The Journal of Systems and Software 123 (2017) 45–63

Contents lists available at ScienceDirect

The Journal of Systems and Software

journal homepage: www.elsevier.com/locate/jss

Approaches to strategic alignment of software process improvement: A systematic literature review

Francisco J.S. Vasconcellos[a,*], Geraldo B. Landre[a], José Adson O.G. Cunha[b], Juliano L. Oliveira[c], Ronaldo A. Ferreira[a], Auri M.R. Vincenzi[d]

[a] College of Computing, Federal University of Mato Grosso do Sul, Campo Grande, MS, Brazil
[b] Centre of Informatics, Federal University of Pernambuco, Recife, PE, Brazil
[c] Institute of Informatics, Federal University of Goiàs, Goiânia, GO, Brazil
[d] Computing Science Department, Federal University of São Carlos, São Carlos, SP, Brazil

ARTICLE INFO

Article history:
Received 30 January 2016
Revised 19 September 2016
Accepted 19 September 2016
Available online 22 September 2016

Keywords:
Software process improvement
Strategic alignment
Business alignment
Systematic literature review

ABSTRACT

Context: Software process improvement (SPI) aims to increase the effectiveness of a software organization. Many studies indicate that the strategic alignment is a critical factor for the SPI success. However, little is known about practical approaches to achieving and maintaining such alignment.

Objective: The goal of this study is to evaluate the validation evidence of the existing approaches to the strategic alignment of SPI.

Method: We develop a search protocol that combines database search and snowballing to perform the systematic literature review and evaluate empirical studies by applying rigor and relevance criteria. To evaluate the efficiency of our protocol, we use a *"quasi-gold standard"* to compute the sensitivity and precision of the search.

Result: We identified 30 studies (18 empirical) and 19 approaches to strategic alignment of SPI from 495 retrieved studies. Only three out of the 18 empirical studies were rated as high in the categories rigor and relevance, suggesting the need for a stronger validation of the approaches.

Conclusion: We conclude that the lack of empirical validation indicates that the results of the existing approaches have not been adequately transferred to practitioners yet, calling for more rigorous studies on the subject.

1. Introduction

Software organizations need to improve their work processes continuously to meet the growing demands for products and services. Many standards and models have been proposed to assist in process improvement strategies, but the selection of the practices to be improved depends on the organizational context, culture, and priorities (Humphrey, 1989). More important, the success of a software process improvement (SPI) initiative hinges on the business results that it can generate (Ebert, 1999), so an SPI initiative needs to be aligned with the organization's business goals. In fact, many studies (Bayona et al., 2012; Dybå, 2000, 2003, 2005) show that the alignment between SPI and business goals is a critical factor for the success of an SPI initiative. For example, Bayona et al. (2012) list the alignment with the business strategy and goals as one of 16 critical factors for the implementation success of process improvement; and Dybå's studies (Dybå, 2000, 2003, 2005) stress that business orientation is one of the key facilitating factors for SPI success.

Both for-profit and non-profit software organizations have organizational goals defined by the senior managers. Improvements in a software process can be derived from the organizational goals in a top-down SPI planning (ISO, 2013) or aligned with the business goals in a bottom-up SPI strategy as in an agile environment (Ringstad et al., 2011; Qumer and Henderson-Sellers, 2008; Lepmets and McBride, 2012). Regardless of the scenario, an SPI initiative invariably demands changes in a software process. Münch et al. (2012) emphasize that changes in software processes must be carefully planned because they rely on humans to be effective, depend on the organizational and business context, and represent

* Corresponding author.
E-mail addresses: francisco.vasconcellos@ufms.br (F.J.S. Vasconcellos), geraldo@facom.ufms.br (G.B. Landre), jaogc@cin.ufpe.br (J.A.O.G. Cunha), juliano@inf.ufg.br (J.L. Oliveira), raf@facom.ufms.br (R.A. Ferreira), auri@dc.ufscar.br (A.M.R. Vincenzi).

http://dx.doi.org/10.1016/j.jss.2016.09.030
0164-1212/

significant costs for an organization. Thus, any change in a software process can be seen as a strategic decision. As the alignment of SPI with business goals involves changes in a software process and is a strategic decision, we call it the *strategic alignment of SPI*.

Despite the importance of strategic alignment of SPI and the existence of several approaches to such alignment (Barreto and Rocha, 2010; Becker et al., 2008; Debou and Kuntzmann-Combelles, 2000; Guzmán et al., 2010; Trienekens et al., 2005; Albuquerque et al., 2009), Lepmets et al. Lepmets et al. (2012) emphasize that it is still not clear how to achieve the strategic alignment in practice. Kuhrmann et al. (2015) go beyond and state of that the SPI research field is being shaped by numerous proposals that provide no evidence on their feasibility, advantages, and disadvantages. The technology transfer model proposed by Gorschek et al. (2006) and Gorschek (2015) defines the steps for validating the outcome of research (i.e., a candidate solution) before its release to practitioners. The model stresses the need for an industry validation, in addition to the academic validation, to provide evidence of the usability and scalability of any candidate solution. We argue that a proper validation of any proposed approach to the strategic alignment of SPI should follow this technology transfer model.

The conjecture we want to investigate with this systematic literature review (SLR) is that there is insufficient evidence of the validation of the proposed approaches to the strategic alignment of SPI. To evaluate the gap between the research proposals and the scientific evidence of strategic alignment of SPI, we review the literature and follow the evidence-based software engineering (EBSE) proposed in Kitchenham et al. (2004) and Dybå et al. (2005) to track down the evidence that can confirm or refute our conjecture. EBSE tries to close the gap between research and practice by stimulating methodological rigor of empirical studies while emphasizing their relevance to practice (Dybå and Dingsøyr, 2008). Our SLR modifies the checklist for rigor-relevance assessment of empirical studies presented by Ivarsson and Gorschek (2010), adapting some criteria to assess the empirical evidence of existing approaches to strategic alignment of SPI.

Following a brief background on strategic management, information technology governance, and decision-making processes for strategic alignment of SPI in Section 2, we detail our method to answer the research questions in Section 3. Our search protocol applies a combination of two well-known search strategies: database search and snowballing. A broader database search defines a snowballing start set to mitigate the risks of researchers' bias. Our protocol also includes quality assurance activities, such as audits by independent researchers and measurement of an inter-observer agreement, to enhance the reliability of our results. These design strategies and quality assurance activities contribute to making our final dataset a strong candidate for a "gold standard", as defined by Zhang et al. (2011).

We answer the research questions in Section 4 based on an extensive analysis of the results. First, we present the results of the search protocol, the classification of the studies in theoretical or empirical, and the values of the rigor and relevance assessment of the empirical studies. Second, we show the distribution over time of the selected studies and the citation graph. The distribution over time contributes to an analysis of the evolution of the research on the subject while the citation graph assists the analysis of influence among approaches by the visualization of the citation-graph connectivity. Third, we present the scores of rigor and relevance of the studies using a bubble chart, which allows a visual analysis of the overall quality of the empirical studies. Fourth, we evaluate the search protocol using the values of sensitivity and precision of the search before and after the snowballing to quantify the snowballing contribution to the strategy efficiency.

In Section 5, we discuss our findings, raise open questions about the SLR results, and discuss the limitations of our study through an analysis of the threats to the validity of our method and results. We also describe implications of our results to practitioners and researchers. We conclude the paper in Section 6 by proposing a research agenda with the challenges that must be met to offer practical support for the strategic alignment of SPI. The list of the selected studies and the criteria used for the rigor-relevance assessment, are presented in the Appendices. In summary, the paper makes the following contributions:

1. A systematic literature review (SLR) protocol that combines database search with snowballing procedures to balance sensitivity (recall) and precision (Section 3).
2. A dataset of existing approaches to strategic alignment of SPI that can be used as a quasi-gold standard in future studies (Section 4).
3. An assessment of the methodological rigor and relevance to practice of the identified empirical studies (Section 4).
4. A discussion of possible reasons for our findings that raises open questions for future studies (Section 5).
5. A proposal, based on the identified gaps of the existing approaches, of a research agenda with the challenges that must be met to offer practical support to the strategic alignment of SPI (Section 6).

2. Background

The objectives of the processes of an organization depend on the business goals, needs, and constraints. Hence, a process improvement initiative must focus on the overall organization objectives. In this section, we place SPI in the context of information technology governance, present concepts about the dynamic of strategic management and decision-making processes, and review strategies for selecting and implementing an SPI initiative. These concepts are important for understanding our criteria used for the data-extraction design (presented in Section 3) and the attributes for rigor and relevance assessment (described in Appendix B).

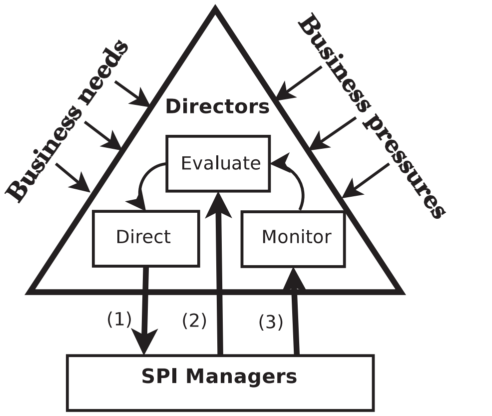
2.1. Information technology governance and strategic alignment

Information technology (IT) governance is the system by which IT is directed and controlled, as defined by the ISO standard 38500:2015 (ISO, 2015). The standard deals with IT management in general, but it recommends the delegation of specific responsibilities to functional managers. We extend the standard to incorporate the role of SPI manager, as illustrated in Fig. 1. The figure, adapted from the ISO standard, shows how the SPI managers interact with the governance body of the organization and highlights the important variables of the SPI governance, which are:

1. two primary stakeholders: directors and SPI managers;
2. business needs and pressures; and
3. three main tasks: Evaluate, Direct, and Monitor.

The *directors* are the people who are accountable for the performance and conformance of the organization. In a software organization, the role of director is performed by its *owners* or *senior managers*. The directors make decisions assisted by *SPI managers*, who are experts in software processes and are aware of the business goals. For an SPI initiative to be successful, the directors and SPI managers must work together to align the SPI initiative with the business goals.

The *business pressures and needs* are important variables because of their impact on business processes, projects and products. In fact, the SPI goals can be derived from the business pressures (e.g., via customer relationship management), or from the business



(1) Directors promulgate policies and strategic goals.
(2) SPI managers elaborate SPI proposals and plans.
(3) SPI managers provide performance data of the SPI initiatives.

Fig. 1. SPI Governance. Adapted from Model for governance of IT (ISO, 2015).

needs (e.g., via project goals and software product requirements). An organization must also consider the business needs to establish and maintain its competitive advantage.

In the *evaluate* task, the directors need to consider pressures that are acting upon the business, such as technological issues, economic trends, and social or political influences for the evaluation of current and future use of IT. In the *direct* task, the directors promulgate policies and strategic goals and assign the tasks for the design and implementation of the SPI plan. The SPI managers provide to the directors the performance data collected from the measurement systems. As the directors are accountable for the SPI results, they *monitor* the SPI performance to ensure that the SPI is following the business objectives.

The two primary stakeholders presented by the ISO 38500:2015 have a straight relationship with the roles of governance and management presented in one of the five principles of the COBIT governance model (ISACA, 2012). The COBIT principle "*Separating Governance From Management*" defines *governance* as a discipline that establishes enterprise objectives, sets direction through prioritization, and monitors performance and compliance against objectives. On the other hand, *management* is defined as a discipline responsible for planning, executing and monitoring activities in alignment with the direction set by the governance body. These two roles define two levels of decision with distinct concerns: the strategic and the tactical. The strategic level needs to respond quickly to external changes, while the tactical level needs to analyze what should be done to meet the established strategic objectives. Mintzberg (1994) considers this subtle difference of perspective between strategic and tactical levels a challenge for the alignment of the strategic goals with the tactical plans.

2.2. The dynamic of strategic management and decision-making process

The strategy has its origins in the military and its epistemological origin defines it as "the art of the general" and the general is a person who defines and controls objectives to be pursued by an army, i.e., who evaluate the situation, direct by objectives and monitor the performance of implementations of tactics. The tactics are the means to achieve strategic objectives through operational actions (Luecke, 2005). The theory of *strategic planning* was incorporated into business administration by Ansoff (1965). Strategic planning is a process that allows repetition of actions that contribute to the success of organizations or, at least, avoid failures.

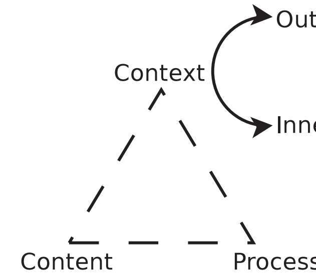


Fig. 2. Pettigrew's framework, by Pettigrew (adapted) (Pettigrew, 1987).

Strategic planning defines goals and plans based on the analysis of Strengths, Weaknesses, Opportunities, and Threats (SWOT) (Pickton and Wright, 1998).

Pettigrew (1987) presented three necessary dimensions to develop, change, and implement strategic plans: context, content, and process (Fig. 2). Developing a new strategy entails managing its context and process. The *context dimension* establishes the reason for an organization to engage in process improvement (*why*). The *outer context* refers to the environment in which the organization operates, whereas the *inner context* refers to the structure, organizational culture, and political context within the company that the strategy has to be implemented. The *content dimension* relates to the organizational aspects that need to be changed (*what*) and the *process dimension* refers to the actions and interactions with various stakeholders as they intend to improve the organization (*how*).

