Information and Software Technology 55 (2013) 1316–1340

Contents lists available at SciVerse ScienceDirect

Information and Software Technology

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

Team building criteria in software projects: A mix-method replicated study

Fabio Q.B. da Silva *, A. César C. França, Marcos Suassuna, Leila M.R. de Sousa Mariz, Isabella Rossiley, Regina C.G. de Miranda, Tatiana B. Gouveia, Cleviton V.F. Monteiro, Evisson Lucena, Elisa S.F. Cardozo, Edval Espindola

Centre for Informatics, UFPE, Cidade Universitária, 50.540-740, Recife, PE, Brazil

A R T I C L E   I N F O

Article history:
Available online 10 December 2012

Keywords:
People management
Software teams
Team effectiveness
Team building criteria
Software engineering

A B S T R A C T

Context: The internal composition of a work team is an important antecedent of team performance and the criteria used to select team members play an important role in determining team composition. However, there are only a handful of empirical studies about the use of team building criteria in the software industry.
Objective: The goal of this article is to identify criteria used in industrial practice to select members of a software project team, and to look for relationships between the use of these criteria and project success. In addition, we expect to contribute with findings about the use of replication in empirical studies involving human factors in software engineering.
Method: Our research was based on an iterative mix-method, replication strategy. In the first iteration, we used qualitative research to identify team-building criteria interviewing software project managers from industry. Then, we performed a cross-sectional survey to assess the correlations of the use of these criteria and project success. In the second iteration, we used the results of a systematic mapping study to complement the set of team building criteria. Finally, we performed a replication of the survey research with variations to verify and improve the results.
Results: Our results showed that the consistent use team building criteria correlated significantly with project success, and the criteria related to human factors, such as personality and behavior, presented the strongest correlations. The results of the replication did not reproduce the results of the original survey with respect to the correlations between criteria and success goals. Nevertheless, the variations in the design and the difference in the sample of projects allowed us to conclude that the two results were compatible, increasing our confidence on the existence of the correlations.
Conclusion: Our findings indicated that carefully selecting team member for software teams is likely to positively influence the projects in which these teams participate. Besides, it seems that the type of development method used can moderate (increase or decrease) this influence. In addition, our study showed that the choice of sampling technique is not straightforward given the many interacting factors affecting this type of investigation.

1. Introduction

The last decades have witnessed a growing emphasis on teamwork in organizations [45] and, consequently, a growing interest in understanding team effectiveness [44]. This has stimulated researchers to study teamwork in a variety of sectors [22,44,58] and to propose several theories and models to explain and predict team effectiveness [78]. In most of these models, team composition has been postulated as a multidimensional factor that affects team performance, mainly through its effect on several team processes [21,47,78]. However, there is neither a consensus about which dimensions should be used to define composition nor how specific compositions affect team process and, ultimately, the different dimensions of team effectiveness [78]. Nevertheless, researchers agree that team composition is built from the combination of individual team member characteristics and the role these individuals play in the team [62]. Therefore, the process of building a work team starts with the identification of individuals with the characteristics needed to create the desired composition in a given context [72].

The goal of this study is twofold. First, to identify the criteria used by project managers and team leaders to select software engineers when building a software project teams in industrial practice. Second, to investigate the relationships of the level of formalization in the use of selection criteria and several dimensions of team effectiveness. In this study, we focused on "the individual in the team", in the sense discussed by Stevens and Campion: "at the individual

* Corresponding author. Tel.: +55 85944085; fax: +55 81 21268438.
E-mail address: fabio@cin.ufpe.br (F.Q.B. da Silva).

0950-5849/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.infsof.2012.11.006

team member level of analysis, as opposed to the group or organizational levels" [75]. That is, our focus was on how to select individuals to be part of teams and not, for instance, on how to configure some optimal combination of individuals within a team. In other words, we are neither addressing the problem of how to apply the selection criteria to create combinations of individuals that would characterize specific team compositions nor investigating the relationship of team composition with team process and effectiveness, as performed for instance by Acuña et al. [4]. Besides, the investigation of role allocation, role and personal diversity and balance [12,15], the matching of task characteristics and team composition, and related issues are also not in the scope of this study. We chose this approach because to effectively study team composition we need first to understand the individual characteristics that are considered in practice. In fact, the team building criteria identified in our investigation provide a starting point for future investigations in how specific uses and combinations of criteria can lead to specific compositions of teams that would enhance effectiveness in different context. We believe that further work is needed to identify adequate matching between team composition and types of tasks in software development, towards the construction of models of team building in software engineering. Our study is one step in this direction.

To achieve our goals, we employed a mix-method [35], replication approach [16], based on two iterations, each one composed of two phases, performed over a five-year period, between 2007 and 2012, in the context of Brazilian software companies. In the first iteration, we conducted a qualitative study [63,73] to understand what team building criteria have been used by project managers and team leaders in practice. This first phase was followed by a cross-sectional survey [54] in which we investigated the correlation between the consistent use of these criteria and a multidimensional definition of project success goals. The results of this first iteration showed that the consistent use of the team building criteria correlated significantly with all dimensions used to define project success. Furthermore, criteria based on human factors, such as personality and behavior, presented the strongest correlations, while criteria related organizational factors represented the weakest correlations.

