified research contexts.

1 INTRODUCTION

Software development is considered to be composed of three essential components: products, processes, and resources (Fenton 1991). Developing software is a long, costly, and complex process. The outcome of this process is not only the final product, but the production of many intermediate and supplementary artifacts during the development endeavor. The quality of this development process significantly impacts the quality of the resulting product (Kitchenham and Pfleeger 1996; Fuggetta 2000; Cugola and Ghezzi 1998).

Measuring the software development process and its outcomes is the only way to gain knowledge about them. In addition, the obtained measurements could be used in models for prediction

This research has been supported by MeGUS Project No. TIN2013-46928-C3-3-R, Pololas Project No. TIN2016-76956-C3-2-R, and SoftPLM Network Project No. TIN2015-71938-REDT of the Spanish Ministry of Economy and Competitiveness.
Authors' addresses: A. Meidan, J. A. García-García, I. Ramos, and M. J. Escalona, University of Seville, Avenida Reina Mercedes s/n, Seville, 41012, Spain; emails: ayman.meidan@gmail.com, julian.garcia@iwt2.org, {iramos, mjescalona}@us.es.

purposes (Lennselius et al. 1987). Software process measurement provides support for better understanding, evaluation, and control of the development process, project, and the resulting product (Ebert et al. 2007). Measurement also enables organizations to have insight into its processes and improve and predict its quality and performance, which gives organizations a better position to make appropriate and informed decisions as early as possible during the development process (García et al. 2006; Abreu Fernando Brito and Carapuça 1994).

In the last decades, software development process has evolved to meet the market needs and to keep abreast of modern technologies and infrastructures that have influenced the product development and its use. These changes in the development processes have increased the importance of the measurement (Bourgault et al. 2002) and caused changes in the measurement process and the used measures (Tihinen et al. 2012).

For instance, cloud computing allowed merging software development, deployment, and operation in what is known as DevOps. Measurement is one of the four DevOps perspectives (collaboration culture, automation, measurement, and sharing) (Bang et al. 2013). In this context, measurement promotes the communication and the common understanding between development and operations. On the other side, today's software is increasingly developed by teams working in different geographic locations, time zones, and cultures. Management of these kinds of projects is more challenging and complicated than traditional one-site development. The measurement is an important element for the success of these development projects (Tihinen et al. 2012).

These evolutions in the development process, technologies, and infrastructures create new challenges and difficulties for the measurement in terms of data collection, storage, analysis, interpretation, and decision-making based on the measurement results. These challenges and difficulties emphasize the importance of further research in this area.

There is a wide variety of software measurement articles available (see the primary study: S156), containing numerous proposals (e.g., methods, processes, models, practices and tools), empirical studies, and experiences that cover different entities related to software development (e.g., process, product, project, or resource). All composed of diverse attributes, and different measurements for the same attributes, the evaluation and classification of them is still a challenge for researchers and practitioners (Habra et al. 2008). These facts make the research in this field a challenging and confusing task.

For this reason, it is essential to choose a rigorous and well-defined method before searching through the current literature related to this field. This article presents the results of a mapping study that was conducted to provide better insight into measurement in the context of software development processes by identifying, classifying, and structuring publications to create a complete updated view on the different subjects of the research in this field, as well as the changes and trends of the research focus over time. This study also tries to outline the main investigated entities and the different measures and methods proposed to characterize and predict their attributes, moreover, identifying the main publication channels where researchers and practitioners can find studies related to this area and highlighting the main research types, contributions (e.g., methods and models) found in the literature, its context, and the principal validation methods used in existing proposals. The results of this study will help researchers understand the measurement area, the trends of the research focus over time, and to discover the main existing proposals, methods, model, or techniques as well as the main contexts discussed in the literature. This allows the researcher to identify gaps and new investigation opportunities.

The article is structured as follows: Section 2 presents the background, and Section 3 analyzes related work. The research design of the systematic mapping is described in Section 4. Section 5 shows the mapping results, and Section 6 presents the validity threats. Section 7 presents a summary of findings. Section 8 discusses the implications for research and practice, and Section 9 states the final conclusions and future work.

2 BACKGROUND

2.1 Measurement

Measurement is the process that assigns values in accordance with certain rules to describe properties of entities (Fenton and Pfleeger 1996). In software engineering, the entity might be an object or an event (e.g., program or project). Entities are commonly classified into three categories: products, processes, and resources (Fenton 1991). An attribute is a measurable characteristic of an entity, e.g., quality of a program or cost of a project, and the experience of the development staff. Two types of attributes can be distinguished: internal and external (Shepperd 1995). Internal attributes are those that can be measured based on the entity itself, such as the elapsed time of the development process and the project size. However, the measurement of the external attributes, such as the product reliability and the resource productivity, depends on the measured entity and its environment.

Software measurement involves broad activities; the measurement process is composed of determining the required data, designing measurement methods and plans, applying the measurement methods, using tools and techniques to gather and analyze the data, and exploiting measurement results.

2.2 Software Process

Software engineering aims to ensure that software development produces the expected outcomes within the estimated cost and schedule (Sommerville 2004). Accomplishing this objective is the main concern of the software process engineering, through the provision of improved approaches and methodologies to enhance the development process conduct.

Defining and executing the software process as a sequence of work activities allows the measurement, control, and improvement of this process. When the execution consistently follows the process definition, the expected outcomes will be produced within the planned quality, cost, and time. This allows the prediction of the process performance and results, which keeps it under statistical control. When the process is under statistical control, every process enactment will produce approximately the same results. In this context, the only way to improve the process outcome is by improving the process itself. It is not possible to obtain sustained improvement until the process is under statistical control (Humphrey 1989; Humphrey 1988).

Measuring the software process is important to monitor, control, assess, manage, and improve it. Besides, it is also relevant to gain insight into the process to detect problems and identify risks early, in addition to facilitating communication and enabling objective planning, estimation, decision-making, and improvements (Jethani 2013).

There are two types of process measurement: in-process metrics (these measurements are carried out within the process stage) and post-activity metrics, which measures the effectiveness of meeting the customer expectations at the end of the process (Bhide 1990).

3 RELATED WORK

Gómez et al. (Gómez Oswaldo and García 2006; S93) carried out a Systematic Literature Review (SLR) about measurement in software engineering. Seventy-eight papers were reviewed to answer the following research questions: "What do we measure?"; "How do we measure?"; "When do we measure?" The criteria on whether or not to include the studies in the review dealt with presenting current and useful measurements. "Product" (79%), "Project" (12%), and "Process" (9%) were the most frequently measured entities, where the principal measured attributes were "Complexity" (19%), "Size" (16%), "Inheritance" (8%), "Defect" (7%), "Structuring" (7%), "Time" (5%), and "Others" (less than 4%). About 46% of the measures were validated using empirical methods, 26% were validated theoretically, and 28% used both (empirical and theoretical) validation methods. The

authors were able to answer the third question by mapping the identified measures to the waterfall lifecycle phases, where the percentage of product measurements in the different phases of the software lifecycle were: 42% during the design phase, 27% during development, 14% through the maintenance, 12% as testing, and 5% over the analysis phase.

Kitchenham [S156] conducted a mapping study to identify the most cited software measurement publications in the period of 2000 to 2005 (according to SCOPUS) and compared them with articles with less citation. Twenty-five papers were included. The author organized the papers based on different aspects as goal, topic, and research type. The study concludes that the journal articles have a higher impact than the conference papers, and it found that the most studied topics were Analysis (27%), Development (15%), Evaluation (24%), Framework (8%), Use (19%), and others (7%). Furthermore, Kitchenham suggests that the empirical approaches used by the software measurement researchers need to be improved to respond to useful empirical questions.

Unterkalmsteiner et al. [S306] carried out a SLR to study and assess the methods used to evaluate the results of the Software Process Improvement (SPI) activities. The authors have included 148 studies issued in the period of 1991 to 2008. These studies were categorized according to the SPI initiative, measurement methods, and the focus of the evaluation. The authors found seven different evaluation approaches. "Pre-Post Comparison" was the most-used approach. Quality, cost, and schedule were the most measured attributes. Project was the most-used measurement perspective. The SLR concluded that most evaluation strategies assess the short and medium impacts of the SPI initiatives, while the long-term measurement such as customer satisfaction and return on investment have a tendency to be less utilized.

The Systematic Mapping Study of Feyh et al. [S213] concentrated on identifying measures and indicators proposed in the literature concerning Lean Software Development (LSD), and structured them according to ISO/IEC 15939. They found that since 2010, the publications related to LSD have increased significantly. The evaluation research and experience reports are the main research types. The authors identify 22 direct measures, 13 indirect measures, and 14 indicators. By mapping the measures to the principles of lean development, the authors found the following: (1) there are not measures relevant to deferring commitment, (2) there is a lack of tools for measuring stakeholders, (3) knowledge creation was not measured objectively, and (4) there is good support for the fast delivery principle.

This Systematic Mapping Study focuses on the measurement of the development process and its execution projects, mainly to give insight on the measurements related to two entities reported in Gómez et al. [S93], "Project" (12%) and "Process" (9%), and to identify and classify publications related to the measurements in the context of these two entities. Moreover, this study tended to answer questions such as: What are the main measured entities and attributes along the development process? What are the main research methods in this area? What are the main research topics and contexts in this area?

4 RESEARCH METHOD

The Systematic Mapping (scoping) Study is the main selected research method for this study (Bailey et al. 2007), because it is more oriented toward trying to understand the area than toward trying to find a solution to a problem (Petersen et al. 2008). The application of this method aims to structure the research area and identify the research quantity, type, and the available results. Also, it works to discover publication trends over time, identify main publication forums, and determine research gaps to guide future investigations. Systematic Mapping studies do not cope with providing recommendations based on strength of evidence as systematic literature reviews do.

As mentioned earlier, the aim of this study is to get a better understanding of measurement in the context of the software development process, to classify and structure the existing works of

research in this area, and to identify the quantity, research types, and results available within it. To accomplish this goal, we formulated the following research questions:

RQ1: How is the research area structured? What are the trends concerning the publication quantity and focus? What are the leading publication channels for this area?
RQ2: What are the main research types and methods found in the studies?
RQ3: What are the main abstractions and attributes measured and what are the frequencies of different purposes of measurement in the research?
RQ4: What are the main base methods, model, or techniques used in the research?
RQ5: What are the main research contexts?

A research process has been developed and conducted to answer these questions. This process has been incrementally improved in two ways: first, by conducting pilot study after the first database searches to refine the process and research string; and second, by reviewing and modifying the search method and tools during the process.