Strategic planning through the SWOT analysis (Pickton and Wright, 1998) works on the inner and outer contexts of the Pettigrew's framework. The analysis of the inner context examines two characteristics of each organization: its strengths and its weaknesses. The analysis of the outer context harnesses opportunities to take advantage of the situation in which the organization operates and to mitigate threats that can jeopardize actions in the future. Besides the SWOT analysis, Hussey (1997) describes other techniques for strategic analysis of an organization and for defining the strategic objectives. However, unsuccessful implementations of planned strategies demonstrate the difficulty in enactment and maintenance of programs arising from strategic planning, as pointed by Kaplan and Norton (2008).

According to Mintzberg (1994), the success of a strategy is not limited to the plan, so the research on strategic planning should focus on the control and management of the strategy. After all, in any business sector, dynamics of actions bring desired and undesired (but necessary) changes in strategy. Planned programs fail to represent strategic value because new options emerge throughout their execution, offering what Mintzberg and Waters (Mintzberg and Waters, 1985) call "*emergent strategy*", which is illustrated in Fig. 3.

The dynamic of strategic management implies constant evaluation and decisions on the previously deliberated alternatives. An emergent strategy can take the place of a deliberate strategy, which becomes an unrealized strategy. This realignment usually implies several costs and the waste of time because of the changes of the actions in progress. Likewise, Eisenhardt and Zbaracki (1992) define a strategic decision as one that is relevant regarding the actions to be taken, the resources needed or the organizational context.

In the context of the software organizations, Münch et al. (2012) state that software process changes are human based and context dependent, representing significant costs for an organization. Therefore, SPI programs have a straight relationship with the strategic decision-making.

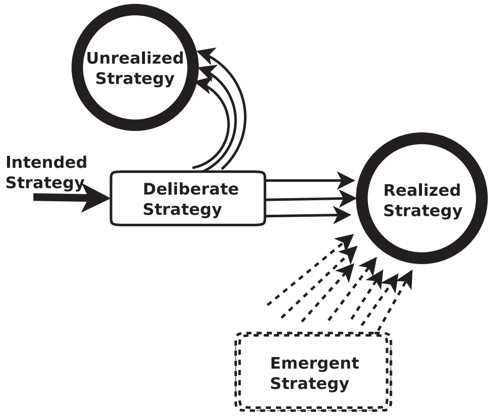


Fig. 3. Synthesis of deliberate and emergent strategies. Adapted from Mintzberg and Waters (1985).

2.3. Software process improvement strategies

Watts Humphrey (1989) argue that software quality can be improved by improving its development process. Following Humphrey's principles, McFeeley (1996) defines the IDEAL model for establishing and carrying out an SPI program. In the IDEAL model, there are two important components for software process improvement that need to be considered: a strategic and tactical ones. The *strategic component* provides directions and prioritization of tactical activities based on the organization's business goals and drivers. In the IDEAL's initiating phase, senior management must determine the business goals and rationale for an SPI program and establish the urgency for those programs.

Generally, organizations seek to implement SPI because of different business goals (ISO, 2004):

- to enhance the quality of the product;
- to minimize wasted time and effort;
- to reduce time to market; and
- to provide predictability.

The mentioned above business goals direct decisions related to process improvement programs, such as the selection of practices to improve as well as the selection of the SPI approach to implement the practices. Münch et al. (2012) define two SPI approaches: *model-based SPI approaches (top-down)* and *continuous SPI approaches (bottom-up)*. The model-based approaches compare organizational processes with a reference model and uses an SPI assessment to identify gaps and potential improvement options. Examples of model-based SPI approaches are CMMI (Chrissis et al., 2011) and ISO 15540 (ISO, 2004). The continuous SPI approaches address specific solutions for identified problems and evaluate specific effects of improvement programs. Münch et al. cite the Quality Improvement Paradigm (QIP) (Basili, 1989) and the Total Quality Management (TQM) (Martínez-Lorente et al., 1998) as examples of continuous SPI approaches. Salo and Abrahamsson (2005) also present the agile SPI as a continuous SPI approach.

The better the alignment between business objectives and process improvement decisions, the greater the likelihood of success of the SPI program (ISO, 2013). Although some proposed SPI approaches point out the importance of business alignment of SPI programs, as far as we know, none describes how to obtain such alignment.

Trienekens et al. (2005) emphasize the central role of the decision-making process for decomposing business strategy and goals into process and team goals. They present a three-level decision: strategic (Why), tactical (What), and operational (How). The strategic decision is related to the prioritization of business goals. The SPI managers are faced with the tactical decisions when they have to select one SPI alternative that better aligns with the strategic decisions made by the directors. The operational decision is related to the choice of tools, procedures, and templates used to support the implementation of an SPI program.

3. Research method

To determine whether a review similar to ours has already been published, we searched the fields keywords, title, and abstract of peer-reviewed publications in the ACM, Science Direct, Compendex and Scopus digital libraries using the following string:

```
(
 ("software process improvement") AND
 ("business" OR "goal") AND
 ("alignment" OR "aligning") AND
 ("systematic literature review" OR "literature review" OR
  "systematic review" OR "mapping study" OR "scoping study")
)
```

The two systematic reviews retrieved are Bayona et al. (2012) and Ullah and Lai (2013), but they do not focus on SPI alignment approaches.

Our research method uses the guidelines provided in Wohlin (2014), Kitchenham and Charters (2007) and Higgins and Green (2011) and the recommendations given in Petersen et al. (2015) and Kitchenham and Brereton (2013), and consists of the following stages, with their respective activities:

I. Planning the review:
  - Specification of research questions; and
  - Development of the search strategy.

II. Conducting the review:
  - Database search and definition of the snowballing start set;
  - Execution of snowballing and data extraction;
  - Evaluation of the protocol;
  - Rigor-relevance assessment of empirical studies; and
  - Synthesis and discussion of issues.

III. Reporting the review.

3.1. Research questions

The overall objective of our systematic literature review is to identify approaches to strategic alignment of software process improvement to obtain a more detailed and comprehensive view on the topic. Furthermore, we seek for usage evidence to assess empirical studies through rigor and relevance.

The Research Questions (RQ) are defined as:

RQ 1. What are the available approaches to strategic alignment of SPI?

RQ 1.1. For each identified approach, is there a method or procedure for dynamic realignment of SPI initiatives in the case of changing business goals (emergent strategies)?

RQ 1.2. For each identified approach, is there a systematic method or procedure to support decision making on both levels (strategic and tactical)?

RQ 2. What is the evidence of the use of these approaches?

RQ 2.1 What are the rigor values of the reported empirical studies?

RQ 2.2 What are the relevance values of the reported empirical studies?

3.2. The search strategy

The following steps are executed to answer the research questions:

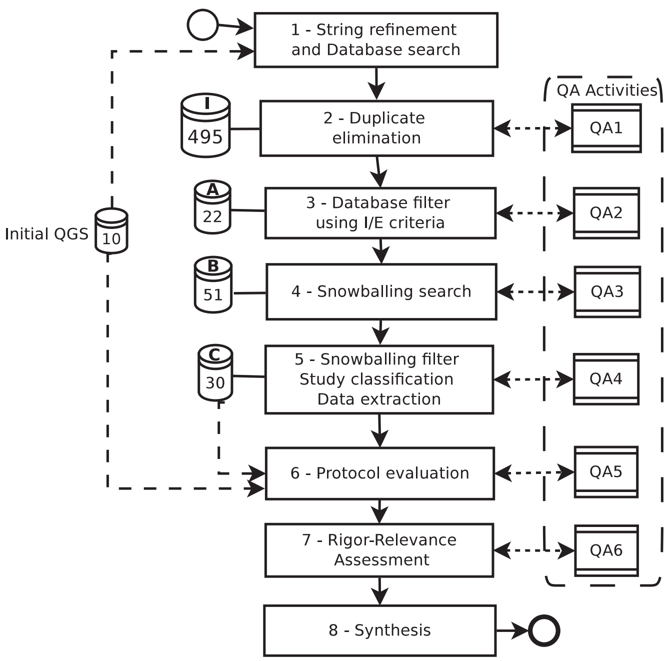


Fig. 4. The protocol. The letter on top of each dataset figure identifies its evolution: I-Initial dataset; A-Filtered dataset; B-Post-snowballing dataset; and C-Final dataset. The number inside of each figure of a dataset represents the dataset size. Dashed lines represent dataset usage as an input of a step while solid lines represent a step dataset output. QA activities are planned after each step to address threats to validity and are surrounded by a dashed line to show independence of the auditors. The dotted lines with arrows illustrate the iterative feedback between the authors and the auditors. The Protocol evaluation step aims to offer an independent check of the performed procedures and analysis and to review the sensitivity and precision values of the search.

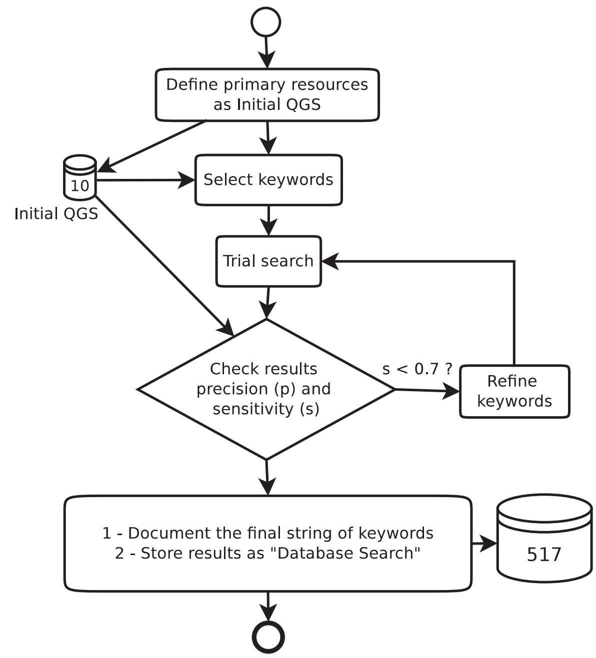


Fig. 5. The string refinement procedure. A dataset with 517 studies was obtained as a result of keyword refinement performed until reaching a 0.7 sensitivity value.

1. Database search and extraction of studies using search strings.
2. Duplicate elimination.
3. Dataset filter by title, keywords, and abstract analysis using inclusion-exclusion criteria.
4. Snowballing search that applies both types of snowball:
   Backward Snowball (BSB) consisting of scanning the reference list of each study.
   Forward Snowball (FSB) consisting of the selection of citations retrieved by Google Scholar.
5. Snowballing filter, study classification and data extraction.
6. Protocol evaluation that measure search sensitivity and precision.
7. Rigor-relevance assessment of empirical studies.
8. Synthesis.

Fig. 4 shows the workflow with all the steps described above, including Quality Assurance (QA) activities to deal with the possible threats to the validity of this research. Each QAx activity is described later in this section. As recommended by Kitchenham and Brereton (2013), we use a dataset of known studies, called Initial Quasi-Gold Standard (Initial QGS), to define the search string. We also use our Final dataset to evaluate our protocol execution.

3.2.1. Database search, duplicate elimination and database filtering

We use the PICOC criteria (Population, Intervention, Comparison, Outcome and Context) (Kitchenham and Charters, 2007; Petticrew and Roberts, 2006) and ten already known studies as the Initial QGS (Zhang et al., 2011; Kitchenham et al., 2011) to select keywords and to compose the search string. However, we do not include the "Comparison" part of the PICOC because the main objective of this literature review is not to retrieve evidence of comparison among approaches. The PICOC criteria derive keywords as following:

- Population: **Software Process Improvement** initiatives
- Intervention: **Strategic** (OR **Business goal**) AND **alignment** AND **approach**
- Comparison: not included
- Outcomes: **Business oriented** SPI
- Context: any

The string derived from the PICOC criteria is refined with identified synonyms and alternative spellings (Kitchenham and Brereton, 2013) through a series of pilot searches (Fig. 5). A trade-off between sensitivity (also known as recall (Manning et al., 2008)) and precision (Higgins and Green, 2011; Zhang et al., 2011) is expected. *Sensitivity* is defined as the number of identified relevant reports divided by the total number of existing relevant reports and *precision* is defined as the number of identified relevant reports divided by the total number of identified reports. Increasing the sensitivity of a search reduces its precision and retrieves more non-relevant primary studies (Higgins and Green, 2011; Kitchenham et al., 2011).

We do not use the keyword "initiatives" nor the SPI acronym because it only decreases precision without an increase of the sensitivity value. In our protocol, the only purpose of the String refinement and Database search step (Step 1 in Fig. 4) is to establish the start set for the Snowballing search step. As we do not expect high precision in this initial search, the String refinement and Database search step ends when the search reaches a 70% of sensitivity and 1.4% of precision values as illustrated in Fig. 5. A better precision value is later expected with our snowballing complementary strategy.

The search string defined after the execution of the string refinement procedure is the following:

Table 1
The Initial Quasi-Gold Standard list.