In the second iteration, we addressed some limitations found in the previous studies and extended our investigation. We performed a systematic mapping study [55,67] to verify and complement the set of team building criteria. The results of the mapping study added two new criteria and confirmed the use of four criteria from the original set. In the fourth phase, we replicated the cross-sectional survey with improved instruments and on a different organizational context. The results of the replication did not reproduce the results of the original survey with respect to the correlations between criteria and success goals. Nevertheless, the variations in the design between the two studies and the difference in the sample of projects allowed us to conclude that the two results were compatible, increasing our confidence on the existence of the correlations. Besides, we found that the type of development method (Agile or Traditional) may have a moderator effect on the correlation between the consistent use of team building criteria and the success goals of the projects.

da Silva et al. [27] presented some of the findings of the first iteration (Phase I and Phase II) and their limitations. In this article, we present an analysis and synthesis of the results of the four phases of our investigation, together with discussions about the implications for research and practice. This article extends the findings presented by da Silva et al. [27] in three important ways. First, two new team building criteria were found as the result of the mapping study developed in Phase III and used in the survey replication. Second, the replication of the survey in the forth phase motivated the discussion about the moderator effect of the development method on the relationship between the team building criteria and project outcomes. Finally, we also discussed the role and impact of variations, in particular the sampling strategy, in the development of replications of empirical studies.

Software teams have been a focus in software engineering since the early propositions of the "Chief Programmer Teams" by Baker [9] and the "Surgical Teams" of Brooks [17]. Over the last three decades, the study of human aspects in software engineering has addressed individual issues like motivation [11,36] and personality [24], as well as team aspects such as cohesion [31] and team structure and typology [30]. However, selection of individuals to build teams has only been directly addressed by a handful of studies and even fewer of them investigated selection from the perspective of the software engineer practitioner in the industrial context. Our study aims to complement the body of knowledge about selection of members to build software teams with an investigation of the industrial practice. This approach allowed us to identify organizational related criteria, such as project and client priorities, individual costs, and availability that have not been addressed in previous studies. On the other hand, it is likely that certain practices and predictors of personnel selection discussed in the scientific literature were not identified by our approach [72]. We tried to overcome this limitation by employing several forms of data triangulation, including method triangulation with a mapping study, and consider that the set of team building criteria identified provided a consistent starting point for future investigations.

The structure of this article is consistent with the iterative approach discussed above. We also followed the guidelines proposed by Carver [20] for reporting replication studies. In Section 2, we briefly discuss some conceptual background about teamwork along with related research about software teams. In Section 3, we summarize our mix-method, iterative research strategy. In Sections 4 and 5, we describe the studies performed in the two iterations of this research. In each section, we present goals, methodology, results, and limitations of each study performed in the corresponding iteration. In Section 6, we present a discussion about the differences of the results of the two iterations together with the limitations of our investigation, and the implications of our results for research and practice. Finally, in Section 7, we summarize our findings, describe potential areas for future work, and present some conclusions.

2. Conceptual background

In this section, we briefly present research about work teams and teamwork in general (Section 2.1) and a summary of the research about software teams related to the goals of this study (Section 2.2). In Section 5.2, we present other related research as part of the results of our mapping study.

2.1. Work teams and teamwork

The growth in knowledge, abilities, and skills needed to solve the tasks that create competitive advantages in today's organizations makes it impossible for individuals to work independently of teams [56]. The prevalence of teamwork in modern organizations has stimulated researchers to study teamwork in a variety of sectors. In the rest of this section, we characterize what constitutes a team for the purposes of our investigation, discuss what we mean by team effectiveness in this study, and provide a brief overview of concepts associated with personnel selection.

2.1.1. Characterizing what constitutes a team

It is likely that there are as many definitions of what constitutes and characterizes a team or work group in the literature as there are authors writing about them. In fact, some authors use team and group to refer to the same concept [22], while others tend to make distinctions between them [10,51]. Although some of the discussion about what exactly constitute a team or a work group, their similarities and differences, are important both from theoretical and practical perspectives, a deeper discussion on this issue is out of the scope of this paper. The interested reader will find plenty of references in three important literature reviews performed about research on work groups [22,44,58].

Nevertheless, we must choose a definition of team to focus our investigation and to make it clear what types of groups are and are not addressed in our studies. In this case, we are interested in teams working in organizations, therefore those teams that, according to Hackman are "(1) real groups (that is, intact social systems complete with boundaries and differentiated roles among members); (2) groups that have one or more tasks to perform, resulting in discernible and potentially measurable group products; and (3) groups that operate within an organizational context" [47].