The research process is composed of the following steps:

(1) Defining search strategy:
    (a) Defining the research string and inclusion and exclusion criteria
    (b) Identifying the research source (databases)
(2) Conducting the research:
    (a) Identifying the relevant studies based on keywords, title, and abstract
    (b) Screening the studies to apply inclusion and exclusion criteria
    (c) Designing a classification scheme
(3) Data extraction, updating the classification scheme, mapping process, and results

The following subsections will describe in detail this research process.

4.1 Defining Search Strategy

Dybå et al. (2007) proposed four steps to define an exhaustive search strategy. Defining the search string is the first step. The second step is defining the inclusion and exclusion criteria. The third step is applying the inclusion and exclusion criteria to the titles and abstracts to find the relevant studies. And the fourth step consists of retrieving the relevant articles for a comprehensive evaluation.

4.1.1 Defining the Search String, Inclusion, and Exclusion Criteria. The search string used to search in the selected databases is defined as suggested by Kitchenham and Charters (2007), who consider population, intervention, and outcome:

—Population. The population is the software development. The keywords ''software development process'' and ''software process'' are used to search for the population.
—Intervention. Intervention refers to the measurement of software development process. For the intervention, we use the keywords "measure," "metric," "indicator," "quantitative," "key process indicator (KPI)" and "process performance indicator (PPI)."
—Outcome. The study is not limited to comparative bodies of research or specific outcomes. Therefore, the research strategy does not consider comparison and outcome.

The search string was formulated as the following:

("software development process" OR "software process") AND (measure OR metric OR indicator OR quantitative OR kpi OR ppi)

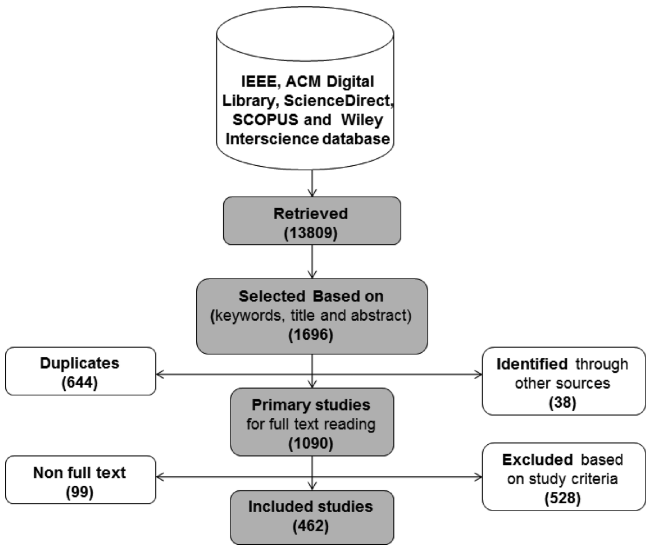

Fig. 1. The selection process.

The inclusion and exclusion criteria were used to identify and include the relevant studies. The inclusion criteria are:

—Topic. It must provide a subject related to measurement in the context of the software development process.
—The literature must be restricted to peer-reviewed journals and conferences.
—Papers must be in English or Spanish.
—Content. The complete text must be available.

And the exclusion criteria are:

—Studies focused on measurement of the final product must be discarded.
—Workshop, book, tutorials must also be excluded.

4.1.2 Identify the Research Source (Databases). The recommendations of Dybå et al. (2007) and Kitchenham and Charters (2007), as well as the search experience reported in the primary study [S163] in the Appendix, have been considered in the selection of the search databases. Therefore, after performing pilot searches, the following databases were selected: ACM Digital Library, IEEE Explore, ScienceDirect (Elsevier), SCOPUS, and Wiley Interscience database.

To find additional relevant literature not discovered during the database search, backwards snowball sampling was used as recommended by Kitchenham and Charters (Kitchenham and Charters 2007). The snowballing process involves gathering all relevant references from the studies found using the database search, then applying the inclusion and exclusion criteria for the title first, then on the abstract. This process was applied to each included study in iterative form, and in the third iteration, no new studies were included.

4.2 Conducting the Research

4.2.1 Identifying the Relevant Studies Based on Keyword, Title, and Abstract. The search on the databases completed on 10 September 2017 has retrieved 13,809 articles, as Figure 1 shows. After applying the selection criteria to the title, abstract, and keywords, 1,686 studies were included as

candidates. Once duplicates and unavailable full-text articles were removed, the total amount of full-text reading articles was 1,013 which includes 38 articles identified through other sources, such as references of candidate articles. Finally, 462 articles were included after applying the inclusion and exclusion criteria.

4.2.2 Classification Scheme. The classification scheme has been constructed to categorize and structure the studies. The scheme consists of several aspects such as: source type, publication year, research type, contribution type, proposal type, validation type, entity/abstraction, and study context.

The classification scheme is created according to the quality attributes mentioned in Lough (2001):

(1) Orthogonality. Clear definition of the categories, which simplify the classification process.
(2) Existing literature. Designing the classification/taxonomy on the basis of comprehensive review and analysis of the area.
(3) Terminology. The classification should use terms inspired by the existing literature.
(4) Completeness. No absent categories; all existing literature can be mapped to a category.
(5) Acceptance. The community approves and realizes the classification/taxonomy.

The classification scheme is designed during the selection process and the full-text reading using a mixed top-down and bottom-up approach. The main topics (i.e., subjects within the measurement domain) and categories are identified based on the research questions. The main aspects of the classification scheme will be described in the following section.

4.2.2.1 Publication Year. Classifying the studies based on the year of publication helps to group the studies and discover trends.

4.2.2.2 Source Type. There are two publication channels identified in the included studies (conference and journal). This classification helps to identify the source of the papers in the research area and the possible target for future publication.

4.2.2.3 Research Approach. The classification of research types provided by Wieringa et al. (2006) was chosen to classify the included studies. Table 1 below displays the classification summary.

Evaluation papers are classified as case study, field study, field experiment, survey, and action research, where the validation study includes work done in the laboratory or office such as experiments, simulation, prototyping, mathematical analysis, mathematical proof of properties, reviews, investigations, and comparisons.

The proposals were classified as methods (including frameworks and approaches), models, techniques, and tools. The proposals were also classified based on their validation methods.

4.2.2.4 Abstraction. The abstraction describes the level at which measurements are conducted. It is essential to consider the abstraction because every entity (e.g., project) has different attributes (e.g., progress and velocity). Moreover, each attribute can be measured in several forms.

— Organization. It refers to the measurement of the software organization comprehensively (e.g., performance and capability).
— Process. It means that measures are directly concerned with work process properties. The time consumed to complete some tasks or phases and the numbers of errors encountered at determined periods are examples of process measures.
— Product. It means that measures are focusing on the attributes of the developed entities during the process.

Table 1. Research Types

Category	Description
Evaluation Research	Evaluate the implementation of a method in a specific context (e.g., industry). Investigate the benefits and drawbacks of implementing the solution in practice.
Validation Research	Investigate the new methods or proposals, which have not been implemented so far (e.g., laboratory works).
Solution Proposal	Propose a solution; this solution could be original or a meaningful extension of an existing method. The consequences of implementing the solution are shown by applying the proposal in practice or by good arguments.
Philosophical Papers	Provide novel forms to see existing subjects by proposing a new conceptual framework, model, or taxonomy.
Opinion Papers	Such papers present personal opinion about a specific method or solution; the opinion could be an assessment of the method or recommendation about it. The authors do not use research methodology or related work to formulate this opinion.
Experience Papers	These papers present a personal experience of the authors about applying a specific method or proposal in the real world.

—Project. It refers to an enactment of a process, involving the management process, which controls the project execution, and the development process, which produces the project outcomes. Typical examples for project measure are the effort, cost, time, and quality.
—Individual. It refers to the people performing the work (e.g., programmer, designer, tester, and project manager).
—A software process model is a real-world software process defined in a specific language that describes its components (e.g., flow, activities, and resources). Some examples of process model measures are complexity and understandability.

4.2.2.5 Measurement Purpose. Lindvall et al. (2005) classify measurement according to the objective of the measurement process, that is to say, to characterize, understand, evaluate, predict, and improve. Characterizing, understanding, and evaluating are the fundamental purposes of measurement as it involves describing or differentiating software processes, explaining the different relations between the development components, and evaluating the achievement of goals or the impact of a technology/process on products. Prediction constructs a model based on existing measures and relations to estimate measures at some future point in time. Improvement is the most sophisticated purpose of the measurement; it identifies the target and activities of the improvement process.

4.2.2.6 Base Methods, Models, and Techniques. The studies are classified according to the base method, model, or the technique used to develop the solution or to carry out the study. Examples of the base methods found in the included studies are: Fuzzy Logic, Goal Question Metric (GQM) method, Capability Maturity Model Integration (CMMI) model, and Process Mining.

4.2.2.7 Context. The context defines the setting of the studied object (measurement in this research) and describes the study environment (Petersen and Wohlin 2009) (e.g., Agile software process, object-oriented project, Model Driven Engineering (MDE), and CMMI organization). It

defines in which circumstances the study results are valid. Thus, it is essential to compare the results of the study with other studies and to evaluate the generalizability of the study's conclusions.

4.2.2.8 Validation Method. The validity describes the extent to which a measure or measurement instrument characterizes the state of the measured attribute (Balch 1974). Measurement validation is required for pragmatic and theoretical reasons (Kitchenham et al. 1995). Fenton and Kitchenham (1991) discussed two different validation views: One view investigates the extent to which a measure characterizes the attribute state, and the other analyzes the appropriateness of a measure for predictive purposes. Ejiogu (1993) proposed five fundamentals for the formal validation of software metrics models: the theoretical examination of the target attribute of software behavior the model measures, the mathematical satisfiability of the postulates of a measure (function), the practical technical experimentation or the metrical vindication of the model beyond the theoretical cross-checking, the examination of results of the measure to obtain any beneficial feedback effects for the potential advancement of productivity, and the confirmation of consistency with dimensional analysis.

This aspect classifies studies according to the method used to validate the proposal (e.g., industry experience and case study).

4.3 Data Extraction and Mapping Process

The data extraction form was developed to store relevant data for this study (e.g., title, publication year, publication channel, research type, and topic). The data have been collected during the selection and full-text reading process. Since the included studies were restricted to peer-reviewed papers, the data (e.g., research type, method, and context) were extracted as mentioned by the original authors, in case the authors did not explicitly mention the requested data, these data were extracted by interpreting the article content according to Table 1 and the research methods defined in the primary study [S306] in the Appendix. The development of the classification scheme and the identification of the topics have been updated and completed during the full-text reading and the data extraction process. When a new topic was discovered, the classification scheme was updated and the classification process was restarted for the consideration of the updated classification scheme.