QGS Id	QGS studies
QGS I	C. Debou, A. Kuntzmann-Combelles, **Linking software process improvement to business strategies: experiences from industry**, Software Process: Improvement and Practice 5 (1), 2000.
QGS II	J. J. Trienekens, R. J. Kusters, B. Rendering, K. Stokla, **Business-oriented process improvement: Practices and experiences at Thales Naval The Netherlands (TNNL)**, Information and Software Technology 47, 2005.
QGS III	V. Basili, J. Heidrich, M. Lindvall, J. Münch, M. Regardie, D. Rombach, C. Seaman, A. Trendowicz, **Bridging the gap between business strategy and software development**, ICIS 2007, Proceedings, 2007
QGS IV	A. L. Becker, R. Prikladnicki, J. L. N. Audy, **Strategic alignment of software process improvement programs using QFD**, in: 1st Int'l Workshop on Business impact of process improvements, ACM, 2008.
QGS V	A. B. Albuquerque, A. R. Rocha, A. C. Lima, **Software process improvement: Supporting the linking of the software and the business strategies**, in: Product-Focused Software Process Improvement, Vol.32, Springer, 2009.
QGS VI	J. J. Trienekens, R. Kusters, D. Kriek, P. Siemons, **Entropy based software processes improvement**, Software Quality Journal 17 (3), 2009.
QGS VII	A. O. S. Barreto, A. R. Rocha, **Defining and monitoring strategically aligned software improvement goals**, in: Product-Focused Software Process Improvement, Vol. 6156, Springer, 2010.
QGS VIII	V. R. Basili, M. Lindvall, M. Regardie, C. Seaman, J. Heidrich, J. Münch, D. Rombach, A. Trendowicz, **Linking software development and business strategy through measurement**, Computer 43 (4), 2010.
QGS IX	J. G. Guzmán, H. A. Mitre, A. Amescua, M. Velasco, **Integration of strategic management, process improvement and quantitative measurement for managing the competitiveness of software engineering organizations**, Software Quality Journal 18 (3), 2010.
QGS X	Y. Sun, X. F. Liu, **Business-oriented software process improvement based on CMMI using QFD**, Information and Software Technology 52 (1), 2010.

```
(
 ("software process improvement") AND
 ("business goal" OR "strategic" OR "goal oriented" OR
  "business oriented" OR "business strategy") AND
 ("alignment" OR  "in line with" OR "geared to" OR
  "aligned with" OR "linking") AND
 ("method" OR "approach" OR "framework" OR "methodology")
)
```

The IEEE Xplore Advanced Search returned more than two million studies with the default string search. We adapted the default string for the IEEE Xplore due to the lack of precision showed. The decrease of the sensitivity caused by this adaptation does not configure a problem because 100% of sensitivity is expected because of our snowballing complementary strategy.

The String adapted for IEEE Xplore:

((“software process improvement”) AND (“business goal”) AND (“alignment” OR “in line with”) AND (“method” OR “approach”))

The following list of known papers is used as the Initial Quasi-Gold Standard:

The database search retrieved seven of the 10 QGS studies (70% sensitivity previously mentioned). The QGS I, III, and X were not retrieved and we proceed to the next protocol step. As mentioned, 100% of sensitivity is expected after the snowballing procedures (Table 1).

The search for primary studies is based on these reference databases: Springer, Scopus, Web of Science, Science Direct, Compendex, IEEE Xplore, and ACM Digital Library. The database search was conducted in August 2015, without any set limits regarding publication year. In July 2016, we performed the Snowballing-search step (Step 4 in Fig. 4)again using the Final dataset as the start set to confirm or update the research data. But, the snowballing search did not return any new study. The following lists of URLs and command types were used:

- IEEE Xplore (Advanced search-Command Search): http://ieeexplore.ieee.org.
- Springer (Advanced Search): http://link.springer.com/advanced-search.
- ACM (Advanced Search): http://dl.acm.org.
- Web of Science (Basic Search): http://apps.webofknowledge.com.
- SCOPUS (Advanced Search) using codes TITLE-ABS-KEY: http://www.scopus.com.
- El Compendex (Expert Search): http://www.engineeringvillage.com.
- Science Direct (Expert Search); http://www.sciencedirect.com.

The Database search and the Duplicate elimination steps build the **Initial dataset** (**I**). Exclusions done by applying exclusion criteria results in the **Filtered dataset** (**A**), which constituted the start set (Wohlin, 2014) used to spot more relevant papers by snowballing. An independent researcher reviews the string usage and the search results, therefore representing the execution of QA1.

3.2.2. Inclusion-exclusion criteria

The first author formulated the inclusion-exclusion criteria that are reviewed for validation by the other authors in the Planning stage. Studies are included in the dataset only if they receive a positive answer to any of the following questions:

- Does the paper discuss SPI and business alignment approaches?
- Does the paper contain evidence of SPI and business alignment?
- Does the paper discuss a particular case of SPI and business alignment?

Subsequently, exclusion criteria are applied excluding papers that:

- are not primary studies;
- are not in English;
- are related to business process improvement or IT alignment in general but not to SPI alignment in particular; or
- are related to SPI but not discuss a business alignment approach. Some studies may offer a measurement approach to SPI, emphasizing its potential use for business alignment, but they are excluded if do not present an alignment procedure.

The first and the third authors applied the rule presented by Badampudi et al. (2015) for the inclusion-exclusion decision process. The rule states that paper is included if either one of the authors accepts it. An invited researcher reviews the performed inclusion-exclusion checking in a random sampling if the mentioned criteria were proper applied. This independent audit represents the execution of QA2.

3.2.3. Snowballing search

The snowballing search strategy becomes significant when a string refinement attempt meets with difficulties due to the lack of standardized keywords. This difficulty is related to the research topic and has an impact in the trade-off analysis between sensitivity and precision of the search.

Unlike the guidelines proposed by Wohlin (2014), a broader database search is planned to establish the start set of papers for the snowballing search. Our protocol does not use the Google Scholar for the start set definition because of its lower precision, possibly caused by gray literature retrieval.

Snowballing is iterative and executed until no new papers are found, generating the **Post-snowballing dataset** (**B**). Backward snowballing (BSB) and Forward snowballing (FSB) are executed in two stages. The first stage (BSB) uses the inclusion-exclusion criteria basing the decision on a review of:

1. The title of the referenced paper;
2. The reference context in the paper (reading the text where the reference is cited for context evaluation); and
3. The abstract of the referenced paper.

In the second stage (FSB) a search is carried out in Google Scholar for each study. The inclusion of citing studies obeys the following method:

1. The title of the citing paper is analyzed.
2. The abstract of the citing paper is reviewed.
3. If the items above are not sufficient, the full text is examined.

The studies included in each iteration are added to the dataset and snowballing (BSB and FSB) of the added studies is executed subsequently and iteratively. This procedure is followed until no new studies are found, thus closing the **Post-snowballing dataset (B)**.

3.2.4. Snowballing filtering and data extraction strategy

Step 5 aims to apply inclusion-exclusion criteria to papers in dataset B thus providing the **Final dataset** (**C**). Subsequently, the data needed to respond the research questions are extracted and empirical studies are identified. A full-text analysis is carried out to confirm the inclusion of each study and to extract data concomitantly. The inclusion-exclusion confirmation by a full-text analysis is necessary since there can be false positive inclusions in step 4, i.e. studies which attended the inclusion criteria by title/abstract/keywords analysis but have to be excluded from a full-text analysis.

Snowballing procedures are performed even for the false positives because, although excluded, those studies can bring good references. In our search, a recently published book by Basili et al. (see S.39 in Appendix A) was found through a false positive study (S.16) FSB. As further illustrated by the citation graph, this book could only be found in our search through snowballing false positive studies.

The following information is extracted from each analyzed study:

I. Research approach type. Approaches can be presented in **theoretical studies** (opinion or philosophical papers, experience reports, and proposals of solutions) or **empirical studies** (validation research and evaluation research) (Wieringa et al., 2006).
II. Approach name. If no name is given to an approach, the first author's name is used.
III. Dynamic realignment method. The method or procedure for an alignment revision in the event of changes of business goals.
IV. Decision-making method. The method or procedure supporting the decision-making process carried out in business goal prioritization (strategic level) and for SPI alternative selection (tactical level).
V. Year. The year of publication.

The data extraction properties mapped by research questions are discussed and defined in the Planning stage (see Table 2).

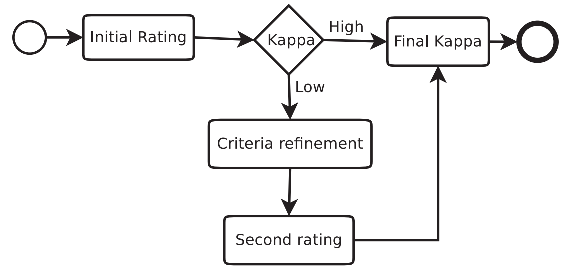


Fig. 6. The Quality Assurance activity of the rigor-relevance assessment consisting of an inter-observer agreement measurement.

The first author performed the snowballing search and filtering, as well as the data extraction, whose results are subsequently inspected by the second and the third authors, representing the quality assurance activities QA3 (concerning the snowballing search results) and QA4 (snowballing filtering and data extraction check). The third author complemented the QA4 activity by reviewing the research approach type classification. Disagreements were resolved in consensus meetings.

3.2.5. Protocol evaluation

Two independent researchers were asked to review the protocol. The protocol establishes the usage of questions, presented by Kitchenham and Charters (2007), as a checklist for this evaluation aiming to contribute to improving the SLR report through suggestions that could make research steps and procedures clear. Further insights could also be gained from data analysis and added to conclusions, contributing to the QA5 activity.

Another action planned for protocol evaluation is checking the sensitivity and precision values of the Final dataset against the values relating to the Initial QGS dataset, as proposed by Zhang et al. (2011). This search process effectiveness evaluation is also recommended by Kitchenham and Brereton (2013). In the Final dataset, a better trade-off between these indicators is expected due to the designed search strategy.

3.2.6. Rigor-relevance assessment

A checklist adapted from the work of Ivarsson and Gorschek (2010) is planned to assess rigor and relevance of the Final dataset empirical studies. The checklist adaptation was carried out by the first author defining specific criteria related to this literature review (see Appendix B). The criteria were validated in the Planning stage by the other authors to minimize subjectivity. Moreover, the first and the second authors rated the studies independently for quality assurance. The second author's ratings aim to measure the inter-observer agreement kappa index for a criteria refinement and for a final agreement evaluation (Fig. 6), representing the QA6 activity.

The QA6 activity consists of two double-ratings carried out independently by two researchers. The initial double-rating aims to establish if a criteria refinement has to be performed. A criteria refinement must be performed only if the kappa value represents a strength of agreement classified as less than Moderate, i.e., is inferior to 0.41 according to Landis and Koch (1977). If the strength of agreement measured after this initial rating is classified as Moderate or higher, no criteria refinement is planned, and the obtained kappa value is the Final Kappa index. In our case, the refinement was necessary due to different interpretation of the Rigor-Context and the Relevance-User/Subject attributes (see Appendix B). The category Rigor was improved by adding the contextual aspects proposed by Petersen and Wohlin (2009), whereas the category Relevance was improved by establishing the inclusion of software organization managers and SPI professionals as the in-

Table 2
Extracted data mapped by the research questions.

Extracted data	Properties	RQ mapping
Research approach type	Theoretical or Empirical	RQ 1 and RQ 2
Approach name	Name given to the approach or Name of the first author	RQ 1
Presence of a dynamic realignment method	Yes or No	RQ 1.1
Presence of a decision-making procedure or method for both levels	Method name or “none”	RQ 1.2
Publication Year	Year	RQ 1

Table 3
Search results.

Database	Search results
IEEE	43
Scopus	15
Springer	141
El Compendex	13
Web of science	10
Science direct	195
ACM	100
TOTAL	***517***
Initial dataset (I)	**495**

Table 4
Data set evolution.

Filtered dataset (A)	22
Post-snowballing dataset (B)	51
Excluded false positives	18
Not available	3
Final dataset (C)	**30**

tended users of the approach. The second double-rating aims to evaluate the rigor-relevance assessment process through another strength of agreement measurement (Final Kappa in Fig. 6). Even if the obtained kappa value represents a strength of agreement classified as less than Moderate, no criteria refinement is planned. This second kappa value becomes the Final Kappa index. The Final Kappa is a quality index aiming to evaluate a researcher bias in the rigor-relevance assessment step.

4. Results and analyses

The results and analyses are reported in this section with a distribution overview of the publication over time and study categorization based on the applied research method.

4.1. Database search results and duplicate elimination

The Database search step retrieved 517 studies and the Duplicate elimination step removed 22 studies considered duplicates. As a result, there were 495 studies in the Initial dataset I (Table 3).

4.2. Definition of the snowballing start set

Selection of studies was performed based on abstracts, titles and keywords and 473 papers in the Initial dataset were excluded after the Database filter step (Step 3 in Fig. 4). The collaborative inclusion-exclusion decision process contributed to reducing the possibility of researcher bias. The Filtered dataset (A), composed of 22 primary studies, was used as a start set and allowed us to detect other relevant papers by snowballing.

4.3. Snowballing results and data extraction

Snowballing added other 29 primary studies to the Filtered dataset. Appendix A lists all of the 51 selected studies that composed the Post-snowballing dataset (B). Database searches retrieved the studies S.1 to S.22 (Steps 1 to 3 in Fig. 4) and snowballing retrieved the studies S.23 to S.51 (Step 4 in Fig. 4). We applied the exclusion criteria in a full-text exam and excluded the following 18 false positive studies:

- Studies S.2, S.11, S.12, S.20, S.26, S.29, S.44, S.45, S.47, and S.50 are measurements or project alignment approaches without a focus on business alignment.
- Studies S.4, S.16, S.22, S.24, S.28, S.33, S.41, and S.43 do not focus on SPI.

Six studies were not accessible free of charge [S.3; S.23; S.25; S.32; S.39; S.46]. We bought the GQM+Strategies book [S.39] and received copies of S.3 and S.32 papers from the authors. Therefore three studies were not fully analyzed due to unavailability [S.23; S.25; S.46] and only FSB was applied to these papers.

The 29 additionally found studies confirmed the snowballing suitability for identification of studies not retrieved by database searches. The 30 studies, representing the Final dataset C, were classified as theoretical or empirical to address the research questions. (Table 4)

4.3.1. Dynamic realignment methods

None of the retrieved approaches present a realignment procedure to be used reacting to changes in the business goals. The GQM+Strategies book has a vague mention about this issue. Although the book describes a grid review options during a plan execution, the authors do not detail procedures for this realignment. The selected empirical study described in the GQM+Strategies book does not deal with a realignment issue.