This is a very inclusive characterization that should comprise most of the types of software teams in industry, including distributed teams. Groups working on open-source software would also be included given that it is often possible to have a good characterization of an (virtual) organizational context. On the other hand, co-acting groups (e.g. individuals working under the same management but with their own individual tasks to perform without interdependencies with other individuals) are excluded because there is no common group task.

2.1.2. Understanding team effectiveness

In this study, we are interested in relating the use of team building criteria to the outcomes of a project as the result of the work of the project team. Therefore, it is necessary to understand what constitutes effectiveness of a work team and derive a conceptual definition of team effectiveness to be operationalized in the research.

According to Campion et al. [18], teamwork reconciles two historically contradictory approaches on the theories about the structure of work: the psychological approach that aims to increase individual's satisfaction with the job and the engineering approach that emphasizes efficiency, effectiveness and productivity. For this reason, most researchers have emphasized that team effectiveness should be understood as a multidimensional construct [21,46,47,53]. In software development, this multidimensional characteristic of team effectiveness has been studies, in particular in the context of Agile teams [65]. Several theories and models have been proposed to understand, explain, and predict team effectiveness. Yeatts and Hyten [78] provided an overview of 10 models of team effectiveness, including their own proposition that was constructed based on empirical evidence found in several industrial case studies. However, there is little agreement about the set of criteria to measure effectiveness [37]. Hackman [47] uses team productivity, team satisfaction, and team continuity as dimensions of effectiveness. Cohen [21] proposed the use of three dimensions related to performance (costs, productivity, and quality), attitudes of the team members (satisfaction and commitment), and withdrawal behaviors (absenteeism and voluntary turnover). Yeatts and Hyten [76] used two dimensions related to customer satisfaction with productivity, quality, timeliness, and costs, and the economic viability of the team and the project.

In our research, we combined the propositions of Hackman [46–48], Hallows [50], Cohen [21], and Yeatts and Hyten [76] to achieve a definition of effectiveness that includes the dimensions of performance (cost, time, and scope) and satisfaction (team members, client or customer, and project manager). An operational definition of this multidimensional criterion of effectiveness will be used to assess the achievement of project success goals by the software teams in our empirical studies.

2.1.3. Conceptualizing personnel selection

Our investigation is primarily concerned about how individuals are selected by project managers or team leader to become part of software teams. The identification of the "right person for the job" is investigated in the literature about personnel selection. Personnel selection is a management discipline that aims to identify individuals to compose the workforce of an organization and is considered a central theme in the study of work behavior [43].

In general terms, the starting point of personnel selection is the identification of what individual characteristics are likely to be related to work performance. This entails the identification of what an individual does in the job and what constitutes or is valued as performance. Once we know what characteristics are needed and how performance will be assessed, we can work on the definition of methods to measure such characteristics and, then, consistently use these methods in personnel selection.

The literature on work behavior uses the term *predictor* to refer to unique combination of an individual characteristic (construct) and the method used for its measurement. Cognitive ability and aptitude tests, personality inventories, biodata, technical skills and job knowledge tests, interviews, etc. have been used as predictors in the research and practice of personnel selection for many years and in a wide range of industrial sectors (see Salgado et al. [72] for an overview).

In this study, we are interested in understanding which predictors have been used in practice to select individuals to build software teams. We consider selection broadly, including hiring new employees or choosing individuals that are already part of the workforce of the organization. We also want to investigate the levels of consistency and formalization of the use of these predictors and whether their consistent use is related to team effectiveness. We use the term "team building criterion" to refer to a predictor because we believe the name is easier to understand in the context of the software engineering practitioner.

2.2. Software teams

In software development, Weinberg [76] was one of the first authors to consider programming as an individual and social activity, i.e., developed by individuals through teamwork. The "Chief Programmer Teams" [9] and the "Surgical Teams" [17] were the first models that dealt explicitly with the structure and composition of software teams. Since then, software teams have been a focus of research by academics and practitioners [13,25]. Lettice and McCracken [57] have reported that the amount of researches related to software team doubled between 1997 and 2007. The increasing interest in management of software teams stems from its reported effects on productivity [13], product quality [59], and project failure [25].

2.2.1. Software team effectiveness

Several authors, including Shneiderman [74], De Marco and Lister [32], Constantine [23], Guinan et al. [42] have investigated and described personal and social factors that can affect the effectiveness of software teams, including: the interaction between the personality of team members, role-related diversity, the effects of the work environment, organization and structure of teams, team processes such as communication, conflict and cohesion, among others. However, we have not identified models or theories of team effectiveness specific for software teams, although some studies have based their investigation on models from other areas, e.g. [4].

Alongside with the characterization of factors that can affect software team effectiveness, a body of research has been produced that addresses software team effectiveness from a variety of perspectives. However, consistent with our focus discussed in the introduction, comprehensive review of the literature about software team effectiveness in general is out of our scope. In the rest of this section, we briefly present a summary of the main research related to our goals.