The extracted data are not the same for all topics. The common data extracted from all the topics are:

—Basic data. Title, authors, publication (year and channel)
—Research type (e.g., evaluation, validation, or experience)
—Research method (e.g., case study, review, and comparison)
—Abstraction/entity (e.g., process and project)
—Attributes (e.g., capability or performance)
—Base method, model, and technique and the study context
—Proposal studies (proposal type and validation method)

Next section describes the identified topics and the data extracted for each topic:

Measurement concepts and practices. Studies related to this topic focus on the fundamental concepts, principles and practices of the measurement (i.e., measures, measurement process, and the main entities and possible attributes to be measured). This topic tries to find answers to questions such as why, what, and how to measure.This data was also extracted for this topic:

—Measurement purpose (e.g., characterize and predict)

Assessment and improvement. The main goal of this topic is to characterize an object of interest (e.g., organization, process, project, and individuals) with respect to a chosen model (e.g., CMM,

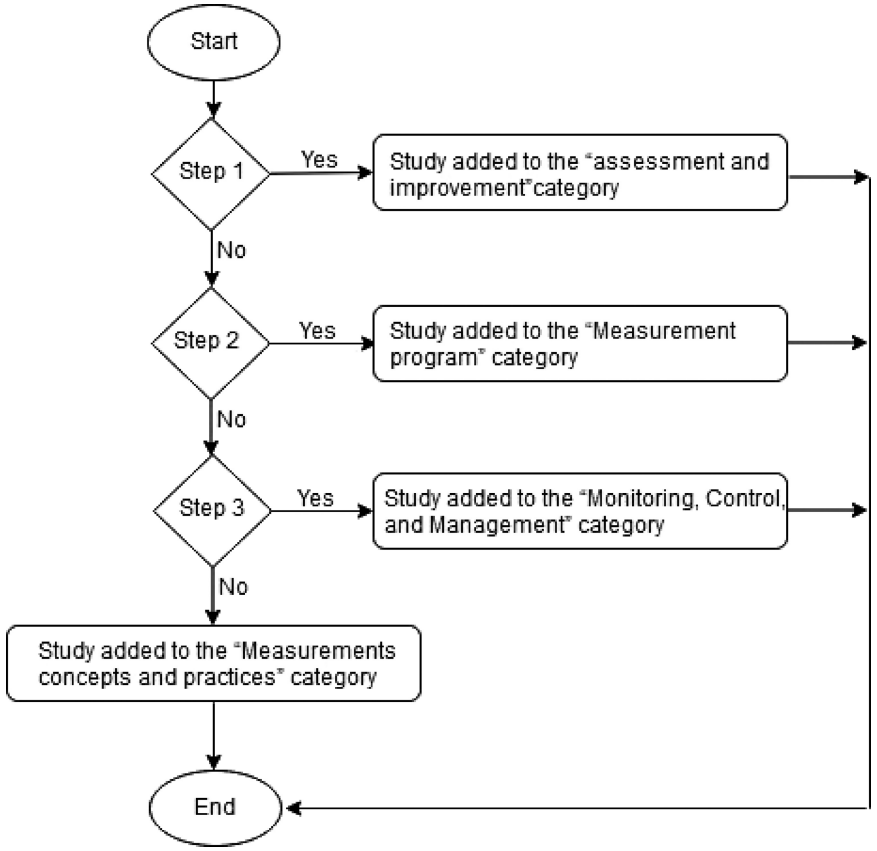

Fig. 2. Categorization process.

ISO/IEC 15504, and Personal Software Process (PSP) (Humphrey 1995)), so that it may be evaluated and improved.

This data was also extracted for this topic:

—Objective (e.g., assessment and improvement)

Measurement program. This topic focuses on how to establish a measurement program within an organization. Studies related to this topic describe experiences, methods and recommendations on establishing a measurement program.

This data was also extracted for this topic:

—Objective (e.g., lesson learned and success factors)

Monitoring, control, and management. This topic focuses on collecting, supervising, analyzing and visualizing data through the execution of the project for management purposes.

The data extracted from the studies of this topic are:

—Objective (e.g., monitoring and management)

Next section describes how primary studies were categorized:

The studies were categorized to topics according to the following process, Figure 2 demonstrates the classification process:

Step1: Does the study focus on assessment or improvement based on a model (standard or customized)?
Yes: Study added to the "assessment and improvement" category.
No: Step2: Does the study focus on measurement program?
Yes: Study added to the "Measurement program" category.
No: Step3: Does the study focus on using the measurements for monitoring, managing, or controlling project?
Yes: Study added to the "Monitoring, Control, and Management" category.
No: Study added to the "Measurements concepts and practices" category.

The study identifies the abstraction levels, measurement purposes, base methods, models, techniques, validation methods, and the contexts found in the included literature. These data were identified and extracted as mentioned by the original authors (during the selection and full-text reading process), when the original authors do not mention explicitly these data, the article content was interpreted to identify and extract it.

Primary studies were screened to identify the measurement purposes addressed by the original authors, after identifying the purpose of the measurement, it was added to the extraction form, then the article was mapped to one of the measurement purposes discussed by Lindvall et al. (2005).

Primary studies were also screened to identify the base methods, models, and techniques used by the original authors; after identifying the used method or model, it was added to the extraction form.

The validation methods used by the original authors to validate the proposals were identified by screening the studies; when the validation method is identified, it was added to the extraction form.

Primary studies were screened also to identify the abstraction level and the attribute addressed by the original authors. The abstraction levels discussed in the primary studies [S163] and [S306] were used as a guide to distinguish and classify the abstraction level and attributes addressed in the included studies.

The contexts of the primary studies were identified, extracted, and classified according to the contexts discussed by Petersen and Wohlin (2009).

5 MAPPING RESULTS

As Figure 1 showed (the selection process) 1,090 articles were identified for full-text reading. After applying the inclusion and exclusion criteria 462 articles were included. Such a number of articles reflect the high importance of the subject (measurement in the context of development process) and the attention it has received from researchers and industry.

5.1 Research Area Structure, Trends, and Publication Forums (RQ1)

5.1.1 Structuring the Area and Trends Over Time. The papers were classified into four topics according to the focus of the study, these topics were measurement concepts and practices (269 papers), assessment and SPI (91 papers), measurement programs (48 papers) and monitoring, control, and management (54 papers). Figure 3 shows the division of the studies into the topics.

Due to the large quantity of studies and the long period covered (the earliest included paper was published in 1984, whereas the most recently included one was published in 2017), articles have been divided into three groups, according to the publication date:

≤2003: Papers published in 2003 and before (123 studies, 27%), and
2004–2009: Papers published between 2004 and 2009 inclusive (139 studies, 30%), and
≥2010: Papers published in 2010 and after (200 studies, 43%).

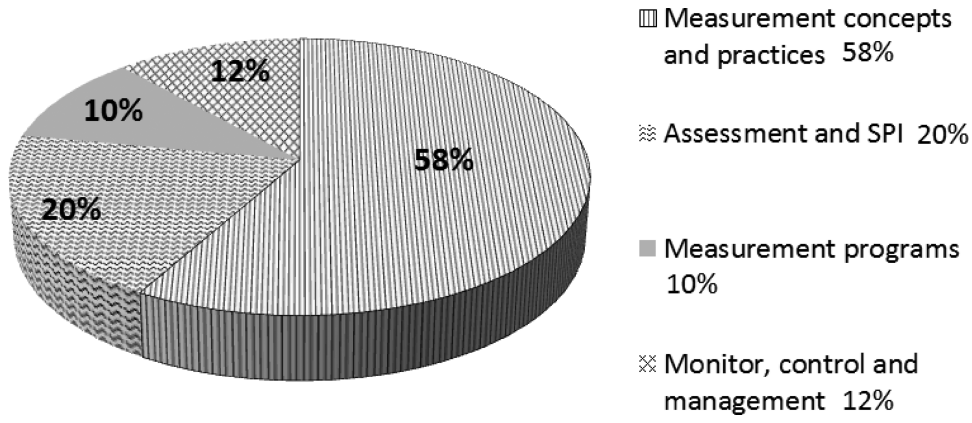

Fig. 3. Percentages of topics.

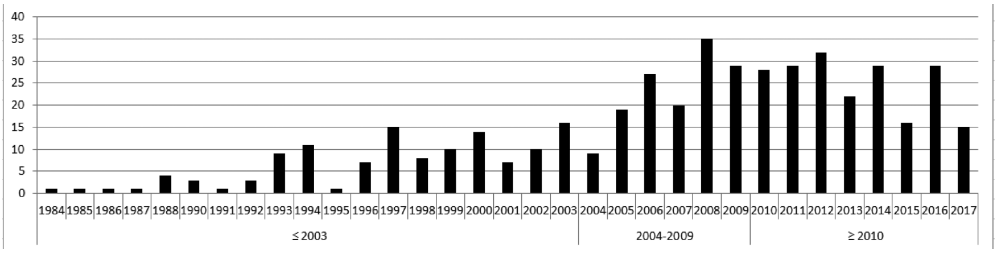

Fig. 4. The included papers along the publication years and periods.

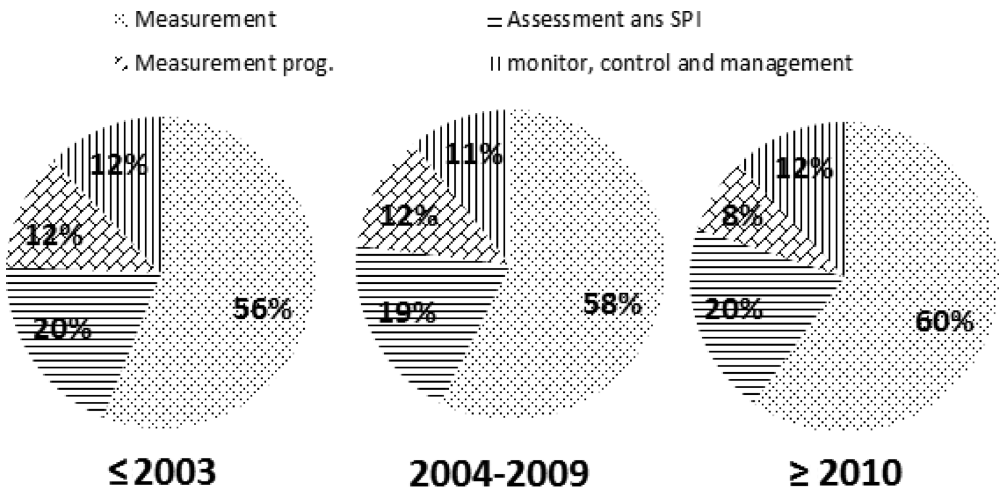

Fig. 5. Distribution of the included papers in topics based on the time period.