4.3.2. Decision-making methods

The process for prioritizing business goals (strategic decision-making) is not supported by any systematic method in any of the retrieved approaches. Almost all approaches describe procedures starting with business goals that have already been defined (e.g. goals stated by existing strategic plans). Nevertheless, they do not deal with the prioritization issue. The process for selecting SPI alternatives (tactical decision-making) is partially supported by a systematic method in Karlström approach (S.27). This study uses the Analytic Hierarchy Process (AHP) proposed by Saaty (1980) for rating certain factors affecting a process improvement goal. However, SPI decisions are not taken as a result of this method because the ratings are applied only for rank factors that influence goals and not for selecting SPI alternatives.

4.4. Approach classification

Two authors performed the study classification according to the previously described research method using the Wieringa et al. (2006) taxonomy. The rating showed an agreement of 93% and a 0.857 kappa value, indicating an “Almost Perfect” strength of agreement (Landis and Koch, 1977). This measurement was only used as part of a quality assurance activity (QA4 in Fig. 4) as the authors consensually classify 12 studies (S.1, S.9, S.13, S.15, S.30, S.34, S.35, S.36, S.37, S.38, S.42 and S.51) as theoretical, excluding

Table 5
The final dataset.

Classification	Study	Approach name	Pub. year
Theoretical	S.1	Waina	2001
	S.9	GQM+Strategies	2010
	S.38		2007
	S.42		2007
	S.51		2007
	S.13	Plösch	2011
	S.15	Reiblein	1997
	S.30	Sun and Liu	2010
	S.36		2005
	S.34	Rosetta Stone Methodology (RSM)	2010
	S.35		2010
	S.37	McCoy	1998
Empirical	S.3	AE-MPS	2008
	S.19		2008
	S.5	GQM+Strategies	2011
	S.39		2014
	S.40		2013
	S.49		2010
	S.6	Barreto	2010
	S.7	Entropy	2009
	S.8	BOQM	2010
	S.10	Sommerville	1999
	S.14	Albuquerque	2009
	S.17	TNNL	2005
	S.48		2004
	S.18	SBPA	2011
	S.21	IDEAL	2000
	S.27	Karlström	2002
	S.31	SCOPE	2010
	S.32	ami	2000

Table 6
Rigor-relevance assessment results. Columns in the table represent values for rigor-relevance assessment. The labels for Rigor attributes are C – Context, D – Design, and V – Validity threats. The labels for Relevance attributes are U – User/Subject, C – Context, S – Scale, and RM – Research Method. Appendix B provides the description of the attributes.

Study	C	D	V	Rigor	U	C	S	RM	Relevance
S.3	1	0.5	0.5	**2**	1	1	0	1	**3**
S.5	0.5	0	0.5	**1**	1	1	0	0	**2**
S.6	1	0.5	0	**1.5**	0	0	1	1	**2**
S.7	1	0.5	0	**1.5**	1	1	1	1	**4**
S.8	1	0.5	0	**1.5**	1	1	1	1	**4**
S.10	0.5	0.5	0.5	**1.5**	0	1	0	0	**1**
S.14	1	0.5	0	**1.5**	0	0	1	0	**1**
S.17	1	0	0	**1**	1	1	0	0	**2**
S.18	0.5	0	0	**0.5**	1	1	0	1	**3**
S.19	1	0	0	**1**	0	1	0	0	**1**
S.21	1	0	0	**1**	1	1	1	0	**3**
S.27	1	1	1	**3**	1	1	1	1	**4**
S.31	0.5	1	0	**1.5**	0	0	0	0	**0**
S.32	0.5	0	0	**0.5**	1	1	1	0	**3**
S.39	1	0.5	0	**1.5**	1	1	0	1	**3**
S.40	1	0.5	0	**1.5**	1	1	1	0	**3**
S.48	1	0	0	**1**	1	1	1	0	**3**
S.49	1	0.5	0.5	**2**	1	1	0	1	**3**

them from rigor-relevance assessment. The Table 5 presents the Final dataset.

4.5. Rigor-relevance assessment results

The first and second authors performed an independent rigor-relevance assessment. As previously described, the initially obtained kappa entailed a refinement of the assessment criteria and a subsequent evaluation. Table 6 presents the first author's grade revision.

The rigor rubric average was 1.4 whereas relevance rubric average was 2.5. These grades were compared with a second au-

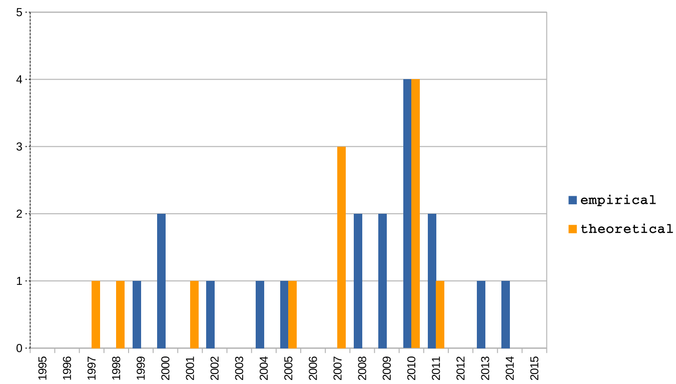


Fig. 7. Study distribution by year.

thor's rating performing the quality assurance activity 6 (QA6 in Fig. 4). The rigor rubric assessment showed an agreement between the researchers in 77.8% of the grades, while the obtained kappa value was 0.687 indicating a "Substantial" strength of agreement. The relevance rubric assessment had an agreement between the researchers in 88.9% of the grades, while the obtained kappa value was 0.851 indicating an "Almost Perfect" strength of agreement as defined by Landis and Koch (1977).

4.6. Answers to research questions

The Final dataset (Table 5) answers the research question 1, which deals with the availability of approaches. We classified the 19 approaches we found as theoretical or empirical. The GQM+Strategies' book partially address a realignment solution (RQ 1.1) although it does not provide a procedure and an evidence of use of the realignment. Concerning the availability of methods for decision-making support (RQ 1.2), only one approach presents a systematic method supporting part of the tactical decision-making. However, the AHP usage in Karlström approach does not answer the research question RQ 1.2 positively because it does not address the selection among SPI alternatives. Table 5 also answers the research question about the empirical evidence of the retrieved approaches (RQ 2). The table shows that 13 approaches present 18 empirical studies, while the other 6 approaches only present theoretical studies. The results of the empirical studies' assessment (presented in Table 6) answer the research questions about the quality of the evidence (RQ 2.1, and RQ 2.2). In Section 5, we discuss these results in more detail and conjecture about possible causes for the findings.

4.6.1. Approaches and citation graph

Fig. 7 illustrates the study distribution by year, also given in Table 5.

The concentration of studies between 2007 and 2011 clearly stands out. Three seminal theoretical studies of GQM+Strategies were published in 2007. The highest concentration of studies was in 2010, with the largest number of new proposed approaches (RSM, BOQM, Barreto, and SCOPE). The eight studies describe six approaches. Both theoretical RSM studies (S.34 and S.35) are very similar. BOQM, Barreto and SCOPE approaches, along with a GQM+Strategies study (S.49) are empirical studies. Another GQM+Strategies study (S.49), as well as a Sun and Liu study (S.30), are theoretical complements of previously published approaches. The two most recent approaches (Plösch and SBPA) were published in 2011 along with another GQM+Strategies empirical study (S.5). Only two empirical studies of GQM+Strategies (S.39 and S.40) were published after 2011 indicating a discontinuity of research on this topic. The empirical study described in the GQM+Strategies' book

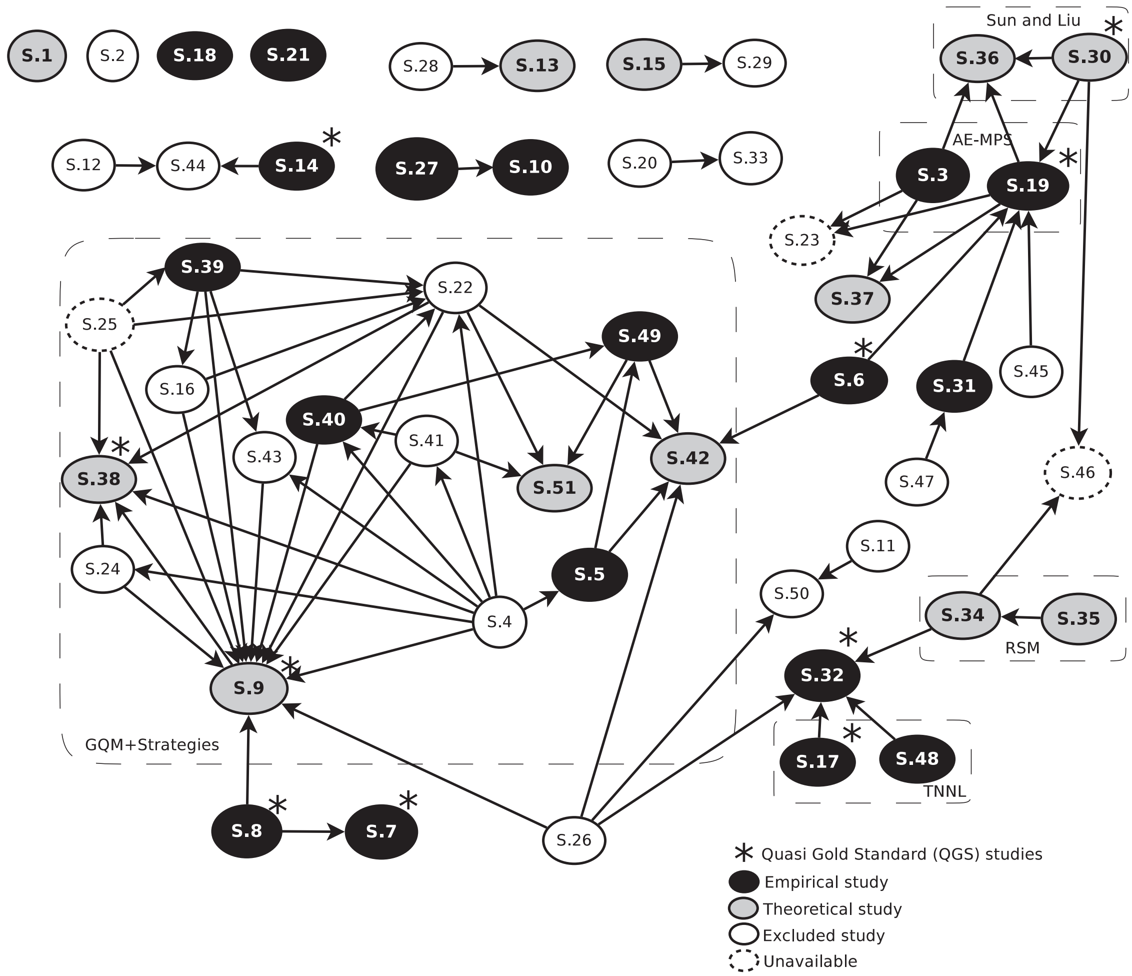


Fig. 8. Citation graph. Empirical studies are represented with a black background, theoretical studies with a gray background, excluded studies with a white background, and unavailable studies with a dotted line and a white background. A dashed box marks a group of studies related to the same approach.

(published in 2014) is quite similar to the case study presented in the S.49 study (published in 2010). AE-MPS (S.3 and S.19) and TNNL studies (S.17 and S.48) also describe a similar evidence.

The Final dataset brings 19 approaches in which only 13 have empirical evidence because Reiblein (S.15), McCoy (S.37), Waina (S.1), Sun and Liu (S.30 and S.36), and RSM (S.34 and S.35) are only theoretical proposals. Fig. 8 shows the citation graph representing empirical studies with a black background, theoretical studies with a gray background, excluded studies with a white background, and unavailable studies with a dotted line and a white background. A dashed box marks a group of studies related to the same approach. Asterisks mark each Initial QGS dataset study.

The GQM$^+$Strategies showed a significant number of citations; nevertheless, only two cited by other approaches of interest (S.9 and S.42). The S.9 theoretical study was cited by 11 other studies although only by one another approach (BOQM - S.8). Despite the study being published three years after the first appearance of the approach, this high citation level could be explained due to the high degree of impact of the vehicle used (IEEE Computer).

Snowballing fully recovered the studies of the Initial QGS dataset, which accounted for nine different approaches. The degree of influence, represented by the number of times another has cited an approach, is illustrated by direct edges in Fig. 8.

Table 7 shows the 19 retrieved approaches, chronologically ordered by the oldest included study, and indicates their degree of influence. Earlier approaches had been expected to have an influence on more recent ones; however, the graph showed a very small number of citations also indicating little comparison among approaches.

AE-MPS was the most influential approach with study S.19 cited by Sun and Liu (S.30), Barreto (S.6) and SCOPE (S.31), but this result is nevertheless misleading due to the weak connectivity illustrated by the citation graph. Although the previously presented graph shows strongly connected subgraph, it represents an intra-approach connection (see the group GQM$^+$Strategies in Fig. 8).

4.6.2. Rigor-relevance assessment results - study distribution

The rigor and relevance ratings are used to answer research question RQ2. As presented by Munir et al. (2015), studies are distributed by the rigor and relevance scores into four quadrants (A, B, C and D).

The following rationale did the classification:

- Studies with scores from 0 to 1.5 are classified as low rigor since high rigor is defined for a score of 2 or higher.
- Studies with scores from 0 to 2 are classified as low relevance since high relevance covers score of 3 or 4.

The quadrants visualized in Fig. 9 represented the four possible combinations:

A. High rigor and high relevance;
B. High rigor and low relevance;
C. Low rigor and high relevance; and
D. Low rigor and low relevance.

Table 7
Approach publication year and influence.