2.2.2. Software team composition and team factors

Team composition based on role allocation has been investigated from different perspectives in software engineering. For instance, Acuña and Juristo [1] and Acuña et al. [2] proposed and tested a model of role allocation in software teams based on required capabilities for tasks of software development. Their results showed that allocating individuals to roles based on the matching of existing and required capabilities produces positive effects on the outcomes of the software development. Using a distinct conceptual framework, Rajendran [69] investigated the effectiveness of certain software team composition using Belbin's Team Role Theory [12]. The author found that the use of Belbin's Theory in the analysis of team composition allowed the characterization of positive and negative features of teams, which important implications for team management. Both studies were conducted in industrial setting with teams of professional software engineers.

Team processes have also been focus of studies about software teams. Acuña et al. [3] investigated the relationships between team climate (operationalized as participative safety, support for innovation, team vision, and task orientation) and software quality, and found that team vision and participative safety are positively related to software quality. The same research group also investigated the relationships between individual characteristics (personality), task characteristics (interdependency and autonomy), team processes (task and social conflict, and cohesion), and team effectiveness (software quality and job satisfaction) [4]. In this study, the researchers used a simplified version of McGrath's input–process–output model [61] to build the hypotheses that guided their investigation. Consistently with the job characteristics theory [46,47], they found positive relationships between task autonomy and job satisfaction. On the other hand, individual satisfaction and team cohesion decreased in teams with high levels of conflict. These two studies were conducted in academic settings with teams of students.

Team cohesion has been postulated as one of the most important factors affecting team effectiveness and, consequently, has gained the attention of researchers studying software teams [31]. However, cohesion has been topic of intense debate in organizational psychology and teamwork research [19,33,41] and several distinct conceptualizations of the construct exist. There is no common understanding of which conceptualization to use in the research about software teams and studies have used a wide range of different definitions, making it very difficult (if not impossible) to compare and integrate their results.

2.2.3. Member personality and team effectiveness

Empirical studies investigating the effect of personality diversity in the composition of software teams have produced a wide range of evidences [24]. For instance, Bradley and Herbert [14] proposed and empirically tested a model to relate personality to team effectiveness. In their model, team performance was affected by four factors: leadership, team cohesion, communication, and heterogeneity of individual characteristics. They proposed a relationship between certain personality types (according to MBTI [66]) and the factors that affected team performance, and tested their propositions on two industrial case studies. According to their findings, certain team compositions with respect to personality have a positive impact on team performance. Gorla and Lam [38] found that personality diversity between the team leader and team members related to higher performance, but that diversity among team members showed no relationship with team performance. Miller and Yin found that a team with a diversity of individual characteristics related to information processing are likely to find more defects during software inspections [64]. White [77] reported that heterogeneous teams (high personality variety) were "optimum" when solving unstructured tasks while homogeneous (low variety) teams were "optimum" when solving structured tasks. On the other hand, Peslak [68] found no relationship between personality variety and project results.

McDonalds and Edwards [60] raised some concerns with respect to the research using personality tests in software engineering. The researchers analyzed 13 empirical studies in software engineering that focused on the use of such tests, including [14,38,64] and an earlier version of [77]. Their findings showed several validity threats in their use of psychometric tests, with obvious negative impact on the reliability of these studies. Therefore, results in this area, and their application, must be analyzed with care by researchers and practitioners.

2.2.4. Individual selection for software teams

Individual selection to build software teams has been much less studied than the other team issues discussed above. In the systematic mapping study described in Section 5.1, we found only 19 studies that explicitly addressed criteria used to select individuals to compose software teams and only eight of them were empirical studies (see Appendix B for the complete list of studies). Among the empirical studies, six were conducted in industry and two were performed with teams of students in academic settings.

Technical competence and skills was the most prevalent selection criteria, addressed by 18 out of the 19 studies. Behavior was found in one theoretical and six empirical studies, being the second most prevalent criteria. The other criteria found in the mapping study were Personality, Task Preference, Availability, and Peer Indication. No organizational related criteria, such as project or client priority, were found in the mapping study.

3. Research strategy

This paper reports the results of an empirical investigation performed between 2007 and 2012. The research strategy was based on the use of multiple research methods, mixing qualitative and quantitative empirical studies, complemented by a systematic mapping study. The research approach was iterative, with two iterations, each one composed of two phases. The results of each phase were used as input and source of improvements for the subsequent phases, as explained in Section 3.3.

3.1. Research questions

Two research questions guided the development of the various studies performed in this research. The first one was an exploratory question, intended to identify criteria for selecting team members used by project managers and team leaders in industry:

RQ1. What are the criteria used by software project managers in practice to select individuals to build software teams?

The second research question was intended to understand the relationships between the level of formalization in the use of the team building criteria and the effectiveness of the team measured by the success of the project:

RQ2. How is the consistent use of team building criteria related to project success?