Figure 4 shows the number of papers included in each year and time period. It also shows that the number of published papers related to the subject has increased over time, which indicates the growing interest in the topic.

Additionally, Figure 5 shows the evolution of the topics along time periods. The papers discuss the measurement main concepts and practices ([S1–S246] and [S420–S442]) has increased in the

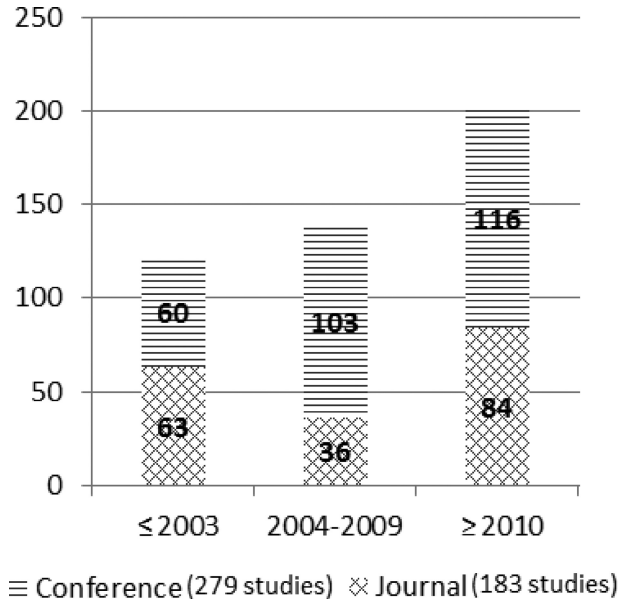

Fig. 6. Studies published in conferences and journals.

last period to be 60% from 56% in the first period, about 20% of the studies is related to assessment and SPI ([S247–S328] and [S443–S451]), the papers related to measurement program topic ([S329–S373] and [S452–S454]) have decreased in the last period to be 8% from 12% in the first period and about 12% of the papers were related to monitor, control, and management ([S374–S419] and [S455–S462]) along these periods. These percentages demonstrate the continuous interest in each of the four topics during all the study period, which means that the researchers and the industry still pay constant attention to each of the main indicated topics.

5.1.2 Publication Forums. About 60% of the included studies were published in conferences, whereas 40% were issued in journals. Figure 6 shows the distribution of forums with respect to time.

The included studies were published in about 220 forums. More than 75 publication forums were identified for 200 papers included in the most recent period of time (≥2010). Due to the large number of forums, Table 2 lists the top publication forums found in the last years (≤2010), and Figure 7 shows the distribution of papers according to the forum and the topic.

5.2 Research Types (RQ2)

Among the selected papers, almost 70% were solution proposal studies, 15% corresponded to evaluation studies, 8% were experience studies, and 7% concerned validation studies. Neither philosophical nor opinion studies were included. Figure 8 shows the research types according to the time period and research topics.

The research methods used in the validation and evaluation studies were distributed as displayed in Figure 9: 23 case studies, 25 reviews (e.g., SLR and mapping studies), 11 surveys, and other research methods.

As mentioned before, around 70% of the included studies were solution proposal (325 studies), 74% of the proposals were classified into methods (to increase readability, this category includes

Table 2. Top Publication Forum in the Recent Years (The period: ≥2010)

Rank	Forums type	Forums	No. of papers
1	Conference	Software Process and Product Measurement	20
2	Journal	Software: Evolution and Process	15
3	Journal	Information and Software Technology	8
4	Journal	Systems and Software	6
4	Conference	Software and System Process	6
5	Conference	IEEE/ACM Intel. Conf. on Automated Software Engineering (ASE)	4
5	Journal	Software Maintenance and Evolution: Research and Practice	4
5	Conference	Product-Focused Software Process Improvement (PROFES)	4

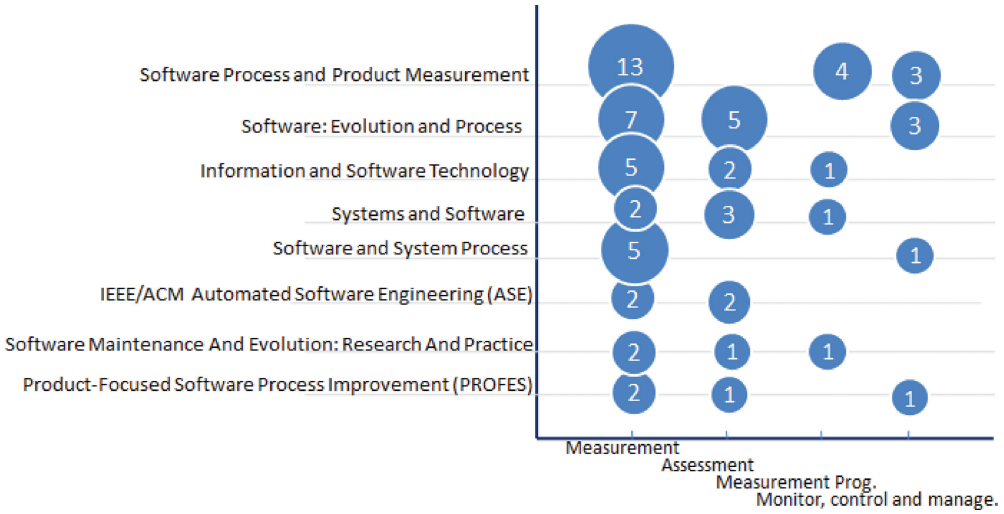

Fig. 7. Distribution of papers in the top forums according to the paper topic.

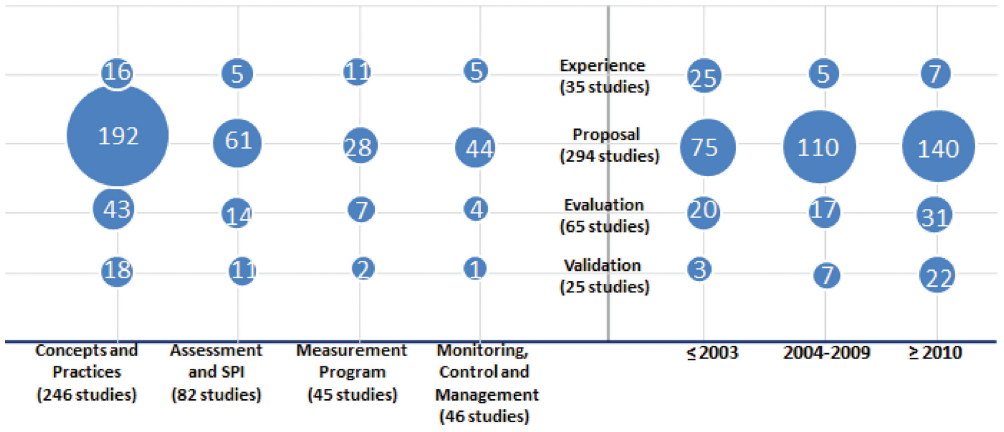

Fig. 8. Distribution of research types.

Fig. 9. Research methods used in the evaluation and validation studies.

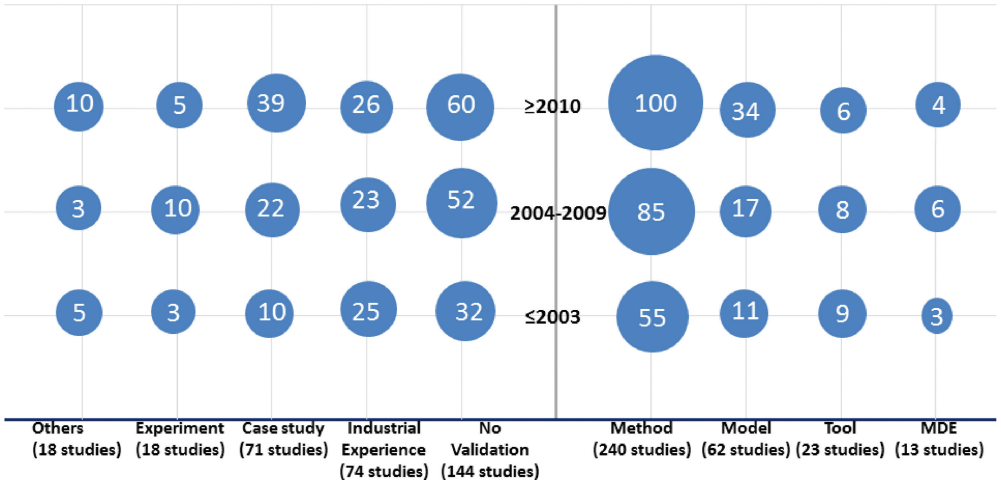

Fig. 10. Proposal types and validation methods based on time periods.

methods, frameworks, approaches, and processes), 19% were models (e.g., conceptual models and mathematical models), and 7% were tools. Thirteen proposal studies, which represented 4% of the total, were based on Model-Driven Paradigm. Besides, 44% of the proposal studies did not mention any proposal validation, whereas 23% of them used industrial experience to validate the proposal, 22% used case study, 6% of them used experiment, and 6% utilized other methods, such as prototypes. Figures 10 and 11 shows more details about the classification of the proposal studies.

5.3 Measured Abstractions, Attributes, and Measurement Purpose (RQ3)

5.3.1 Measurement Concepts and Practices. Five abstractions were found in this topic: Project (110 papers, 24 attributes), Process (76 papers, 17 attributes), Developer (12 papers, 9 attributes),

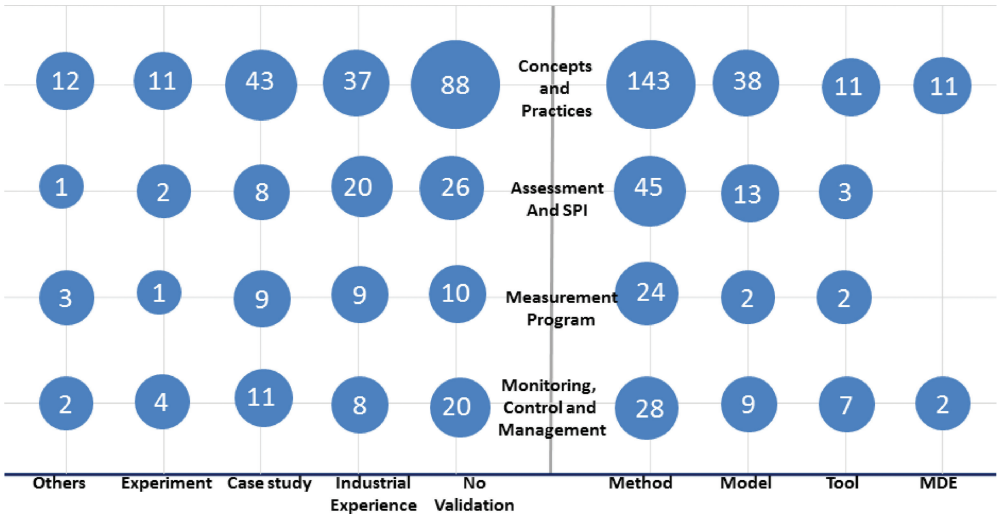

Fig. 11. Proposal types and validation methods based on topics.