Studies	Approach name	1st pub. year	Influence
S.37	McCoy	1998	2
S.10	Sommerville	1999	1
S.21	IDEAL	2000	0
S.32	ami	2000	2
S.1	Waina	2001	0
S.27	Karlström	2002	0
S.17 and s.48	TNNL	2004	0
S.30 and S.36	Sun and Liu	2005	1
S.5, S.9, S.38, S.39, S.40, S.42, S.49 and S.51	GQM^{+}Strategies	2007	2
S.3 and S.19	AE-MPS	2008	3
S.14	Albuquerque	2009	0
S.15	Reiblein	2009	0
S.7	Entropy	2009	1
S.6	Barreto	2010	0
S.8	BOQM	2010	0
S.31	SCOPE	2010	0
S.34 and S.35	Rosetta Stone Methodology	2010	0
S.13	Plösch	2011	0
S.18	SBPA	2011	0

Table 8
Evolution of sensitivity and precision values, relating to the Initial QGS dataset.

Data set	Studies	Initial QGS	Retrieved	Sensitivity (%)	Precision (%)
Initial dataset (I)	495	10	7	70	1.4
Filtered dataset (A)	22	10	7	70	31.8
Post-snowballing dataset (B)	51	10	10	100	19.6
Final dataset (C)	30	10	10	100	33.3

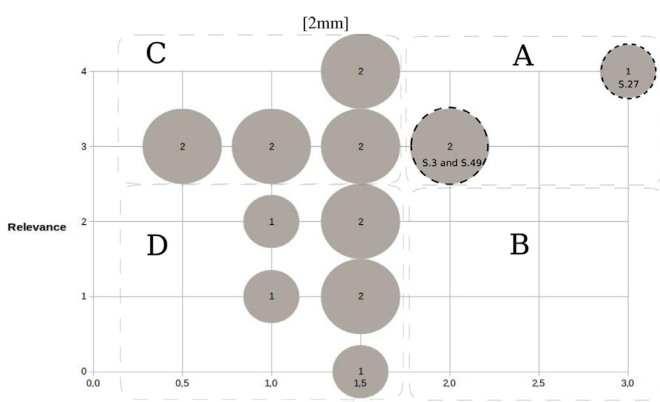


Fig. 9. Rigor-relevance assessment results - study distribution.

Only three studies were found to have the highest rigor and relevance grades (see area A in Fig. 9). No study was classified in the B category (high rigor and low relevance). Moreover, eight studies were ranked in the C category (high relevance but low rigor) and seven studies in the D category (low relevance and low rigor). S.27 (Karlström approach) was the best-rated study. It was retrieved by snowballing and did not present any influence on other approaches despite its age (13 years).

4.7. Protocol evaluation

The Final dataset was used to evaluate the protocol as established in our protocol (see Step 6 in Fig. 4). Sensitivity and precision values were calculated based on the Initial QGS and Final datasets. The Final dataset presents 20 new studies to the Initial QGS dataset, eight retrieved by the database search and 12 by snowballing. Table 8 represents the evolution of the sensitivity and precision values relating to the Initial QGS dataset and Table 9 illustrates the evolution of the sensitivity and precision values relating to the Final dataset.

The sensitivity of the database search decreased when considering the Final dataset. A better string refinement would have to be done to increase sensitivity if using only database search. As a result, the string should become more comprehensive and complex, considerably decreasing the precision. Low precision implies lower efficiency due to the nonlinear increase of false positives; therefore, it can be concluded that snowballing contributed to the protocol efficiency.

Conversely, the precision could not reach 100% only snowballing the Initial QGS dataset. Seven studies (S.1, S.10, S.13, S.15, S.18, S.21, and S.27) would not be retrieved if we used the Initial QGS dataset as a snowballing start set. The use of the database search to configure the snowballing start set made possible the recovery of these studies. The citation graph (see Fig. 8) illustrates this by the subgraphs analysis. Subgraphs without an Initial QGS study (indicated by an asterisk) as a vertex, would not be retrieved. We observed that the same 100% precision could be obtained by applying the database search step only on ACM, IEEE and SCOPUS, reducing the Initial dataset to 158 studies, therefore increasing the protocol efficiency. Nevertheless, it would be unknown which database to disregard *a priori* because the Duplicate elimination step did not show any unnecessary databases as each contributed to at least one unique study. The decrease of the precision in the evolution from the dataset A (Filtered) to dataset B (Post-snowballing) always occurs due to false positives insertions. But the analyses based on the Final dataset (see Table 9) shows the contribution of the snowballing to the efficiency of the search. The snowballing decreased the precision value only 9.4 percentage points (68.2–58.8%) and increased the sensitivity 50 percentage points (50–100%).

5. Discussions and threats to validity

The results of our SLR confirmed our conjecture that there is insufficient evidence of the validation of the proposed approaches

Table 9
Evolution of sensitivity and precision values, relating to the final dataset.

Data set	Studies	Final dataset	Retrieved	Sensitivity (%)	Precision (%)
Initial dataset (I)	495	30	15	50	3.0
Filtered dataset (A)	22	30	15	50	68.2
Post-snowballing dataset (B)	51	30	30	100	58.8
Final dataset (C)	30	30	30	100	100

to the strategic alignment of SPI and brought up some open questions, contributing to a proposal of a future research agenda. In this section, we discuss our findings and the limitations of our SLR because of possible biases.

5.1. Findings

5.1.1. No new proposals since 2011

Our investigation to answer the research question 1 (RQ 1) resulted in 19 approaches to the strategic alignment of SPI. As we can see in Table 5, there was a significant number of proposals from 1997 until 2011, but, surprisingly, no new proposal since then. This last finding was unexpected. Our perception is that practitioners still have no support for the alignment of the SPI with the business goals by a proper technology transfer (Gorschek et al., 2006), yet the scientific community has silenced about the problem. A conjecture is that the popularity of the SPI area itself has decreased, affecting the continuity of research in any topic related to this field. However, Kuhrmann et al. (2016) contradict this conjecture by a mapping study in which they show that the number of studies published annually has increased for the past 20 years. We also refute the conjecture and argue that the continuous evolution of technology implies in new practices of software development and, consequently, in new SPI initiatives.

Another possibility is that the alignment of SPI with business goals is not a research-worthy problem because the published approaches have already solved it. For a problem to be considered solved, it requires a validated approach. The validation of candidate solutions has three steps in the technology transfer model proposed by Gorschek et al. (2006) and later summarized in Gorschek (2015). The first step involves an *academic validation* to test the proposed solution and to get feedback for improvements. Gorschek et al. (2006) argue that a solution validation is not complete if it is carried out only in an academic setting. We agree with them and penalize the empirical studies carried out only in academic environments (e.g., in laboratory trials) by decreasing their score in the attribute Relevance-Research Method (see Appendix B). The two subsequent steps (static and dynamic validations) occur in the industry. The *static validation* involves training of the practitioners and support by the solution proponents throughout the evaluation of the solution. The *dynamic validation* consists in handing over the solution to the industry without any interference of the solution proponents. Only eight of 18 empirical studies assessed describe a static validation, and none of the evaluated studies presents a dynamic validation. Therefore, we can conclude that the alignment of SPI with business goals is not a problem solved by the proposed approaches.

Other probable causes for the lack of new proposals can be the challenges on university–industry collaborations mentioned by Lethbridge et al. (2008), the knowledge transfer problem presented by Shapiro et al. (2007), or the challenge of interdisciplinary research pointed by Starkey and Madan (2001). We explore these aspects hereafter.

Lethbridge et al. (2008) present a list of challenges of a university–industry collaboration in programs involving software engineering empirical studies. Dealing with conflicting interests of researchers and company managers, establishing adequate experimental controls, and choosing the appropriate metrics are examples of challenges in these initiatives. The conflict of interests in this type of research is usually related to the requirements of data confidentiality (i.e., Non-Disclosure Agreements), different definitions of success, different perceptions of the problem, and different perspectives of commitment to the project. Research may be jeopardized or its results not be published if these conflicts are not negotiated or previously determined by either party.

Shapiro et al. (2007) discuss the knowledge-transfer problem and separate this problem into two groups: “lost in translation” and “lost before translation”. The former relates to the synthesis quality of the studies, which is directly linked to our findings of low rigor values of the assessed studies. The later refers to the production difficulty associated to the challenges in the university-industry collaboration aforementioned. The scarcity of new proposals might be a consequence of the perceived risks related to these two groups of problems.

Starkey and Madan (2001) state that developing an interdisciplinary researcher requires more time than developping researchers in single disciplines. Indeed, Lethbridge et al. (2008) point that software engineering researchers are generally not trained in management. This lack of expertise might keep researchers away from the management issues related to SPI, preventing them from developing new approaches to strategic alignment of SPI.

To conclude, our results show that the interest in strategic alignment of SPI decreased without a consistent and practical approach to achieving such alignment.

5.1.2. Lack of methods for realignment and decision-making support

The lack of a realignment method in the evaluated approaches (see Section 4.3.1) indicates that the realignment remains an open issue, answering, therefore, the research question 1.1 (RQ 1.1). The use of the evaluated approaches, in situations that require a review of the SPI actions in progress, needs to be addressed by new empirical studies. Static implementations, i.e., considering a frozen business analysis, are not relevant because a complete solution should address the emergent strategies defined by Mintzberg and Waters (1985). In fact, a change of business objectives and the revision of the established improvement plans involve decisions about reallocation of resources and redefinition of priorities that must be supported by a realignment method. Regardless of the difficulty that realignment represents, it is a research challenge that must be met to offer practical support for the strategic alignment of SPI.

Another finding is the lack of a method for decision-making support in the existing approaches which answers the research question 1.2 (RQ 1.2). As mentioned in Section 4.3.2, only one of the 19 approaches presents a systematic support of decision making partially used in the strategic alignment of SPI. Even though the strategic alignment of SPI is a decision-making problem, no other study describes how to make decisions in this process. We argue that there could be two main reasons for this lacuna: the interdisciplinarity challenge and the decision-making paradox.

The first reason is again related to the lack of complementary expertise necessary in this type of research. Decision-making is a subject of other disciplines, such as operations research, psychology, economics, and cognitive science. March (1991) points that developing a theory of decision-making in organizations can be supported by different perspectives. Each of the above disciplines deals with a particular view of the problem, and the researcher needs to know which theory has to be applied in the context and to the problem under investigation.

The other reason for the lack of decision-making methods in the studies about the strategic alignment of SPI can be the challenge of deciding on the best decision-making method, also known as the *decision-making paradox*. Triantaphyllou and Mann (1989) discuss the comparative value of four multi-criteria decision-making (MCDM) methods to establish this paradox. The conclusion of their study is that the four methods yield different results even when they are fed with the same data and decision problem. Moreover, Montibeller and Franco (2010) argue that MCDM has limitations that may interfere with its use for supporting a strategic decision process, claiming that changes to MCDM methods require the consideration of other technical and social aspects.

An MCDM method works in the analysis of a finite set of alternatives that are evaluated regarding well-defined criteria, indicating a well-structured problem. Nevertheless, some decision problems are ill-structured and occur in dynamic environments in which time pressure and uncertainty are the norms. These situations demand another type of decision-making support and motivate the naturalistic decision-making (NDM) research. Moreover, Klein (2008) argues that decision support systems developed by formal methods do not improve decision quality. In contrast, the NDM researchers try to discover how people decide instead of using a formal model of decision making.

Regarding SPI, Moe et al. (2012) state that moving from a plan-driven to an agile development can be seen as moving from rational to naturalistic decision-making, showing the need for more empirical studies of NDM in agile development. Indeed, the agility expected in the tactical decisions (the selection of practices to improve) seems to deal with a naturalistic decision while a formal decision-making process still shows suitability to the strategic decisions (the prioritization of organizational goals). We argue that the institutionalization of agile practices is not a problem for NDM and an SPI program, even in an agile environment, still demands an MCDM method.

5.1.3. Empirical studies scarcity and low rigor and relevance values

Our SLR results reinforce the need to close the gap between research and practice, as six out of 19, or 31.6%, retrieved approaches are only theoretical proposals. The other 13 approaches present 18 empirical studies (See Table 5), answering the research question 2 (RQ 2). But, three approaches report the same evidence twice – AE-MPS (S.3 and S.19), GQM$^+$Strategies (S.39 and S.49), and TNNL (S.17 and S.48). Even if the reports are complementary in some way, the execution context of the approach is the same. This duplicity reduces our empirical sampling to only 15 contexts of use offered by 13 approaches, in which only the GQM$^+$Strategies approach presents more than one context of use. It is an insufficient sample to transfer any solution to practice properly.

In addition to the scarcity of the evidence, the results of the Rigor-Relevance assessment showed a low rigor average (1.4 in a 0–3 scale), which answers the research question 2.1 (RQ 2.1). The low rigor evaluation was mainly a consequence of the attributes Rigor-Design and the Rigor-Validity threat. Despite the description of the threats to validity be important to ensure the reliability of an empirical study and to contribute to the understanding of the challenges related to the use of an approach, only one out of the 18 empirical studies details the limitations of its research. Regarding the Design attribute, only two out of the 18 empirical studies completely describe their design and nine out of the 18 studies provide an incomplete description of the design. The other seven empirical studies do not describe their designs. The lack of proper measured variables and measurement criteria are the most common perceived flaws. It could be argued that it is difficult to determine an outcome of an approach to the strategic alignment of SPI. An SPI program demands time, and its long-term results could present many confounding variables that it becomes often impossible to measure the consequences of the decisions properly. However, the focus of a validation of such type of approach should be the decision-making process instead of the decisions' outcomes. Indeed, like any other strategic planning, it is often impossible to determine a cause-effect relationship between the decision-making process with the results of the implementation of the action plans, but an assessment of how much an approach could assist the decision-making process could be valuable.