3.2. Research methods

A mix-method approach was used to investigate how teams were built in practice and how the criteria used to build teams were related to several outcomes of project success. Methods and techniques of qualitative research [63,73] and systematic mapping study [55,67] were used in the identification of the team building criteria used in the practice of software engineering. Cross-sectional survey research [54] was used to find the correlations between the consistent use of team building criteria and project success goals. In the case of the survey research, we performed an internal [16], non-identical [52] replication to verify and extend the results of the original study.

3.3. Research iterations and phases

Fig. 1 illustrates how the main results of each phase, in Iterations 1 and 2, were used to inform and guide the subsequent phases.

3.3.1. Iteration 1: Qualitative interviews and survey research

In Phase I, we investigated the following research question:

RQ1. What are the criteria used by software project managers in practice to select individuals to build software teams?

We used semi-structured interviews with project managers and team members from various software companies to understand which criteria were used in practice to select software engineers to build software teams. We interviewed one project manager and two team members that worked in the same project. We used the data from the manager's interview as the primary source of data and the data from the team members to triangulate the data provide by the managers. Therefore, we collected data from 16 participants, six managers and 12 team members, from six software companies. The interviews were recorded and the audio was then transcribed and the resulting text was analyzed using coding techniques, and checked for consistency and reliability using member checking and triangulation. Eight categories representing team building criteria emerged from the data analysis.

In Phase II, our goal was to find answers to the following research question:

RQ2. How is the consistent use of team building criteria related to project success?

Towards this goal, an operational definition of the variables representing the consistent use of team building criteria was created to provide the operations necessary to measure, categorize, and manipulate these variables. Complementing the operational definition, a survey questionnaire was constructed to collect data from project managers regarding their consistent or inconsistent use of the team building criteria. Using this questionnaire, a cross-sectional survey was carried out with project managers from 24 software companies located in nine Brazilian States and the data from 48 projects were analyzed using statistical methods.

da Silva et al. [27] reported some central results of Phases I and II. In that article, limitations and threats to validity were identified, that provided opportunity for further investigations. We then addressed some of these limitations in Phases III and IV, as described below.

3.3.2. Iteration 2: Systematic mapping study and replicated survey research

Consistently with our goal of improving previous results, in Phase III we readdressed the research question RQ1. In this phase, we used a systematic mapping study to find empirical and theoretical studies addressing the use of team building criteria. From a set of 19 studies, we found six team building criteria used in empirical studies in industry. Four of them have already been found in Phase I and we added the two new ones to our set of criteria to be used in the next phase.

Finally, in Phase IV we readdressed the research question RQ2. We performed a replication of the survey research of Phase II, introducing variations to improve reliability of the results and also to verify the previous findings under a different set of conditions. In this phase, we collected data from 40 projects, from 10 software companies located in the City of Recife, State or Pernambuco, Brazil.

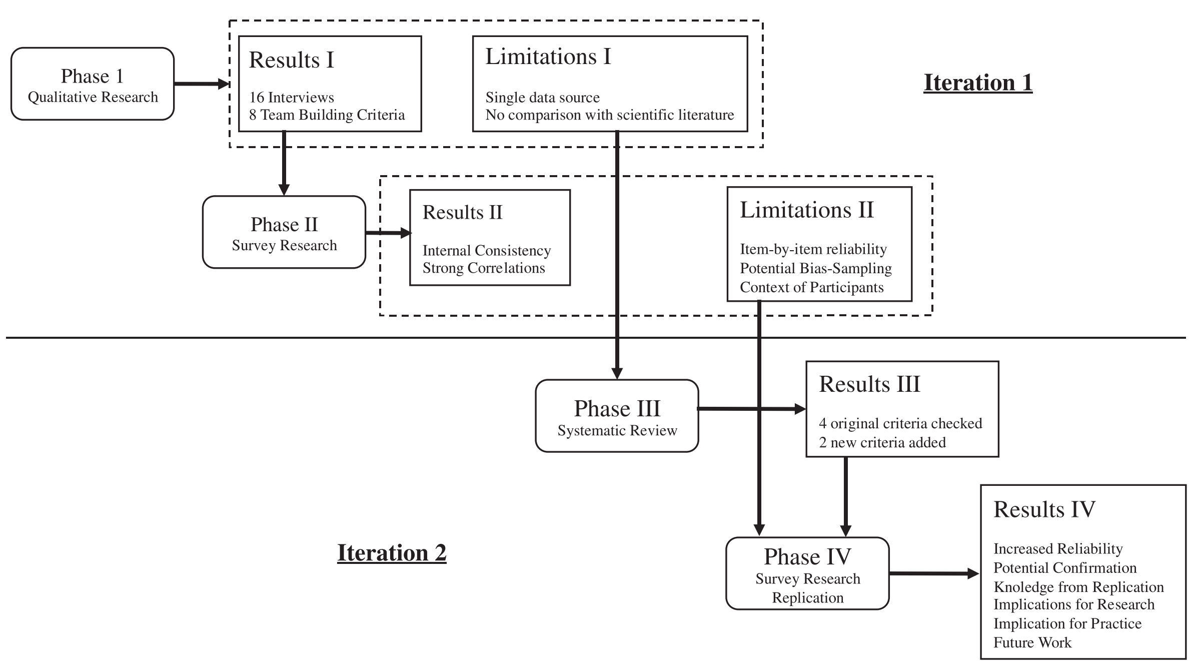


Fig. 1. Iterations and phases of the research.