Table 3. Abstractions Studies with Respect to Time Periods

	Project	Process	Developer	Process Model	Organization
≥2010	43	42	9	2	
2004–2009	43	18	1	4	1
≤2003	24	16	2	1	

Process Model (6 papers, 3 attributes), and Organization (1 paper, 1 attribute). Table 3 shows the frequency of the studies related to each abstraction according to time period.

The most investigated attributes were project effort (21 studies), project productivity (8 studies), project risk (8 studies), process productivity (7 studies), and process quality (7 studies). Tables 4 to 8 show the attributes of each entity and the studies related to each attribute.

The studies related to this topic were also classified according to the purpose of measuring the entities. Table 9 lists the classification results.

5.3.2 Assessment and SPI. This topic focuses on four abstractions: process (69 papers, 13 attributes), developer (7 papers, 3 attributes), project (4 papers, 1 attribute), and organization (3 articles, 2 attributes). The most investigated attributes were: process capability (19 studies), process maturity (15 studies), process performance (14 studies), and developer performance (5 studies). Table 10 details the abstractions and attributes.

5.3.3 Measurement Program. This topic mainly focuses on the execution environment of software development process. Organization and project were the main abstractions found. Table 11 shows the abstractions and the distribution of the studies based on the time periods.

5.3.4 Monitoring, Control, and Management. Monitoring studies constituted about 46% of the studies related to this topic, 22% focused on control, whereas 31% of the studies were oriented toward management. Table 12 demonstrates the division of the studies based on the objectives and time periods.

Table 4. Process Attributes

Process attribute	≥2010	2004–2009	≤2003	Total studies
Productivity	S163, S166, S221, S239	S106	S2, S7	7
Quality	S150, S180, S231, S426	S77, S140	S28	7
Performance	S177, S191, S227, S436		S17	5
Cost	S150	S81	S28	3
Sustainability	S197, S200, S216			3
Risk	S183, S214, S226			3
Agility	S152, S442	S104		3
Issue, Bugs	S194, S237			2
Conformance	S438	S115	S46	3
Security	S196			1
Documentation	S223			1
Activities	S242			1
Information leakage		S80		1
Cycle time	S243			1
Product Value	S186			1
Suitability	S207			1
Commonality	S441			1

The majority of the studies, about 74%, were based on the project, whereas 19% of them focused on the process. The main investigated attributes in this topic were: project bottleneck (e.g., the primary study [S408]), project progress (e.g., the primary studies [S392] and [S380]), project performance (e.g., the primary study [S395]), and project risk (e.g., the primary study [S415]).

5.4 Base Methods, Models, and Techniques (RQ4)

Many methods, models, techniques, and standards were used in the studies as a base to build the solution proposal or to carry out the research. The most used methods were GQM, applied to 37 studies, CMMI model used in 13 studies, and Statistical Process Control (SPC) carried out in nine studies. Table 13 displays the main methods according to the research topic. Below, Figure 12 shows in alphabetical order the main models, methods, and techniques used according to time periods.

5.5 Research Context (RQ5)

The context of the study defines the environment in which the research is carried out and also determines the validity conditions of the study contribution.

The organization is the most frequent context dimension found in the primary studies, while the process dimension is the second. Table 14 categorizes the contexts found in the literature according to the context dimensions mentioned in Petersen and Wohlin (2009).

Different contexts have been distinguished in the included studies; Agile and Lean development (35 studies) and Small- and Medium-Size Enterprise (22 studies) were the most frequent contexts in the studies. Table 15 lists the identified contexts according to the topic and Figure 13 shows them in alphabetical order according to time periods.

6 THREATS OF VALIDITY

The discussion on validity threats is necessary to assess the conclusion's reliability. The main threats for this study were researcher bias, the absence of related articles and the classification

Table 5. Project Attributes

Project attribute	≥2010	2004–2009	≤2003	Total studies
Effort	S161, S162, S173, S175, S179, S185, S420, S421, S424, S427, S428, S430, S433, S435	S73, S85, S87, S101, S135	S32, S69	21
Productivity	S236	S125, S139, S147, S98	S29, S31, S64	8
Risk	S234	S72, S74, S88, S94, S122, S123, S105		8
Size	S219, S182	S71, S98, S100	S5	6
Performance	S169, S437	S137, S145	S43, S36	6
Release readiness	S222, S232	S127, S129, S141		5
Success	S181, S187, S225		S56	4
Cost		S92, S113, S120	S13	4
Progress	S189	S78, S134		3
Schedule	S215	S92	S13	3
Faults, defects	S159	S142		2
Similarity	S157	S126		2
Conformance	S434	S146, S149		3
Agility		S148		1
Quality		S114		1
Coordination delays	S188			1
Sustainability	S198			1
Velocity		S138		1
Customer satisfaction			S50	1
Rework			S16	1
Reuse			S37	1
Visibility index			S27	1
Communication			S33	1
Complexity	S425			1

Table 6. Developer Attributes

Developer attribute	≥2010	2004–2009	≤2003	Total studies
Team performance	S184, S439		S41	3
Performance	S233, S423, S429			3
Productivity			S3	1
Emotions	S202			1
Human aspects		S112		1
Time estimation performance	S190			1
Confirmation biases	S212			1
Creativity	S431			1
Team Climate	S431			1

Table 7. Process-model Attributes

Process-Model attribute	≥2010	2004–2009	≤2003	Total studies
Quality	S178, S206, S217, S172			4
Complexity		S86		1
Maintainability		S76		1

Table 8. Organization Attributes

Organization attribute	≥2010	2004–2009	≤2003	Total studies
Performance		S109		1

Table 9. Studies According to Measurement Purpose

Measurement purpose	Characterize and understand	Evaluate	Predictand estimate	Improve	Not applicable
≥2010	52	13	25	11	22
2004–2009	45	6	15	5	9
≤2003	29	3	9	13	15
Total	126	22	49	29	46

scheme. Some actions were taken to reduce and control these threats. Thus, other researchers would draw the same conclusions if they were to repeat this mapping study. In this section, the validity threats are discussed in terms of construct, internal, external, and conclusion validity, according to Wohlin et al. (2000).

Construct validity concerns about to what extent the study really represents researchers' goals, and to what extent the carried out investigation supports achieving these objectives. To minimize this threat, the research questions were carefully defined to consider the study objectives and to clarify its scope. Another aspect of the construct validity is the absence of important literature related to the research subject as well as incorrect data extraction; both affect the validity of the study and the generalizability of the conclusions. To reduce this threat, the backwards snowball was conducted to supplement the database search. Also, the main pre-existing works related to this study have been identified.

Internal validity is related to data analysis. Since the analysis uses only descriptive statistics, the threats are in the minimal limits. Additional aspect of the internal validity is the researcher bias; the research process was conducted by an individual researcher, which posed a threat. Mitigating this threat was impossible, but some actions were taken to reduce it:

— Other authors reviewed the research protocol to assure objective clarity and understandability of the research process (especially inclusion and exclusion criteria and data extraction process).
— The first author has conducted the selection process under the supervision of the other authors as recommended by Kitchenham (2004) for Ph.D. students. The first researcher is who has decided which articles should be included or excluded. In the case of uncertainty, the article would be classified for other authors' review to avoid excluding relevant work and reducing bias.
— The test-retest validation [S156] was applied to reduce the errors in both inclusion/exclusion process and data extraction process.

Table 10. Assessment and SPI Abstractions and Attributes

Abstraction	Attribute	≥2010	2004–2009	<2003	Total
Process					
	Capability	S298, S300, S304, S310, S311, S315, S319, S320, S321, and S324	S273, S275, S280, S288, S291, S294, and S297	S265 and S268	19
	Maturity	S304, S309, S311, S317, S320, S321, S323, and S324	S276	S247, S252, S253, S256, S248, and S260	15
	Performance	S298, S302, S303, S305, S307, S313, and S444	S279, S280, S283, S286, S292, and S293	S269	14
	Quality	S307, S308, S443, and S450	S276 and S278		6
	Risk	S313,S299,	S288		3
	Strengths and Weaknesses	S299, and S326			2
	Consistency			S258 and S270	2
	Conformance	S327, S445, and S447			3
	Credibility	S316			1
	Management		S284		1
	Security		S274		1
	Effectiveness			S249	1
	Productivity	S443			1
Organization					
	Quality	S314		S266	2
	Performance		S296		1
Project					
	Performance	S312	S271 and S282	S267	4
Developer					
	Performance		S285	S251, S254, S257, and S262	5
	Team maturity	S328			1
	Productivity		S277		1

External validity is related to the generalizability of the study's results. The findings of this study depend on the retrieved articles using the defined search strings, search databases, and the research method; to minimize this threat, we have formulated the search string and selected the search databases according to the suggestions and previous experiences mentioned in Section 4. Another threat is the categories defined to classify the primary studies; this threat was reduced by the careful development of the classification scheme according to the quality attributes mentioned

Table 11. Measurement Program Abstractions

Abstract/Entity	≥2010	2004–2009	≤2003	Total
Organization	S361, S362, S363 S364, S365, S366 S367, S368, S369 S370, S371, S372, S373, S453, and S454	S345, S347, S353 S354, S355, S356 S357, and S360	S332, S333, S334 S338, S339, S340 S341, S342, and S343	32
Project	S364	S344, S346, S349 S351, S352, and S359	S330, S335, S336 and S337	11
Process			S331	1

Table 12. Studies Objective

	Monitoring	Control	Management
≥2010	12	1	11
2004-2009	8	4	3
≤2003	5	7	3

in Section 4.2.2. The findings of this study are consistent with the results of another recent study [S451], which gives more credibility to the results of this study. The long period covered by the study and the number of studies screened (1,090 papers) and included (462 papers) also support the generalizability of results of this study, because it represents a big sample of the literature.