Finally, the relevance average of 2.5 in a 0–4 scale (See Table 6) answers the research question 2.2 (RQ 2.2). This low average resulted from a lack of details of the descriptions related to the attributes Scale and Research Method. As we will explain later, this lack is a threat to the validity of our results because half of the evaluated studies do not clearly report the research method and the scale of the sample of the research. Our criteria penalize such lack of description because it hinders the understanding of the empirical studies, hence compromising their validation.

5.2. Threat to validity

Both internal and external dimensions of the validity (Kitchenham and Charters, 2007; Higgins and Green, 2011) were considered to minimize researchers' bias and to ensure the reliability of this SLR. The *internal dimension* of a study's validity represents how well the study answers its research questions and the bias degree of the study. The external validity, on the other hand, is related to the generalizability or applicability of a study's findings.

5.2.1. The quality of evidence and potential biases in the review process - internal validity

This study's internal validity was strengthened by a protocol designed with verifiable procedures that are based on well-known systematic review guidelines and recommendations (Wohlin, 2014; Kitchenham and Charters, 2007; Higgins and Green, 2011; Petersen et al., 2015; Kitchenham and Brereton, 2013) as well on guidelines for quality assessment criteria (Ivarsson and Gorschek, 2010).

To mitigate the selection bias, we designed a search strategy that defined the research terms and the inclusion-exclusion criteria. The search string was refined before the execution of the search, as shown in Fig. 5, to ensure that the selection process was as comprehensive and unbiased as possible. The search strategy was developed by the first author and reviewed by the others. Additionally, the results of the database search were checked by the use of the string in a parallel search carried out by an invited researcher. Finally, two invited researchers evaluated the inclusion-exclusion criteria applied to the initial dataset. They performed this quality assurance activity independently and reviewed 129 of the 495 studies. Their decisions about the inclusion or exclusion of a study were the same as those reached by the first author.

Another threat to validity refers to the existence of potential confounding factors, i.e., whether the results of a study are really based on the collected data (Easterbrook et al., 2008). The lack of a standard terminology on the subject was an important concern in this study. A string refinement was carried out supported by the PICOC criteria and the Initial QGS dataset. The Initial QGS dataset was also used for the search performance evaluation through the measurement of sensitivity and precision of the search. Moreover, the data properties and relationship with the research questions were established to address data extraction accuracy (see Table 2), enabling the agreement in the classification of the studies. The other authors later reviewed this data-extraction strategy proposed by the first author.

Regarding the search and the selection of studies, we combined the snowballing and database search strategies to balance the sensitivity and precision values of our search. Seven digital libraries were used to provide, in an unbiased manner, the start set of the snowballing search. The snowballing was also used to mitigate the risk of missing relevant studies. The snowballing contributed to the reliability of our SLR because we retrieved six approaches and 50% of the selected studies through this complementary strategy. Moreover, the Snowballing-search and Snowballing-filter Steps (Steps 4 and 5 in Fig. 4) were performed by the first author and the results were inspected by two other authors independently. These quality-assurance activities (QA 3 and 4 in Fig. 4) were carried out hiding the outcomes of each researcher to minimize their performance bias. All questions raised by the two authors who performed the QA activities were properly answered by the first author and no disagreements were observed in the selection and filtering of the studies in the snowballing steps.

Limitations of this SLR still exist because there could be studies (and approaches) not retrieved if they do not cite or have not been cited by a study of our post-snowballing dataset. This risk is related to the lack of a standard terminology on the subject of the research questions and to the natural trade-off between database search sensitivity and precision. Another concern about the reliability of this SLR exists because future search replications may yield different results due to the inconsistencies of search engines mentioned by Brereton et al. (2007), affecting the definition of the start set for the snowballing step. Moreover, the snowballing search also presents limitations, especially in the forward snowball (FSB), consisting of the selection of citations retrieved by Google Scholar, due to inconsistencies of the Google Scholar. We performed another Snowballing-search Step based on the Final dataset before the submission of this paper (in July 2016), and we obtained an incorrect register of a citation. The result of FSB in Google Scholar retrieved a citation of the S.51 by the S.24 study, but the list of references of S.24 does not include the S.51 study. A cross-checking between the results of the FSB and BSB detected this error.

To minimize the performance bias in the Rigor-Relevance Assessment Step, the checklist proposed by Ivarsson and Gorschek (2010) was extended by the first author, as described in Section 3.2.6. The adapted criteria were validated by the other authors. Additionally, the first and second authors rated the studies independently for quality assurance. The initially obtained kappa demanded a second rating that was performed after a refinement of the assessment criteria. The Final Kappa index indicated a "Substantial" strength of agreement of the assessment of rigor and an "Almost Perfect" strength of agreement of the assessment of relevance. Reaching such level of inter-observer agreement was possible after the refinement of the assessment criteria. However, some classification failures are conceivable because the authors had to make inferences about some studies" content due to the lack of information in the reports. Another threat to the validity of our SLR is the fact that the same researchers performed the initial rating and the second rating, hence contributing to a higher kappa value (Sim and Wright, 2005).

5.2.2. Overall completeness and applicability of evidence - external validity

External validity is concerned with the relevance of the evidence. The evidence obtained and evaluated by our SLR were relevant to answer the research questions and to confirm our conjecture that there is insufficient evidence of the validation of the existing approaches to the strategic alignment of SPI. We argue that it is unlikely that our SLR missed a relevant study because it is expected that a relevant study cites, or is cited by, another relevant study that was retrieved. Our results are applicable to both practitioners and researchers. The contribution of this study for a practitioner is an overview of existing approaches to strategic alignment of SPI. Furthermore, the identified empirical studies describe the contexts where each approach can be applied. The contributions of this study for researchers are summarized at the end of Section 1.

6. Conclusion and future work

The objectives of our systematic literature review were to identify the available approaches to strategic alignment of SPI and evaluate empirical studies to confirm or refute our conjecture that there is insufficient evidence of the validation of the proposed approaches to the strategic alignment of SPI. This SLR identified 19 approaches to the strategic alignment of SPI in 30 selected studies. We assessed 18 empirical studies using a rigor and relevance criteria and confirmed our conjecture because of the scarcity of empirical studies and the low rigor values of the assessed ones. We also found that there is no method to support the decision-making processes in the strategic and tactical levels of a strategic alignment of SPI. Another finding of our SLR was that none of the retrieved approaches addresses the realignment that is necessary to react to emergent strategies.

The main conclusion of our study is that the lack of empirical validation indicates that the results of the existing approaches have not been properly transferred to practitioners yet, calling for more rigorous studies on the subject. From the findings of this SLR, we list the following goals that we will pursue in the near future,

- Conduct a qualitative research to collect data on the perceived values of the formal and naturalistic approaches to the decision-making process of strategic alignment of SPI. The NDM research method and a constructivist approach may be appropriate in an agile context of SPI, while the formal methods may be more appropriate to plan-driven scenarios.
- Extend an in-progress NDM research that is investigating how software project managers decide (Cunha et al., 2016) to investigate how an agile team decides on the improvement of its software process.
- Conduct empirical studies to determine the usefulness and usability of a formal decision-making to support the strategic alignment of SPI, both in agile and in plan-driven scenarios.
- Compare the sensitivity and precision of our SLR search strategy with a strategy based strictly on database search and with a strategy based strictly on snowballing.
- Propose a new set of rigor-relevance assessment criteria consolidate from the criteria proposed in Ivarsson and Gorschek (2010), Dybå and Dingsøyr (2008) and Martins and Gorschek (2016)
- Conduct a qualitative study to determine factors that can change an SPI program because of the organizational emergent strategies and to identify the criteria used for the realignment decision of the SPI.

Acknowledgments

The authors would like to thank the Brazilian Funding Agencies: CAPES (grant no. 1345078) and CNPq.

Francisco J. S. Vasconcellos is a Ph.D. student at the Federal University of Mato Grosso do Sul College of Computing, where he receives a scholarship from CAPES.

We would like to thank Professor Débora M.B. Paiva and Ph.D. candidate Vanessa Araújo Borges for their contribution as independent reviewers of the protocol.

Appendix A. Analyzed studies

S.1	R. B. Waina, A business goal-based approach to achieving systems engineering capability maturity, in: Digital Avionics Systems, 20th Conference, IEEE, 2001, pp. 4B2–1 – 4B2–13. doi: http://doi.org/dzq6gx.
S.2	F. G. Wilkie, F. Mc Caffery, D. McFall, N. Lester, E. Wilkinson, A low-overhead method for software process appraisal, Software Process: Improvement and Practice 12 (4) (2007) 339–349. doi:http://doi.org/dv5dmp.
S.3	A. L. Becker, J. L. N. Audy, R. Prikladnicki, An Approach to Support the Strategic Alignment of Software Process Improvement Programs, in: ICEIS, Proceedings of the Tenth International Conference on Enterprise Information Systems, IEEE, 2008, pp. 66–73. doi:http://doi.org/c6rz8d.
S.4	K. Petersen, C. Gencel, N. Asghari, S. Betz, An elicitation instrument for operationalising GQM$^+$Strategies (GQM$^+$ S-EI), Empirical Software Engineering 20 (4) (2015) 968–1005. doi:http://doi.org/f248fv.
S.5	T. Kaneko, M. Katahira, Y. Miyamoto, M. Kowalczyk, Application of GQM$^+$Strategies in the Japanese Space Industry, in: Software Measurement, Joint Conference of the 21st Int'l Workshop on (IWSM-MENSURA), IEEE, 2011, pp. 221–226. doi:http://doi.org/fzrnhf.
S.6	A. O. S. Barreto, A. R. Rocha, Defining and Monitoring Strategically Aligned Software Improvement Goals, in: Product-Focused Software Process Improvement: 11th International Conference, PROFES 2010, Limerick, Ireland, June 21–23, 2010. Proceedings, Springer Berlin Heidelberg, 2010, pp. 380–394. doi:http://doi.org/bshtt6.
S.7	J. J. Trienekens, R. Kusters, D. Kriek, P. Siemons, Entropy based software processes improvement, Software Quality Journal 17 (3) (2009) 231–243. doi:http://doi.org/ckbxxh.
S.8	J. G. Guzmán, H. A. Mitre, A. Amescua, M. Velasco, Integration of strategic management, process improvement and quantitative measurement for managing the competitiveness of software engineering organizations, Software Quality Journal 18 (3) (2010) 341–359. doi:http://doi.org/bxh2p3.
S.9	V. R. Basili, M. Lindvall, M. Regardie, C. Seaman, J. Heidrich, J. Münch, D. Rombach, A. Trendowicz, Linking Software Development and Business Strategy Through Measurement, Computer 43 (4) (2010) 57–65. doi:http://doi.org/bqh5bc.
S.10	I. Sommerville, P. Sawyer, S. Viller, Managing process inconsistency using viewpoints, IEEE Transactions on Software Engineering 25 (6) (1999) 784–799. doi:http://doi.org/b787zf.
S.11	Q. Wang, M. Li, Measuring and improving software process in China, in: 2005 International Symposium on Empirical Software Engineering, IEEE, 2005, pp. 183–192. doi:http://doi.org/bcwgf6.
S.12	P. V. Martins, A. R. da Silva, ProPAMet: A Metric for Process and Project Alignment, in: Software Process Improvement: 15th European Conference, EuroSPI 2008, Dublin, Ireland. Proceedings, Springer Berlin Heidelberg, 2008, pp. 201–212. doi:http://doi.org/b5mkvc.
S.13	R. Plösch, G. Pomberger, F. Stallinger, Software Engineering Strategies: Aligning Software Process Improvement with Strategic Goals, in: Software Process Improvement and Capability Determination: 11th International Conference, SPICE 2011, Dublin, Ireland. Proceedings, Springer Berlin Heidelberg, 2011, pp. 221–226. doi:http://doi.org/cpxmvg.
S.14	A. B. Albuquerque, A. R. Rocha, A. C. Lima, Software Process Improvement: Supporting the Linking of the Software and the Business Strategies, in: Product-Focused Software Process Improvement: 10th International Conference, PROFES 2009, Oulu, Finland. Proceedings, Springer Berlin Heidelberg, 2009, pp. 347–361. doi:http://doi.org/bm6hhb.
S.15	S. Reiblein, A. Symons, SPI:'I can't get no satisfaction'- Directing process improvement to meet business needs, Software Quality Journal 6 (2) (1997) 89–98. doi:http://doi.org/dr2jxj.
S.16	V. Mandić, V. Basili, M. Oivo, L. Harjumaa, J. Markkula, Utilizing GQM$^+$Strategies for an Organization-Wide Earned Value Analysis, in: 36th EUROMICRO Conference on Software Engineering and Advanced Applications, 2010, pp. 255–258. doi:http://doi.org/c2rcvf.
S.17	J. J. Trienekens, R. J. Kusters, B. Rendering, K. Stokla, Business-oriented process improvement: practices and experiences at Thales Naval The Netherlands (TNNL), Information and Software Technology 47 (2) (2005) 67–79. doi:http://doi.org/ctnbw9.
S.18	H. C. Esfahani, E. Yu, M. C. Annosi, Strategically Balanced Process Adoption, in: Proceedings of the 2011 International Conference on Software and Systems Process, 2011, pp. 169–178. doi:http://doi.org/dmgd8v.
S.19	A. L. Becker, R. Prikladnicki, J. L. N. Audy, Strategic Alignment of Software Process Improvement Programs Using QFD, in: Proceedings of the 1st International Workshop on Business Impact of Process Improvements, 2008, pp. 9–14. doi:http://doi.org/c6rz8d.
S.20	O'Connor, Rory V. and Lepmets, Marion, Exploring the Use of the Cynefin Framework to Inform Software Development Approach Decisions, in: Proceedings of the 2015 International Conference on Software and System Process, 2015, pp. 97–101. doi:http://doi.org/bkw7.
S.21	K. Kautz, H. W. Hansen, K. Thaysen, Applying and Adjusting a Software Process Improvement Model in Practice: The Use of the IDEAL Model in a Small Software Enterprise, in: Proceedings of the 22Nd International Conference on Software Engineering, 2000, pp. 626–633. doi:http://doi.org/d3rrkn.
S.22	V. Mandić, V. Basili, L. Harjumaa, M. Oivo, J. Markkula, Utilizing GQM$^+$Strategies for Business Value Analysis: An Approach for Evaluating Business Goals, in: Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, 2010, pp. 20:1–20:10. doi:http://doi.org/fnjgbg.
S.23	M. Biró, C. Tully, Better Software Practice for Business Benefit: Principles and Experiences, IEEE Computer Society Press, 1999, Ch. The Software Process in the Context of Business Goals and Performance, pp. 15–27.
S.24	V. Mandić, M. Oivo, Sas: A tool for the GQM$^+$Strategies grid derivation process, in: Product-Focused Software Process Improvement: 11th International Conference, PROFES 2010, Limerick, Ireland. Proceedings, Springer Berlin Heidelberg, 2010, pp. 291–305. doi:http://doi.org/b2v6qf.
S.25	F. Cocozza, E. Brenes, G. L. Herrera, M. Jenkins, A. Martínez, Application of GQM$^+$Strategies in a Small Software Development Unit, in: Product-Focused Software Process Improvement: 15th International Conference, PROFES 2014, Helsinki, Finland. Proceedings, Springer International Publishing, 2014, pp. 108–118. doi:http://doi.org/bk9j.
S.26	M. Unterkalmsteiner, T. Gorschek, A. K. M. M. Islam, C. K. Cheng, R. B. Permadi, R. Feldt, A conceptual framework for SPI evaluation, Journal of Software: Evolution and Process 26 (2) (2014) 251–279. doi:http://doi.org/f24nrr.
S.27	D. Karlström, P. Runeson, C. Wohlin, Aggregating viewpoints for strategic software process improvement- a method and a case study, IEE Proceedings - Software 149 (5) (2002) 143–152. doi:http://doi.org/fs3s9v.
S.28	F. Stallinger, R. Plösch, R. Neumann, S. Horn, J. Vollmar, Development and evaluation of systems engineering strategies: An assessment-based approach, in: Software Quality. Increasing Value in Software and Systems Development: 5th International Conference, SWQD 2013, Vienna, Austria. Proceedings, Springer Berlin Heidelberg, 2013, pp. 215–229. doi:http://doi.org/bk9k.
S.29	D. Hinley, S. Reiblein, A goal-oriented approach for managing software process change, WIT Transactions on Information and Communication Technologies, 13, WITPress, 1995. URL http://www.witpress.com/elibrary/wit-transactions-on-information-and-communication-technologies/13/10565 [Accessed July 20, 2016].