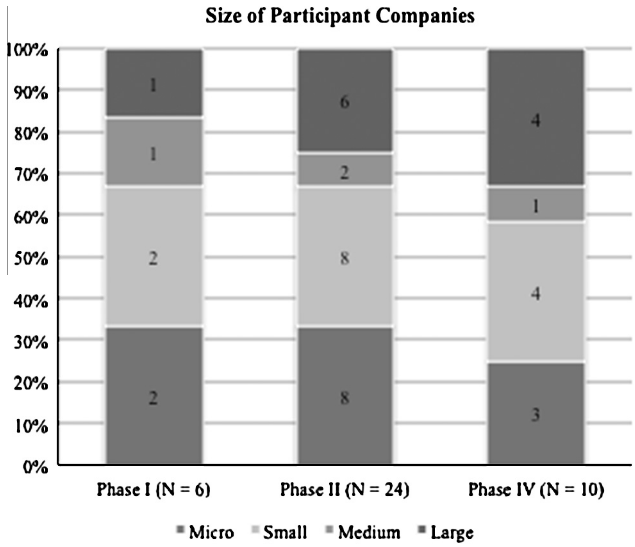


Fig. 2. Characterization of the participants according to company size.

Table 1
Size categories based on company's gross revenue.

Category	Annual gross revenue (US$ 1,000.00)
Micro	⩽600
Small	>600 and ⩽ 6,000
Medium	>6,000 ⩽ 18,000
Large	>18,000

3.4. Research context

This research was performed in Brazil, with a research team based on the Federal University of Pernambuco (UFPE). The empirical studies in Phases I, II, and IV were performed with Brazilian software companies and Fig. 2 shows the relative distribution of participants with respect to company size. The size categories were based on 2007 version of the official classification of the Brazilian Development Bank[1] (BNDES) and are described in Table 1. The systematic mapping study in Phase III was performed at UFPE, using the online library resources of the University.

There was no intersection between the companies that participated in Phases I, II, and IV. This was a deliberated strategy employed to allow variation in the sources of data used. We believe that this strategy is important to strengthen the results of empirical studies in general and, in particular, those performed in diversified industrial contexts.

4. Iteration 1: Qualitative interviews and survey research

In this section, we describe the design, results, and limitations of the studies conducted in Phase I (Section 4.1, 4.2, and 4.3) and Phase II (Section 4.4, 4.5, and 4.6) of Iteration 1.

4.1. Study design for Phase I: Qualitative research

The objective in this phase was to find initial answers to the research question RQ1 (Section 3.1). Taking a constructivist stance, this investigation was undertaken by looking into how project managers actually build their teams through the use of criteria to select individuals. The investigation was planned based on the guidelines presented by Seaman [73]: choice of data collection technique; construction of data collection instruments; participant selection; data collection; data analyses and synthesis. The main steps are described below.

4.1.1. Data collection technique

Seaman [73] suggests that interviewing is a suitable data collection technique to get "opinions, feelings, goals, and procedures" and the structure of the RQ1 indicates that this was the case. Therefore, we chose interviews as the primary data collection technique.

The primary source of data was defined to be the project manager and we certified, by asking the participant before scheduling the interview, that he or she was the person in charge of assembling the teams in the organization. It was found that data triangulation should be used to decrease the risk of managers answering what they wished or liked to happen instead of what actually happened in the practice of team building. Therefore, two team members working with each manager were also interviewed to provide data that could be checked against what was collected from their managers. Therefore, six project managers and 12 software engineers participated in Phase I.

4.1.2. Data collection instrument

The format of the interviews was semi-structured [63], guided by a script (interview guide presented in Appendix A). The script had 10 questions used to guide the conversation, structured in a "funnel shape": starting from general and more open questions and gradually moving to more specific, closed ones [73]. The audio of the interviews was recorded and the interviewers also took notes to help data analysis and synthesis.

4.1.3. Participant selection

In Phase I, we selected 18 software development professionals from 6 software companies located in Recife, State of Pernambuco, to participate in the interviews. Using purposive sampling, we chose companies with different sizes (Fig. 2) to increase the variability of the context. In each company, we purposively selected managers with different professional maturity. We tried to maximize diversity of professional maturity and management styles among the participants to increase the richness and diversity of the data.

4.1.4. Data collection

In a qualitative research, the goal is to interview as much participants as necessary to achieve diversity of information and category saturation. Therefore, the number of interviewees was not defined *a priori*. Instead, data was collected and primary analysis was undertaken after each interview checking for category saturation. The interviews were stopped when the researchers found that enough diversity had been achieved and the answers became repetitive without new information added.