Conclusion validity focus on whether the data are collected and the analysis is conducted in a way that affects the ability to draw correct conclusion. The research strings and the databases were defined and selected carefully according to the recommendations and the prior experiences mentioned in Section 4.1. Furthermore, the systematic mapping study procedures were first piloted and incrementally improved to refine the search string and the selection criteria as well as to ensure the common understanding of selection criteria and the data extraction process among the authors. The classification scheme and misclassification are another source of threats to the conclusion reliability, these threats have been addressed by the careful development and testing of the classification scheme. For this, the authors jointly studied titles, keywords and abstracts of the articles that appeared in the pilot study and in the selection process. The scheme was refined during the classification process, taking into account the quality attributes mentioned in Lough (2001). To reduce the risk of misclassification, it was double-checked by two researchers. Also, the classification process was repeated each time the classification scheme was updated.

7 SUMMARY OF FINDINGS

This section discusses the principal findings based on the results.

7.1 Publication Trend and Forums

The number of publications found in conferences and journals indicates that this research area was very seriously considered by researchers. Two hundred twenty forums were identified in the included studies (conferences 60% and journals 40%). This diversity of forums reflects the importance and the interest of different disciplines in the subject.

To analyze the trends, the studies were divided into three time periods. The results show that the number of studies was increasing over time, which reflects the growing interest in the area by

Table 13. Base Models Methods and Techniques

Type	Model, Method, and Technique	Concepts and Practices	Assessment and SPI	Measurement Program	Monitoring, Control and Management
Models and Standards					
	CMM		4		1
	CMMI	5	7	1	
	ISO/IEC 15504		4	1	
	ISO/IEC 33000	1			
	Personal Software Process (PSP)		4	1	2
	K-model		2		
	ISO/IEC 15939	2		3	
	ISO/IEC 9126	1			
Methods and Techniques					
	Analogy-Based Estimation (ABE)	1			
	Balanced Scorecard	3		1	
	Bayesian Network	5			
	Data Envelopment Analysis (DEA)	5			
	Data-mining techniques	2			1
	Earned Value Management (EVM)	2			3
	Function Points	1			
	Fuzzy logic	7			1
	Goal Question Metric (GQM)	23	1	11	2
	Goal-Driven Software Measurement	1			
	Grey Relational Analysis (GRA)	1			
	Neural Networks	3			
	Practical Software Measurement (PSM)	2		1	
	Six Sigma	2	2	1	1
	Socio-Technical Systems Theory	1			
	Statistical Process Control (SPC)	1			8
	Use Case Points (UCP) and COSMIC-FFP Methods	1			

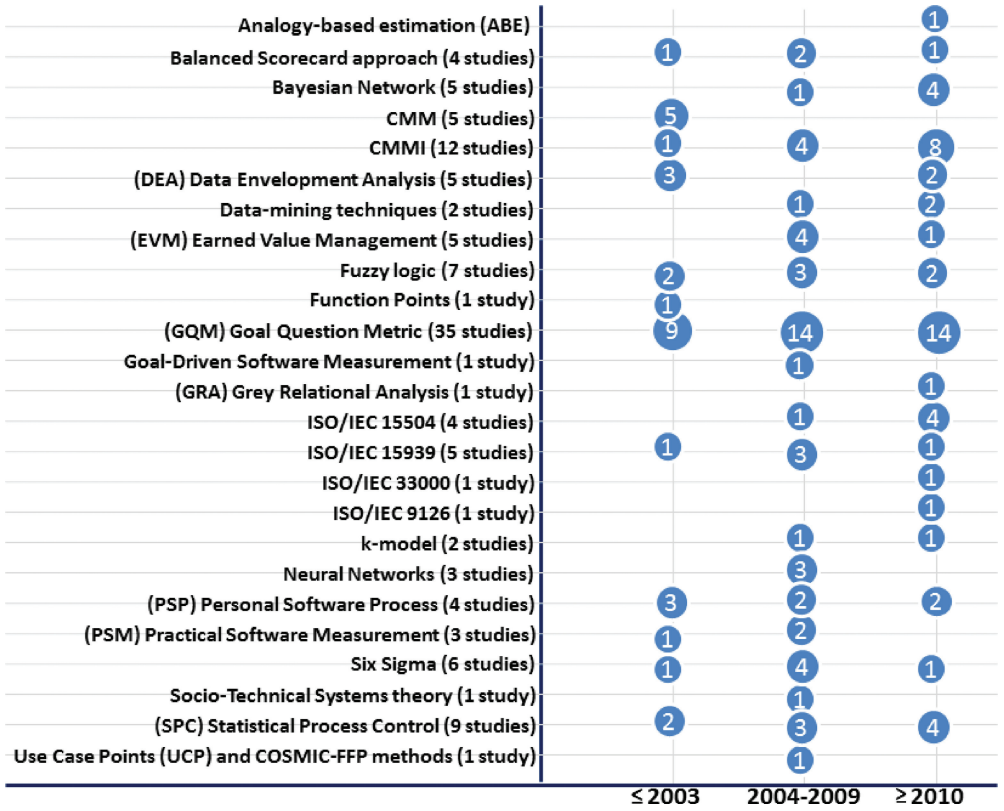

Fig. 12. Base methods, models, and techniques.

researchers and practitioners as well as the constant and continuous focus on the four topics that structure the research area.

We can see that the percentage of the studies related to the topics "measurement program" has decreased in recent years. In the same period the studies related to the topic "monitoring, control, and management" has increased, which means that there are growing interest in monitoring the development process and using monitoring data to control and manage it. Studies related to "assessment and improvement" have increased in the last period, which reflect the growing awareness of researchers and practitioners regarding the importance of assessment and improvement, and the studies related to "measurement concepts and practices" have also increased, which indicates that researchers are still studying the entities, their attributes, and developing methods to measure them. Furthermore, we note that, in the last years, top publication forums have published more studies related to the topics "measurement concepts and practices" and "assessment and improvement" than studies related to "measurement programs" and "monitor, control, and management."

7.2 Area Structure

The area was divided into four topics based on the focus of the studies.

7.2.1 Measurement Concepts and Practices. Studies related to this topic focus on the main measurement principles, concepts and methods. This topic tries to answer questions such as "why,

Table 14. Context Dimensions

Dimension	Context
Market	No studies found
Organization (63 studies)	CMM CMMI ISO/IEC 15504 Outsourcing Small and Medium-Size Enterprises (SME)
People	No studies found
Practices, Tools, Techniques (24 studies)	Application of Metrics in Industry (AMI) Goal Question Metric (GQM) GQM + Strategies Object-Oriented Development Practical Software Measurement (PSM) Personal Software Process (PSP) Test-Driven Development
Process (45 studies)	Agile and Lean Development Incremental and Iterative Software Development Open Source Software Development Global software development (GSD)
Product (1 study)	Embedded System

what and how to measure?", by investigating the development process (e.g., inputs, outputs, artifacts, and entities), the available measurement methods and studying the relations among the different entities and attributes to develop new models and methods to evaluate or estimate specific attributes. Based on the results, the main focus was on project effort, productivity and risks. The main purpose of analyzing effort is to estimate and control the project cost, which in turn increases the productivity. Consequently, it can be stated that making the development process cost effective is the main objective of studies related to this topic.

7.2.2 Assessment and Improvement. Studies of this topic focus on the organization's processes and projects, and investigate the way to increase the capability, maturity and the performance. For that, studies related to this topic focus on maturity standards like CMMI and ISO/IEC 15504, develop and validate assessment methods to evaluate the organization's processes to determine its maturity level, and propose practices to reach a high maturity level. Additionally, studies related to this topic also aim to develop customized models to meet specific contexts like agile development and small and medium-size organizations. Thus, the final objective of this topic is to improve the development process capability and maturity, which in turn enhances the final product quality.

7.2.3 Measurement Program. Studies of this topic investigate applying the measurement methods, models, and techniques to establish measurement procedures and infrastructure in real-world organizations. For this reason, the studies analyzed how to implement the measurement program

Table 15. Contexts

Dimension	Context	Concepts and Practices	Assessment and SPI	Measurement program	Monitoring, Control and Management	Total
Organization						
	CMM	2	8			10
	CMMI	9	9	1		19
	ISO/IEC 15504		9			9
	Outsourcing	2			1	3
	Small and Medium-Size Enterprises (SME)	3	15	3	1	22
Practices, Tools, and Techniques						
	Application of Metrics in Industry (AMI)	2				2
	Goal Question Metric (GQM)	3				3
	GQM + Strategies	2		1		3
	Object-Oriented Development	2				2
	Practical Software Measurement (PSM)	1		1		2
	Personal Software Process (PSP)	3	7			10
	Test-Driven Development	2				2
Process						
	Agile and Lean Development	22	7	1	5	35
	Incremental and Iterative Software Development	5				5
	Open Source Software Development	1			1	2
	Global Software Development (GSD)	1	1		1	3
Product						
	Embedded System			1		1

in organizations, proposed methods to ensure the successful introduction of measurement into the organization's processes, and investigated linking the measurement program with the organization's strategies and goals. The main objectives of the literature related to this topic are to describe implementation experiences (25%), highlight lessons learned from previous experiences of implementing measurement programs (25%), and discuss the success factors that affect the implementation of the measurement program (23%).

7.2.4 Monitoring, Control, and Management. This topic focuses on monitoring the project execution to measure progress, detect/control deviations from the project plan, and for the quantitative project management in specific contexts (e.g., small organization). Monitoring the development project was also carried out to establish a baseline to be a starting point for improvement process and to evaluate the results of this process.

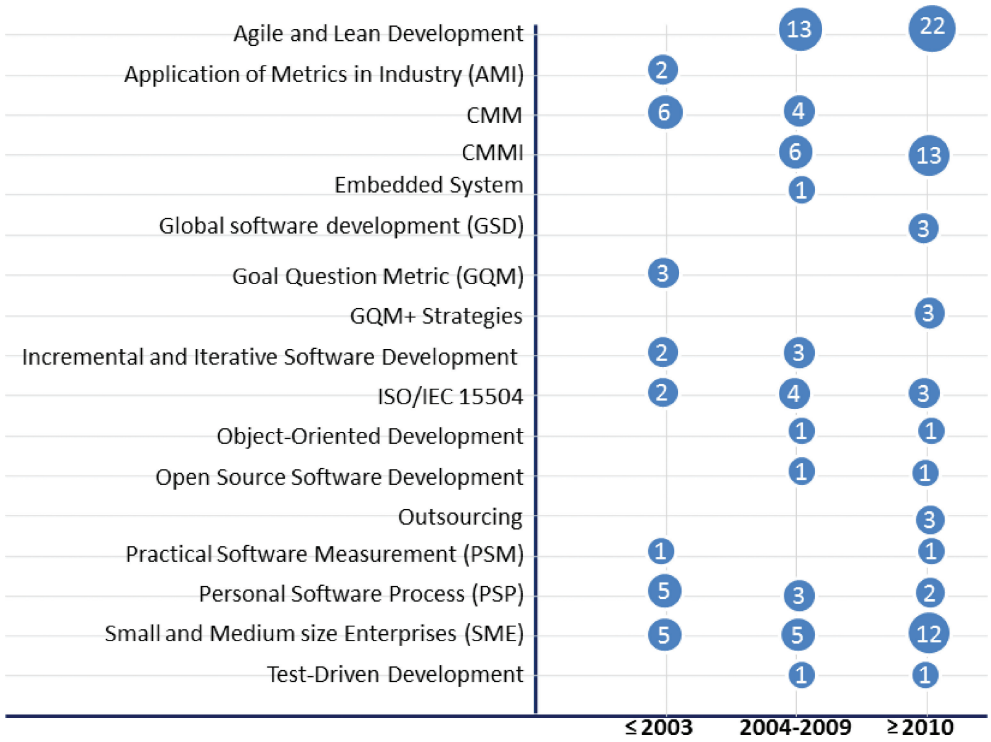

Fig. 13. Contexts according to time periods.