(continued on next page)

(*continued*)

S.30	Y. Sun, X. F. Liu, Business-oriented software process improvement based on CMMI using QFD, Information and Software Technology 52 (1) (2010) 79–91. doi:http://doi.org/fsznbq.
S.31	O. Armbrust, Determining organization-specific process suitability, in: New Modeling Concepts for Today's Software Processes: International Conference on Software Process, ICSP 2010, Paderborn, Germany. Proceedings, Springer Berlin Heidelberg, 2010, pp. 26–38. doi:http://doi.org/b2nrzr.
S.32	C. Debou, A. Kuntzmann-Combelles, Linking software process improvement to business strategies: experiences from industry, Software Process: Improvement and Practice 5 (1) (2000) 55–64. doi:http://doi.org/csvvkc.
S.33	M. Lepmets, R. V. O'Connor, A. Cater-Steel, A. L. Mesquida, T. McBride, A Cynefin Based Approach to Process Model Tailoring and Goal Alignment, in: Quality of Information and Communications Technology (QUATIC), 9th International Conference on the, 2014, pp. 166–169. doi:http://doi.org/bk9p.
S.34	F. McLoughlin, I. Richardson, The Rosetta Stone Methodology – A Benefits Driven Approach to Software Process Improvement, in: Product-Focused Software Process Improvement: 11th International Conference, PROFES 2010, Limerick, Ireland. Proceedings, Springer Berlin Heidelberg, 2010, pp. 366–379. doi:http://doi.org/dchhqg.
S.35	F. McLoughlin, I. Richardson, The Rosetta Stone Methodology – A Benefits-Driven Approach to SPI, in: Systems, Software and Services Process Improvement: 17th European Conference, EuroSPI 2010, Grenoble, France. Proceedings, Springer Berlin Heidelberg, 2010, pp. 201–212. doi:http://doi.org/dv93kd.
S.36	X. F. Liu, Y. Sun, G. Kane, Y. Kyoya, K. Noguchi, QFD Application in Software Process Management and Improvement Based on CMM, in: Proceedings of the Third Workshop on Software Quality, ACM, 2005, pp. 1–6. doi:http://doi.org/fp3vkd.
S.37	W. L. McCoy, Interfacing three complementary technologies: strategic planning, process modeling, and system dynamics, in: Systems, Man, and Cybernetics. 1998 IEEE International Conference on, Vol. 3, 1998, pp. 2620– 2624. doi:http://doi.org/d9szp7.
S.38	V. Basili, J. Heidrich, M. Lindvall, J. Münch, M. Regardie, D. Rombach, C. Seaman, A. Trendowicz, Bridging the Gap between Business Strategy and Software Development, ICIS 2007 Proceedings. URL http://aisel.aisnet.org/icis2007/25 [Accessed July 20, 2016].
S.39	V. Basili, A. Trendowicz, M. Kowalczyk, J. Heidrich, C. Seaman, J. Münch, D. Rombach, Aligning Organizations Through Measurement: The GQM+Strategies Approach, Springer Publishing Company, 2014.
S.40	J. Münch, F. Fagerholm, P. Kettunen, M. Pagels, J. Partanen, Experiences and Insights from Applying GQM+Strategies in a Systems Product Development Organisation, in: 39th Euromicro Conference on Software Engineering and Advanced Applications, 2013, pp. 70–77. doi:http://doi.org/bk9r.
S.41	J. Münch, F. Fagerholm, P. Kettunen, M. Pagels, J. Partanen, The Effects of GQM+Strategies on Organizational Alignment, in: Proceedings of the DASMA Software Metric Congress (MetriKon 2013), 2013. URL http://arxiv.org/abs/1311.6221 [Accessed July 20, 2016].
S.42	V. Basili, J. Heidrich, M. Lindvall, J. Münch, M. Regardie, A. Trendowicz, GQM+Strategies – Aligning Business Strategies with Software Measurement, in: First International Symposium on Empirical Software Engineering and Measurement (ESEM 2007), 2007, pp. 488–490. doi:http://doi.org/djx3kj.
S.43	A. Trendowicz, J. Heidrich, K. Shintani, Aligning Software Projects with Business Objectives, in: Software Measurement, 2011 Joint Conference of the 21st International Workshop on and 6th International Conference on Software Process and Product Measurement (IWSM-MENSURA), 2011, pp. 142–150. doi:http://doi.org/fx6hnw.
S.44	P. V. Martins, A. R. Silva, ProPAM: SPI based on Process and Project Alignment, in: Managing Worldwide Operations Communications with Information Technology, Information Resources Management Association, USA, 2007. URL http://www.irma-international.org/viewtitle/33250/ [Accessed July 20, 2016].
S.45	O. Armbrust, M. Katahira, T. Kaneko, Y. Miyamoto, Y. Koishi, Which Processes Are Needed in Five Years? Strategic Process Portfolio Management at the Japan Aerospace Exploration Agency (JAXA), in: Proceedings of the International SPICE Days, 2010. URL http://www.ove-armbrust.de/pubs.html [Accessed July 20, 2016].
S.46	X. F. Liu, Y. Sun, G. Kane, Y. Kyoya, K. Noguchi, Business-oriented software process improvement based on CMM using QFD, Software Process: Improvement and Practice 11 (6) (2006) 573–589. doi:http://doi.org/fcxvdm.
S.47	T. Birkholzer, C. Dickmann, J. Vaupel, A Framework for Systematic Evaluation of Process Improvement Priorities, in: 37th EUROMICRO Conference on Software Engineering and Advanced Applications, 2011, pp. 294–301. doi:http://doi.org/cx6nm6.
S.48	J. J. M. Trienekens, R. J. Kusters, B. Rendering, K. Stokla, Business Objectives as Drivers for Process Improvement: Practices and Experiences at Thales Naval The Netherlands (TNNL), in: Business Process Management: Second International Conference BPM, Proceedings, Springer Berlin Heidelberg, 2004, pp. 33–48. doi:http://doi.org/dxrxbg.
S.49	M. Kowalczyk, J. Münch, M. Katahira, T. Kaneko, Y. Miyamoto, Y. Koishi, Aligning Software-related Strategies in Multi-Organizational Settings, in: Proceedings of the International Conference on Software Process and Product Measurement (IWSM/MetriKon/Mensura), 2010, pp. 261–274. URL http://arxiv.org/pdf/1401.1910v1.pdf [Accessed July 20, 2016].
S.50	M. Murugappan, G. Keeni, Blending CMM and Six Sigma to meet business goals, IEEE Software 20 (2003) 42–48. doi:http://doi.org/dx4mxq.
S.51	V. Basili, J. Heidrich, M. Lindvall, J. Münch, M. Regardie, D. Rombach, C. Seaman, A. Trendowicz, GQM+Strategies: A comprehensive methodology for aligning business strategies with software measurement, in: Proceedings of the DASMA Software Metric Congress (MetriKon 2007), 2007, pp. 253–266. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.720.886 [Accessed July 20, 2016].

Appendix B. Rigor-relevance assessment criteria

A.1 - Rigor = C + D + V

- **Context (C)**

I. Strong description: The context is described to the degree where it becomes comparable to other settings. This involves description of market and organization structure (Petersen and Wohlin, 2009). If all these factors are highlighted, then C is evaluated to 1.
II. Medium description: The context that the study is performed is mentioned or presented briefly but not described sufficiently so that a reader can understand and compare it to another context. C is evaluated to 0.5.
III. Weak description: If no description of context is provided in the study, then C is evaluated to 0.

- **Design (D)**

I. Strong description: The study design is described in enough detail for the reader to understand. To be specific, if the study presents measured variables, measurement criteria, treatments, the number of subjects, and sampling, then D is evaluated to 1.
II. Medium description: If the study lacks any of the factors related to design and data collection is missing (see above), then D evaluates to 0.5.
III. Weak description: If no design description is provided at all, D is evaluated to 0.

- **Validity threats (V)**

I. Strong description: If different types of threats to validity (i.e., internal, external, conclusion, and construct validity) are evaluated and reflected upon, then V is evaluated to 1.
II. Medium description: If the study only highlights a subset of the relevant threat categories or the validity is mentioned but not described in detail, then V is evaluated to 0.5.
III. Weak description: If the study does not have any validity discussion, then V is evaluated to 0.

A.2 - Relevance = U + C + S + RM

- **User/Subject (U)**

I. Contribution to relevance: If the subjects used in the evaluation are representative of the intended users of the approach, i.e., software organization managers and SPI professionals, then U is evaluated to 1.
II. No contribution to relevance: If the subjects used in the evaluation are not representative of the envisioned approach users, then U is evaluated to 0.

- **Context (C)**

I. Contribution to relevance: If the evaluation is performed in a setting representative of the intended usage, i.e., industrial setting, then C is evaluated to 1.
II. No contribution to relevance: If the evaluation is performed in a laboratory or other settings not representative of a real usage, then C is evaluated to 0.

- **Scale (S)**

I. Contribution to relevance: If the scale of the applications used in the evaluation is of realistic size, i.e., it covers the whole organizational unit, then S is evaluated to 1.
II. No contribution to relevance: If the evaluation is performed in a simulation or using applications of unrealistic size (e.g., a project), then S is evaluated to 0.

- **Research Method (RM)**

I. Contribution to relevance: If the research method used in the evaluation is one that facilitates the investigation of real situations (action research, case study, field study, interviews, descriptive, or exploratory surveys), then RM is evaluated to 1.
II. No contribution to relevance: If the research method used in the evaluation does not lend itself to investigate real situations (e.g., Laboratory trials), then RM is evaluated to 0.

References

Albuquerque, A.B., Rocha, A.R., Lima, A.C., 2009. Software process improvement: supporting the linking of the software and the business strategies. In: Product-Focused Software Process Improvement: 10th International Conference, PROFES 2009, Oulu, Finland, June 15–17, 2009. Proceedings, Springer Berlin Heidelberg, pp. 347–361. doi: http://doi.org/bm6hhb.

Ansoff, H.I., 1965. Corporate Strategy: Business Policy for Growth and Expansion. McGraw-Hill.

Badampudi, D., Wohlin, C., Petersen, K., 2015. Experiences from using snowballing and database searches in systematic literature studies. In: Proceedings of the 19th International Conference on Evaluation and Assessment in Software Engineering, ACM, p. 17.

Barreto, A.O.S., Rocha, A.R., 2010. Defining and monitoring strategically aligned software improvement goals. In: Product-Focused Software Process Improvement: 11th International Conference, PROFES 2010, Limerick, Ireland, June 21–23, 2010. Proceedings, Springer Berlin Heidelberg, pp. 380–394. doi: http://doi.org/bshtt6.

Basili, V.R., 1989. Software development: a paradigm for the future. In: Computer Software and Applications Conference, 1989. COMPSAC 89., Proceedings of the 13th Annual International, pp. 471–485. doi:10.1109/CMPSAC.1989.65127.