Interviews happened in the work place of the project managers, during working hours. Each interview lasted 45 min on average. Two researchers conducted all interviews, one in the role of interviewer and the other taking notes, in the role of scriber.

4.1.5. Data analysis and synthesis

All interviews were recorded and the audio recordings were verbatim transcribed. NVivo™ 2[2] was used to support data analysis and synthesis. Data analysis began with open coding of the interview transcripts. Post-formed codes were constructed as the coding

[1] http://www.bndes.gov.br/SiteBNDES/bndes/bndes_en.

[2] http://www.qsrinternational.com.

Interview transcript: "*First the* *technical* *criteria. Who have worked with java, jsp, ejb, etc. A guy who has those* *competencies.*" [PM1]

Code: Technical competence

Interview transcript: "*We always look for the appropriate* *knowledge of technical aspects* *to see if it meets the demand that we have.*" [PM6]

Fig. 3. Open coding of interview transcripts.

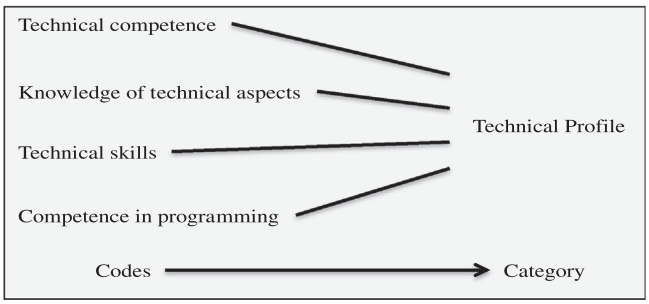


Fig. 4. Category building from the codes.

progressed and were attached to particular pieces of the text. Then, the codes arising from each interview were constantly compared to codes in the same interview and from other interviews. From the constant comparisons of the codes, we grouped codes into categories that represented the team building criteria. We present an example of the construction of the category C_1 – Technical Profile. We started from the interview transcripts by marking and coding relevant parts of the text (Fig. 3).

Then, we compared all codes built in the first step and grouped the codes referring to the same concept into mutually exclusive categories (Fig. 4). To perform this second step, we looked again into the coded interview transcripts to check the context in which the codes emerged, making sure that the apparently similar codes in fact referred to the same concept. Each team-building criterion was constructed in a similar way.

4.2. Results of Phase I

4.2.1. Categories describing the team building criteria

The team building criteria emerged from coding the data resulting from the interviews with project managers and the results were triangulated with the data collected from team members, as described in the previous section, resulting in eight categories. As pointed out in [73], some quantitative data extracted after the open coding process may provide valuable information. In Table 2, we show the frequency of occurrences of the team building criteria across the participants, using only the six interviews performed with project managers. The interviews (I_1–I_6) are presented in temporal order, from the first to the last.

Table 2
Frequency of occurrences of each criterion in the interviews.

Team building criteria	I_1	I_2	I_3	I_4	I_5	I_6	Total
C_1 – Technical Profile	X	X	X	X	X	X	6
C_5 – Personality	X	X	X	X	X	X	6
C_6 – Behavior		X	X	X	X	X	5
C_8 – Customer Importance	X	X	X		X		5
C_3 – Productivity	X			X	X	X	4
C_4 – Availability	X	X	X	X			4
C_2 – Individual Cost				X	X	X	3
C_7 – Project Importance			X	X	X		3

One of the most frequently cited criteria were C_1 – Technical Profile and C_5 – Personality. This confirmed that the managers tend to value technical competence as one of the most important criteria in team building, and this is consistent with the technical nature of the job. Besides, all managers mentioned the use of individual personality to build teams, which is consistent with the increasing number of researches on this theme, as described by Cruz et al. [24].

Another relevant information that can be obtained from Table 2 is related to the introduction of new criterion in an interview. This is important because it indicates whether the number of interviews was enough to achieve saturation of the concepts. In this case, interview I_4 was the last one that introduced a new criterion (C_2 – Individual Cost) which had not appeared before. Interviews I_5 and I_6 did not introduced any new criteria. This information was used to guide the decision of stopping the interviews after I_6. Therefore, six project managers and two team members working with each manager, in a total of 12 team members, were interviewed.

Data from the interviews with team members were used to triangulate the answers given by the project manager to increase reliability of the results. In the situations where explicit tests were applied (e.g., programming aptitude tests or personality questionnaires), team members were able to precisely define the criteria used and their information was consistent with what the managers answered. When a criterion was used informally (in the sense defined in Section 4.4.4), team members did not identified its use. In such cases, we investigated, when available, with the human resources department of the company to recheck the information. We only used data from the managers' interviews that we could check either from the triangulation or from other sources of information. This procedure did not leave out any criterion, but excluded occurrences of some criteria from two interviews. In these cases, managers mentioned the use of behavior (I_1) and productivity data (I_3), but we could not find corroborating evidence from a second data source.