7.3 Research Types

Solution proposal was the most frequent research type identified in the studies. About 56% of the proposal studies were validated, 23% used industrial experience to validate the proposal, and 22% used case studies.

The validated proposals have increased in the last years. In the period (≥2010), about 57% of the proposals were validated, whereas 53% were validated in the previous period (2004–2009).

The following section discusses the research types in each topic.

7.3.1 Measurement Concepts and Practices. Seventy-one percent of the studies in this topic are proposals (46% of these proposals were not validated), about 6% of the studies are review studies, 5% empirical studies, 5% case studies. Few surveys (five studies) and comparisons (two studies) were found in the literature related to this topic.

7.3.2 Assessment and Improvement. Sixty-seven percent of the studies related to this topic are proposals, 43% of these proposals were not validated, and the industrial experiences were the most used validation method in this topic. This low percentage of the validated proposals shows the necessity to focus more on validating the proposals of this topic. There are six reviews in the literature of this topic in the last period; this research type was not used in this topic in the previous periods, and this means that researchers are working to analyze the previous work in this topic to discover gaps and new investigation lines.

7.3.3 Measurement Program. Proposals constitute 58% of this topic, and 36% of these proposals were not validated. Few reviews and surveys found in this topic raise the necessity to focus on these research methods, so more focus on validating the proposals are needed.

7.3.4 Monitoring, Control, and Management. Eighty-one percent of the studies related to this topic are proposals, and 45% of these proposals were not validated. Only one review found in this topic, neither surveys nor comparisons were found in the literature of this topic. More studies focusing on these research methods and more focus on validating the proposals are needed.

7.4 Measured Abstractions and Attributes

The most frequently measured entities were project (167 studies) and process (156 studies). For the project, the most measured attributes were: effort, performance, risk, productivity, and progress. While the most measured process attributes were capability and performance, maturity, quality, productivity, risk, and conformance. Performance was the most measured attribute for the developer, the quality attribute for the process model, and for the organization, the performance and the quality were the most measured attributes.

The following sections discuss the abstractions and the attributes for each topic.

7.4.1 Measurement Concepts and Practices. In this topic, the focus on process and the developer has increased in recent years. For process, there are more studies in recent years (period 2010) about process (productivity, quality, and performance). Also there are studies focusing on the sustainability and the risks. Furthermore, there are few studies focusing on the process agility, security, and the documentation. And only one study discusses the development of the software as a service, which demonstrates the necessity of more focus on these subjects.

Studies related to the developer have increased in the recent years, reflecting the increased awareness of the importance of the human factor in the development process. The total studies that focus on the developer in this topic were 13 studies (about 5%).

Moreover, there are few studies that focus on the process model (6 studies). Furthermore, there is a scarcity of studies that focus on the definition and integration of the measures as an element of the process model. This integration supports the communication and the common understanding of the measurement plan among the participants. Also, few studies use the MDE paradigm (12 proposal studies, 4%) for the definition and integration of the measures in the development process lifecycle. MDE paradigm can support the formal definition and automation of the measure lifecycle.

The focus on measuring the project effort and success has increased, with researchers still focusing on measuring the project size and the release readiness. Also, we observe a lack of studies focusing on the measurement of the project rework, reuse, communication, progress, and customer satisfaction, which are important attributes that affect the project success and support the global understanding and evaluation of the project performance.

The studies related to this topic were also classified according to the purpose of measuring the entities. As shown in Table 9, the majority of the studies were focusing on characterizing/understanding 56% and predicting/estimating (22%) the measured entity.

7.4.2 Assessment and Improvement. In recent years, research is focusing more on the assessment and improvement of the process capability, maturity, and performance. This indicates the increasing awareness of the importance of improvement and following the best practices. On the other side, there is a scarcity of studies focusing on the assessment and improvement of the process security, management, and risk management. Also, there are few studies focusing on the assessment and improvement of the developer as an individual and as a team. Studies in the literature

focus only on the developer performance, maturity, and productivity. More studies are needed to address these aspects, also, assessing and improving the communication inside the development team and between the development team and the other teams is also needed.

7.4.3 Measurement Program. This topic mainly focuses on the implementation and integration of the measurement programs within the execution environment of the development process in the organization and the project levels. Main studies focus on the organization level (63%), then the project level (23%).

7.4.4 Monitoring, Control, and Management. This topic focuses on the project and process. More studies are needed on the project performance, bottlenecks, deviation, progress and risk. We note that in the last period the focus on the management has increased. For process, more studies are needed on monitoring the stability of the process and on establishing the control limits for the Statistical Process Control.

7.5 Base Methods, Models, and Techniques

The most identified base methods, models, and techniques were GQM, CMMI, and SPC. The only common method used in all the topics was Six Sigma.

The following sections discuss the base methods, models, and techniques found in each topic.

7.5.1 Measurement Concepts and Practices. GQM is the most used method in this topic (9% of the studies), then the data mining and analysis methods like fuzzy logic, Bayesian network, and Data Envelopment Analysis (DEA). There is a limited number of studies that focus on measurement methods like PSM, Goal-Driven Software Measurement, and the international standard ISO/IEC 15939, so further investigation on these methods is needed. Also, there is a lack of comparisons and surveys discussing experiences of applying these methods.

7.5.2 Assessment and Improvement. CMMI was the most-used model in this topic. There is a lack of studies focusing on the international standard ISO/IEC 15504. Due to the increasing adoption of the agile methods (which focus more on the developer), the need for more research on methods focusing on assessing and improving the developer aspects (e.g., PSP) is increasing.

7.5.3 Measurement Program. GQM and ISO/IEC 15939 are the main methods used in this topic. There is a lack of studies related to methods that address different levels and aspects of the organizations such as balanced scorecard.

7.5.4 Monitor, Control, and Management. SPC is the most used method in this topic. There is a lack of studies related to methods that focus on monitoring and enhancing the variability, quality, and the performance, such as Six Sigma.

7.6 Research Context

Only 29% of the 462 studies have described the context in which the study was carried out. The highest proportion of studies that describe their context was found in assessment and SPI topic (62%), then 22% in the measurement concepts and practices topic, 17% for measurement program topic, and 17% for studies related to the monitor, control, and management topic.

Agile/lean development and small-/medium-size enterprise are the most common context in the literature. CMMI and ISO/IEC 15504 are the main maturity standards found in the last years.The following sections discuss the research contexts found in each topic.

7.6.1 Measurement Concepts and Practices. Major contexts used by the researchers in this topic are Agile/Lean Development and CMMI. No studies were found related to the embedded system

and ISO/IEC 15504 contexts. Also, there is a lack of studies related to contexts like SME, open source development, and outsourcing.

7.6.2 Assessment and Improvement. Studies of this topic are mainly distributed between the following context: SME (15 studies), ISO/IEC 15504 (9 studies), CMMI (9 studies), and Agile/Lean Development (7 studies). No studies were found related to the outsourcing context. Assessment and improvement methods related to this context could support the organizations to demonstrate their strength and support the clients in their evaluation of organizations that work in this context.

7.6.3 Measurement Program. The most used context in this topic is SME. No studies were found related to open source development, ISO/IEC 15504 and outsourcing contexts, and only one study was related to the agile/lean context and CMMI context.

7.6.4 Monitor, Control, and Management. The most investigated context in this topic is agile/lean context. Recently, there is focus on outsourcing and open source development contexts. Only one study was related to SME. And no studies were found related to CMMI and ISO/IEC 15504 contexts. More studies about these contexts are needed.

8 IMPLICATIONS FOR RESEARCH AND PRACTICE

The findings of this study illustrate that there is relatively constant interest in the domain and in the main indicated topics. As shown, the majority of the studies were published in conferences, which mean that researchers tend to publish their work in conferences more than in journals.

Studies related to the measurement program topic has decreased in the last period, and in the same time the studies related to monitoring, control, and management has increased, this means that in recent years establishing the measurement program was the major challenge for the academic and industrial community, but the focus now has changed to be on how to use the data provided by these programs for better monitoring, control, and management of the development process and its projects. Another evidence of the growing interest in using the data provided by measurement programs is the growing focus on using real data for prediction and estimation purposes.

Solution proposal was the most frequent research type found in the literature. Such result means that investigators are still searching for better methods to measure the development process. A large section of these proposals was not validated; the last period shows more validated proposals, and this indicates that more researchers are focusing on validating their proposal. Although the number of validated proposals is increasing, the current situation demonstrates the need for more attention on validating the proposals. Furthermore, in recent years, more case studies were used to validate the proposals, which indicate that researchers tend to consider the case study as the main validation method.

In recent years, there are more surveys and case studies, which show a trend toward these research methods. Survey studies are important to describe and explain knowledge over large populations. Also, comparisons highlight the strength and weakness of each method. The presence of a large number of proposals increases the need for comparisons and surveys, because they provide more information about the different proposals and demonstrate the results of their use in different contexts, which helps researchers and practitioners identify which of these proposals is more suitable to meet their needs. Therefore, more surveys and comparison studies and more focus on validating the proposals are needed.

Moreover, there are more systematic mapping studies, which indicate that researchers tend to get first a big image about the literature to discover possible gaps, and then apply systematic literature review on a subset of the studies included in the mapping study.

Five abstractions and 64 attributes were identified; Performance was the most measured attribute in the literature. There is growing interest in measuring the developer as individual and the development team, in the last period there are more studies and new attributes measured; this could also be related to the growing adoption of the agile models. The main subjects and attributes investigated in relation to the developers were: the social-technical congruence, the developer and the team performance, developer-emotion and confirmation biases. Recently, new attributes such as creativity and team climate have emerged. Given the significant expansion of the development methods that focus on the developer, we believe that it is necessary to increase the research related to this abstraction.