Bayona, S., Calvo-Manzano, J.A., Feliu, T.S., 2012. Critical success factors in software process improvement: a systematic review. In: Software Process Improvement and Capability Determination: 12th International Conference, SPICE 2012, Proceedings, Springer Berlin Heidelberg, pp. 1–12. doi: http://doi.org/bfgd.

Becker, A.L., Prikladnicki, R., Audy, J.L.N., 2008. Strategic alignment of software process improvement programs using QFD. In: Proceedings of the 1st International Workshop on Business Impact of Process Improvements, BiPi '08, ACM, pp. 9–14. doi: http://doi.org/c6rz8d.

Brereton, P., Kitchenham, B.A., Budgen, D., Turner, M., Khalil, M., 2007. Lessons from applying the systematic literature review process within the software engineering domain. J. Syst. Softw. 80 (4), 571–583. doi: http://doi.org/bt6brd.

Chrissis, M.B., Konrad, M., Shrum, S., 2011. CMMI for Development: Guidelines for Process Integration and Product Improvement, third ed. Addison-Wesley Professional.

Cunha, J.A.O.G., Silva, F.Q.B.d., Moura, H.P., Vasconcellos, F.J.S., 2016. Towards a substantive theory of decision-making in software project management: Preliminary findings from a qualitative study. In: Proceedings of the Empirical Software Engineering and Measurement (ESEM). To appear.

Debou, C., Kuntzmann-Combelles, A., 2000. Linking software process improvement to business strategies: experiences from industry. Softw. Process 5 (1), 55–64. doi: http://doi.org/csvvkc.

Dybå, T., 2000. An instrument for measuring the key factors of success in software process improvement. In: Empirical Software Engineering, vol. 5. Kluwer Academic Publishers, pp. 357–390. http://doi.org/cztmpq.

Dybå, T., 2003. Factors of software process improvement success in small and large organizations: an empirical study in the scandinavian context. In: Proceedings of the 9th European Software Engineering Conference Held Jointly with 11th ACM SIGSOFT International Symposium on Foundations of Software Engineering, ACM, pp. 148–157. doi: http://doi.org/b3b8cz.

Dybå, T., 2005. An empirical investigation of the key factors for success in software process improvement. In: IEEE Transactions on Software Engineering, Vol. 31. IEEE, pp. 410–424. http://doi.org/dwph3p.

Dybå, T., Dingsøyr, T., 2008. Strength of evidence in systematic reviews in software engineering. In: Proceedings of the Second ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, ESEM '08, ACM, pp. 178–187. doi: http://doi.org/cxqnk3.

Dybå, T., Kitchenham, B.A., Jorgensen, M., 2005. Evidence-based software engineering for practitioners. IEEE Softw. 22 (1), 58–65. doi: http://doi.org/bvfs42.

Easterbrook, S., Singer, J., Storey, M.-A., Damian, D., 2008. Selecting Empirical Methods for Software Engineering Research. Springer London, pp. 285–311. doi: http://doi.org/fksbhs.

Ebert, C., 1999. Technical controlling and software process improvement. J. Syst. Softw. 46 (1), 25–39. doi: http://doi.org/fgxtr2.

Eisenhardt, K.M., Zbaracki, M.J., 1992. Strategic decision making. Strategic Manage. J. 13, 17–37. doi: http://doi.org/d4hbj6.

Gorschek, T., 2015. How to increase the likelihood of successful transfer to industry: going beyond the empirical. In: Proceedings of the Third International Workshop on Conducting Empirical Studies in Industry, CESI '15. IEEE Press, pp. 10–11. doi: http://doi.org/bf8j.

Gorschek, T., Garre, P., Larsson, S., Wohlin, C., 2006. A model for technology transfer in practice. IEEE Softw. 23 (6), 88–95. doi: http://doi.org/bww4sw.

Guzmán, J.G., Mitre, H.A., Amescua, A., Velasco, M., 2010. Integration of strategic management, process improvement and quantitative measurement for managing the competitiveness of software engineering organizations. Softw. Qual. J. 18 (3), 341–359. doi: http://doi.org/bxh2p3.

Higgins, J., Green, S., 2011. Cochrane handbook for systematic reviews of interventions. Accessed: July 20, 2016. URL http://handbook.cochrane.org/.

Humphrey, W.S., 1989. Managing the Software Process. Addison-Wesley Longman Publishing Co., Inc.

Hussey, D.E., 1997. Glossary of techniques for strategic analysis. Strategic Change 6, 97–115. doi: http://doi.org/fmw7m3.

ISACA, 2012. COBIT 5: A Business Framework for the Governance and Management of Enterprise IT, ISACA.

ISO, 2004. ISO/IEC 15504-4:2004 - Information technology – Process assessment – Part 4: Guidance on use for process improvement and process capability determination.

ISO, 2013. ISO/IEC TR 33014:2013 - Information technology – Process assessment – Guide for process improvement.

ISO, 2015. ISO/IEC 38500:2015 - Information technology – Governance of IT for the organization.

Ivarsson, M., Gorschek, T., 2010. A method for evaluating rigor and industrial relevance of technology evaluations. Emp. Softw. Eng. 16 (3), 365–395. doi: http://doi.org/drn5nb.

Kaplan, R.S., Norton, D.P., 2008. The Execution Premium: Linking Strategy to Operations for Competitive Advantage. Harvard Business Press.

Kitchenham, B., Brereton, P., 2013. A systematic review of systematic review process research in software engineering. Inf. Softw. Technol. 55 (12), 2049–2075. doi:10.1016/j.infsof.2013.07.010.

Kitchenham, B., Charters, S., 2007. Guidelines for performing systematic literature reviews in software engineering. Available at http://cdn.elsevier.com/promis_misc/525444systematicreviewsguide.pdf (Accessed 20.07.16).

Kitchenham, B.A., Dybå, T., Jorgensen, M., 2004. Evidence-based software engineering. In: Proceedings of the 26th International Conference on Software Engineering, ICSE '04, pp. 273–281. doi: http://doi.org/crc8n5.

Kitchenham, B.A., Li, Z., Burn, A.J., 2011. Validating search processes in systematic literature reviews. In: EAST 2011 - Proceeding of the 1st International Workshop on Evidential Assessment of Software Technologies, In conjunction with ENASE 2011,Beijing, China, June 2011, pp. 3–9.

Klein, G., 2008. Naturalistic decision making. Hum. Factors 50 (3), 456–460. doi: http://doi.org/bn54qh.

Kuhrmann, M., Diebold, P., Münch, J., 2016. Software process improvement: a systematic mapping study on the state of the art. PeerJ Comput. Sci. 2, e62.

Kuhrmann, M., Konopka, C., Nellemann, P., Diebold, P., Münch, J., 2015. Software process improvement: where is the evidence?: initial findings from a systematic mapping study. In: Proceedings of the 2015 International Conference on Software and System Process, ICSSP 2015, ACM, pp. 107–116. doi: http://doi.org/bf5j.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33 (1), 159–174. doi: http://doi.org/dtzfj3.

Lepmets, M., McBride, T., 2012. Process improvement for the small and agile. In: Systems, Software and Services Process Improvement: 19th European Conference, EuroSPI 2012, Vienna, Austria. Proceedings, Springer Berlin Heidelberg, pp. 310–318. doi: http://doi.org/bmng.

Lepmets, M., McBride, T., Ras, E., 2012. Goal alignment in process improvement. J. Syst. Softw. 85, 1440–1452. Elsevier Inc. doi: http://doi.org/bfgj.

Lethbridge, T.C., Lyon, S., Perry, P., 2008. The Management of University–Industry Collaborations Involving Empirical Studies of Software Engineering. Springer London, p. 257—281. doi: http://doi.org/dnw5pc.

Luecke, R., 2005. Harvard Business Essentials: Strategy: Create and Implement the Best Strategy for Your Business. Harvard Business School.

Manning, C.D., Raghavan, P., Schütze, H., 2008. Introduction to Information Retrieval. Cambridge University Press.

March, J.G., 1991. How decisions happen in organizations. Hum. Comput. Interact. 6 (2), 95–117. doi: http://doi.org/cfx9vw.

Martínez-Lorente, A.R., Dewhurst, F., Dale, B.G., 1998. Total quality management: origins and evolution of the term. TQM Mag. 10 (5), 378–386. doi:10.1108/09544789810231261.

Martins, L.E.G., Gorschek, T., 2016. Requirements engineering for safety-critical systems: a systematic literature review. Inf. Softw. Technol. 75, 71–89. doi: http://doi.org/bhfs.

McFeeley, R., 1996. IDEAL: A User's Guide for Software Process Improvement. Technical Report cmu/sei-96-hb-001. Software Engineering Institute.

Mintzberg, H., 1994. Rise and Fall of Strategic Planning. Simon and Schuster.

Mintzberg, H., Waters, J.A., 1985. Of strategies, deliberate and emergent. Strategic Manage. J. 6 (3), 257–272. doi: http://doi.org/bs2pm3.

Moe, N.B., Aurum, A., Dybå, T., 2012. Challenges of shared decision-making: a multiple case study of agile software development. Inf. Softw. Technol. 54 (8), 853–865. doi: http://doi.org/bx7sxk.

Montibeller, G., Franco, A., 2010. Handbook of Multicriteria Analysis. Springer Berlin Heidelberg, p. 25—48. Chapter Multi-Criteria Decision Analysis for Strategic Decision Making. doi: http://doi.org/bfv232.

Münch, J., Armbrust, O., Kowalczyk, M., Soto, M., 2012. Software Process Definition and Management. Springer Berlin Heidelberg. doi: http://doi.org/bf8f.

Munir, H., Wnuk, K., Runeson, P., 2015. Open innovation in software engineering: a systematic mapping study. Emp. Softw. Eng. 1–40.

Petersen, K., Vakkalanka, S., Kuzniarz, L., 2015. Guidelines for conducting systematic mapping studies in software engineering: an update. Inf. Softw. Technol. 64, 1–18. doi:10.1016/j.infsof.2015.03.007.

Petersen, K., Wohlin, C., 2009. Context in industrial software engineering research. In: Proceedings of the 2009 3rd International Symposium on Empirical Software Engineering and Measurement, IEEE Computer Society, pp. 401–404.

Petticrew, M., Roberts, H., 2006. Systematic reviews in the social sciences: a practical guide. John Wiley & Sons.

Pettigrew, A.M., 1987. Context and action in the transformation of the firm. J. Manage. Stud. 24 (6). http://doi.org/fmds3g.

Pickton, D.W., Wright, S., 1998. What's swot in strategic analysis? Strategic Change, vol. 7. John Wiley & Sons, Ltd. doi: http://doi.org/fspztg.

Qumer, A., Henderson-Sellers, B., 2008. A framework to support the evaluation, adoption and improvement of agile methods in practice. J. Syst. Softw. 81 (11), 1899–1919. doi: http://doi.org/cw6cfp.

Ringstad, M.A., Dingsøyr, T., Moe, N.B., 2011. Agile process improvement: diagnosis and planning to improve teamwork. In: Systems, Software and Service Process Improvement: 18th European Conference, EuroSPI 2011, Roskilde, Denmark. Proceedings, Springer Berlin Heidelberg, pp. 167–178. doi: http://doi.org/czp8g8.

Saaty, T.L., 1980. The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation. McGraw-Hill.

Salo, O., Abrahamsson, P., 2005. Integrating agile software development and software process improvement: a longitudinal case study. In: International Symposium on Empirical Software Engineering, pp. 193–202. doi:10.1109/ISESE.2005.1541828.

Shapiro, D.L., Kirkman, B.L., Courtney, H.G., 2007. Perceived causes and solutions of the translation problem in management research. Acad. Manage. J. 50 (2), 249–266.

Sim, J., Wright, C.C., 2005. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys. Ther. 85 (3), 257–268.

Starkey, K., Madan, P., 2001. Bridging the relevance gap: aligning stakeholders in the future of management research. Br. J. Manage. 12 (s1), S3–S26.

Triantaphyllou, E., Mann, S.H., 1989. An examination of the effectiveness of multi-dimensional decision-making methods: a decision-making paradox. Decis. Support Syst. 5 (3), 303–312. doi: http://doi.org/bn9m29.

Trienekens, J.J., Kusters, R.J., Rendering, B., Stokla, K., 2005. Business-oriented process improvement: practices and experiences at thales naval the netherlands (TNNL). Inf. Softw. Technol. 47 (2), 67–79. doi: http://doi.org/ctnbw9.

Ullah, A., Lai, R., 2013. A systematic review of business and information technology alignment. ACM Trans. Manage. Inf. Syst. 4 (1), 4:1–4:30. doi: http://doi.org/bf8h.

Wieringa, R., Maiden, N., Mead, N., Rolland, C., 2006. Requirements engineering paper classification and evaluation criteria: a proposal and a discussion. Requirements Eng. 11 (1), 102–107.

Wohlin, C., 2014. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering - EASE '14. ACM Press, pp. 1–10. doi: http://doi.org/f22cpf.

Zhang, H., Babar, M.A., Tell, P., 2011. Identifying relevant studies in software engineering. Inf. Softw. Technol. 53 (6), 625–637. doi: http://doi.org/bpmvnj.

Francisco J. S. Vasconcellos is a Ph.D. student at the Federal University of Mato Grosso do Sul, Brazil. He received his M.Sc. Degree in Computer Science from the State University of Campinas, Brazil. He has worked with software process improvement since 2001 as a consultant and appraiser of the Brazilian model for software process improvement (MPS.BR). He also has worked as a consultant on strategic planning since 2004. His research interests are the decision-making processes of software process improvement, agile methods, software process assessment, empirical software engineering, strategic management, project management and portfolio management.