The majority of the studies related to the topic: measurement concepts and practices focus on characterizing/understanding and predicting/estimating the measured entity. This could be due to the invisible and intangible nature of the development process and its resulting product, so the measurement could be used as a tool to make the development process and its results more transparent [S112]. There is a growing focus on using the measurement as a predicting and estimating tool, which could be seen as a result of the increasing use of measurement to characterize and understand the actuality of the development process. As a result, researchers and practitioners are focusing more and more on the exploitation of the collected measures and previous experience to improve the ability to foresee aspects of the future development processes. Prediction and estimation methods are important, because they support and complement any predictions made during the development process [S158]. Due to the number of studies related to this subject and its importance, more studies are needed.

CMMI and GQM were the most identified model and method in the literature. The most prominent methods identified were data mining and analysis techniques (e.g., Bayesian network, fuzzy logic, artificial neural networks, and regression analysis), which manifest the growing need for analyzing and mining the measurement data to support management activities (e.g., planning and decision making). The frequency of using the models and methods demonstrates that the industrial models and methods (e.g., CMMI and GQM) have a higher impact on the literature than the international standards (e.g., ISO 15504 and 15939).

Seventy-one percent of the studies did not describe the context in which the study was carried out. This result highlights the need for more focus by researchers on describing the context of their studies. Agile/lean development and small/medium-size enterprise are the most common context in the literature. Both contexts are linked and mutually promoted, because SMEs constitute the majority of software development organizations (e.g., 99% of all European businesses are SMEs [S241]) and Agile/Lean development is more appropriate in the case of a small development team, project, or product [S245].

CMMI and ISO/IEC 15504 are the main maturity standards found in recent years, which indicates that organizations are still adopting these standards. Moreover, there is a lack of studies focusing on the measurement of the open source development, outsourcing, and test-driven development, which highlights these contexts for more investigations in the future.

In the recent period appear studies related to the global software development context, which demonstrates the correlation between the evolution of the development process and the measurement research.

9 CONCLUSIONS AND FUTURE WORK

This systematic mapping study focuses on the measurement in the context of software development process. It is an endeavor to structure the area and to provide categorized references to be used as a starting point for more detailed studies in this domain. Included were 462 studies; those studies were divided into three time periods and four topics based on the main focus to facilitate

the study and to identify trends. Five abstractions and 64 attributes were identified, and 25 methods and 17 contexts were distinguished.

The majority of the identified studies were published in conferences. The study results demonstrate that the solution proposal was the main research type. About 56% of these proposals were validated. This result highlights the necessity for more focus on validating the proposals. The industrial experience and the case studies were the main methods used to validate the proposals.

Process capability, performance, and maturity were the most measured process attributes, while project effort, performance, and risk were the most measured project attributes. The results also show the necessity of more focus on the process (agility, security, documentation, sustainability, and risks) and on the project (rework, reuse, communication, progress, and customer satisfaction). Further studies focusing on the developer are also required, as well as more studies focusing on the prediction and estimation.

GQM, CMMI, and SPC were the main base methods and models used in the studies. Data mining and analysis techniques were the most prominent methods. The results also show limited use of the international standards like ISO 15504 and 15939 against the industrial standards like CMMI and GQM.

The majority of the studies did not describe the research context (71%), which highlights the need for more emphasis on describing the context of the study by researchers. Agile/lean development and small/medium-size enterprise were the most frequently identified research contexts.

For researchers and practitioners, this study allows better understanding of the state of the art of development process measurement, as a first step for further studies. Future research can be oriented toward a specific topic, entity, attribute, method, model, or context to cover the aforementioned necessities. Future work could also be oriented to specific types of research, such as surveys to explain the results of this study, e.g., why there is scarcity of studies related to specific topic, method, or context. Also, conducting comparisons between the different measures, proposals, and methods in specific contexts could support researchers and practitioners to choose between them.

ELECTRONIC APPENDIXES

—The primary studies related to this article can be found at http://iwt2.org/wp-content/uploads/2017/10/ELECTRONIC-APPENDIX-Primary-Studies.pdf.

—Spreadsheets used in the mapping study can be found at http://iwt2.org/wp-content/uploads/2017/10/Spread_Sheets.rar.

REFERENCES

Abreu Fernando Brito and Rogério Carapuça. 1994. Object-oriented software engineering: Measuring and controlling the development process. In *Proceedings of the 4th International Conference on Software Quality*. 1–8.

J. Bailey, D. Budgen, M. Turner, B. Kitchenham, P. Brereton, and S. Linkman. 2007. Evidence relating to object-oriented software design: A survey. In *Proceedings of the 1st International Symposium on Empirical Software Engineering and Measurement (ESEM'07)*. 482–484. DOI: https://doi.org/10.1109/ESEM.2007.58

George I. Balch. 1974. Multiple indicators in survey research: The concept "sense of political efficacy." *Polit. Methodol.* (1974), 1–43.

Soon K. Bang, Sam Chung, Young Choh, and Marc Dupuis. 2013. A grounded theory analysis of modern web applications: Knowledge, skills, and abilities for DevOps. In *Proceedings of the 2nd Annual Conference on Research in Information Technology*. 61–62.

Sandhiprakash Bhide. 1990. Generalized software process-integrated metrics framework. *J. Syst. Softw.* 12, 3 (1990), 249–254.

Mario Bourgault, E. Lefebvre, Louis A. Lefebvre, Robert Pellerin, and Elie Elia. 2002. Discussion of metrics for distributed project management: Preliminary findings. In *Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS'02)*.

G. Cugola and C. Ghezzi. 1998. Software processes: A retrospective and a path to the future. In *Software Process Improvement and Practice*. 101–123.

Tore Dybå, Torgeir Dingsøyr, and Geir Kjetil Hanssen. 2007. Applying systematic reviews to diverse study types: An experience report. In *Proceedings of the International Symposium on Empirical Software Engineering and Measurement (ESEM'07)*. 225–234.

C. Ebert, R. Dumke, M. Bundschuh, and A. Schmietendorf. 2007. Best practices in software measurement: How to use metrics to improve project and process performance. Springer Berlin Heidelberg, 1–295.

Lem O. Ejiogu. 1993. Five principles for the formal validation of models of software metrics. *ACM SIGPLAN Not.* 28, 8 (1993), 67–76.

Norman E. Fenton. 1991. *Software Metrics: A Rigorous Approach*. Chapman & Hall.

Norman E. Fenton and Shari Lawrence Pfleeger. 1996. *Software Metrics: A Rigorous and Practical Approach* (2nd ed.). International Thomson Computer Press, London.

Norman Fenton and Barbara Kitchenham. 1991. Validating software measures. *Softw. Testing, Verif. Reliab.* 1, 2 (1991), 27–42.

Alfonso Fuggetta. 2000. Software process: A roadmap. In *Proceedings of the Conference on the Future of Software Engineering*. 25–34.

Félix García et al. 2006. Towards a consistent terminology for software measurement. *Inf. Softw. Technol.* 48, 8 (2006), 631–644.

Hanna Oktaba, Mario Piattini, Gómez Oswaldo, and Félix García. 2006. A systematic review measurement in software engineering: State-of-the-art in measures. In *Proceedings of the International Conference on Software and Data Technologies*. 165–176.

Naji Habra, Alain Abran, Miguel Lopez, and Asma Sellami. 2008. A framework for the design and verification of software measurement methods. *J. Syst. Softw.* 81, 5 (2008), 633–648.

Watts S. Humphrey. 1988. Characterizing the software process: A maturity framework. *IEEE Softw.* 5, 2 (1988), 73–79.

Watts S. Humphrey. 1989. *Managing the Software Process*, Addison-Wesley.

Kanhaiya Jethani. 2013. Software metrics for effective project management. *Int. J. Syst. Assur. Eng. Manag.* 4, 4 (2013), 335–340.

B. Kitchenham and S. Charters. 2007. Guidelines for performing systematic literature reviews in software engineering. EBSE Technical Report, version 2.3.

Barbara Kitchenham. 2004. Procedures for performing systematic reviews. *Keele, UK, Keele University* 33, 2004 (2004), 1–26.

Barbara Kitchenham and Shari Lawrence Pfleeger. 1996. Software quality: The elusive target. *IEEE Softw.* 13, 1 (1996), 12.

Barbara Kitchenham, Shari Lawrence Pfleeger, and Norman Fenton. 1995. Towards a framework for software measurement validation. *IEEE Trans. Softw. Eng.* 21, 12 (1995), 929–944.

B. Lennselius, C. Wohlin, and C. Vrana. 1987. Software metrics: Fault content estimation and software process control. *Microprocessors and Microsystems* 11, 7 (1987), 365–375.

Mikael Lindvall, Paolo Donzelli, Sima Asgari, and Vic Basili. 2005. Towards reusable measurement patterns. In *Proceedings of the 11th IEEE International Software Metrics Symposium (METRICS'05)*. IEEE Computer Society, Washington, DC. 21. DOI: https://doi.org/10.1109/METRICS.2005.49

Daniel Lowry Lough. 2001. *A Taxonomy of Computer Attacks with Applications to Wireless Networks*. Ph.D. Disertation. Virginia Polytechnic Institute and State University.

K. Petersen and C. Wohlin. 2009. Context in industrial software engineering research. In *Proceedings of the 3rd International Symposium on Empirical Software Engineering and Measurement (ESEM'09)*. 401–404. DOI: https://doi.org/10.1109/ESEM.2009.5316010

Kai Petersen, Robert Feldt, Shahid Mujtaba, and Michael Mattsson. 2008. Systematic mapping studies in software engineering. In *Proceedings of the 12th international Conference on Evaluation and Assessment in Software Engineering*.

Martin Shepperd. 1995. *Foundations of Software Measurement*. Prentice Hall.

Ian Sommerville. 2004. *Software Engineering* (7th ed.). Pearson Addison Wesley.

Maarit Tihinen, Päivi Parviainen, Rob Kommeren, and Jim Rotherham. 2012. Metrics and measurements in global software development. *International Journal on Advances in Software* 5 (2012), 278–292.

Roel Wieringa, Neil Maiden, Nancy Mead, and Colette Rolland. 2006. Requirements engineering paper classification and evaluation criteria: A proposal and a discussion. *Requir. Eng.* 11, 1 (2006), 102–107. DOI: https://doi.org/10.1007/s00766-005-0021-6

Claes Wohlin, Per Runeson, Martin Höst, Magnus C. Ohlsson, Björn Regnell, and Anders Wesslén. 2000. *Experimentation in Software Engineering: An Introduction*. Kluwer Academic Publishers, Norwell, MA.