4.2.2. Meanings and grouping of the team building criteria

Table 3 presents the meaning of the team building criteria. The meaning of each criterion was defined using the interview data coded during the data analysis, together with formal definitions of the criterion found in the literature. The definitions from the literature were used to clarify or formalize the meanings and solve ambiguities present in the ground data. Member checking was carried out to verify whether the participants agreed with the final definition of the criteria.

We then performed a deeper analysis of the criteria, looking into specialized literature on organizational psychology [6,7] with the objective of finding more abstract categories (Table 3). We realized that two large groups could be identified: one group related to the individual factors and another related to organizational factors. In the first group, some criteria were related to innate characteristics of the individuals (e.g., C_5 – Personality) while others were technically oriented criteria (e.g., C_6 – Technical Profile). In the second group, some criteria were internal to the organization (e.g., C_4 – Availability) while others were external or related to strategic business issues (e.g., C_8 – Customer Importance). In Section 4.5.3, we show that this conceptual structure has a very good match with a component structure that emerged using principal component analysis on the survey data related to the use of the criteria.

From the perspective of the practitioner, this grouping is relevant because it identifies those criteria that can be subject to

Table 3
Meaning and grouping of the team building criteria.

Groups	Sub-groups	Criteria	Meaning
Individual Factors	Innate Characteristics	C_5 – Personality	*Personaliy* is what distinguishes one individual from other individuals, that is, the character, thoughts, and feelings possessed by a person that "uniquely influences his or her cognitions, motivations, and behaviors in various situations" [5]
		C_6 – Behavior	*Behavior* is the set of actions and reactions observed in individuals in certain circumstances and in relation to the environment
	Technical Aspects	C_1 – Technical Profile	*Profile* is the appearance, character, or the general characteristics of a person. *Technical* means the specialized area of activity or human knowledge in a particular area. Therefore, *technical profile* is is connected directly to the technical capacity of the individual in a particular technology, language, platform, etc. This also includes expert knowledge in a module of a system or business process
		C_3 – Productivity	*Productivity* of the software engineer is the ratio between what an individual is capable of producing (in terms of lines of code, test cases tested, etc.) by a given amount of time.
Organizational Factors	Internal (or Operational)	C_2 – Individual Cost	*Individual Cost* refers to the impact on the project costs exerted by adding that individual to the project team.
		C_4 – Availability	*Availability* refers to the amount of time an individual is available to work in a new project.
	External (or Strategic)	C_7 – Project Importance	*Project Importance* refers to the strategic, competitive, or financial importance a project has to the company at the time the team is being assembled.
		C_8 – Customer Importance	*Customer Importance* refers to the strategic, competitive, or financial importance the customer for which the project will be developed has to the company at the time the team is being assembled.

managerial actions and those that cannot. Innate characteristics are thought as being relatively stable and difficult to change [70], while technical aspects can be subject to improvement, e.g., through training. On the organizational level, internal criteria can be managed by the organization, while external criteria is directed by broad competitive strategies and are, consequently, more difficult to manage.

4.3. Limitations of Phase I

We adopted strategies to decrease threats to credibility (internal validity) and consistency in the qualitative phase of the research, as proposed by Merriam [63]. Our main concern was with the category saturation of the team building criteria. We compared the interviews looking for new categories and only stopped adding new participants after two interviews that did not introduced any new category. Furthermore, we prompted the participants in the survey in Phase II to add new criteria that they thought were missing. No new criterion was added in that phase, increasing our confidence in the saturation. We used triangulation between the data from project managers and team members to check the veracity of the information provided by the managers. We also used member checking to verify the consistency of our data analysis and synthesis.

Despite these strategies, some limitations still remained. We collected the set of team building criteria from practitioners in the software industry in a given context and we did not check it against the scientific literature. Although this assured that the criteria were relevant to the industrial practice, it did not assure completeness. It was possible that other criteria had been researched in different contexts and were not being used in practice in the investigated context. For instance, the obtained criteria did not explicitly include issues related to collaborative values or competence in teamwork [75], for instance. This could mean that the participants did not consider these aspects as important, which reveals a limitation of the professionals with respect to these issues in that particular context. We tried to reduce this limitation by interviewing managers with different levels of experience and background. This limitation could also be originated from the types of questions asked in the interviews. Therefore, it would be important to extend the search for other team building criteria using other research methods (method triangulation). We performed this triangulation by adding a question in the survey instrument used in Phase II (and subsequently in Phase IV), and also by performing a systematic mapping study in Phase III.

4.4. Study design for Phase II: Survey research

The empirical research method used in Phase II was a cross sectional survey and the survey design is presented in the following subsections.

4.4.1. Goals and variables

Once the team building criteria were constructed from the practice of project managers in industry, the goal became to investigate how the use of the criteria in team building was related to project success. This is stated as the second research question RQ2 (Section 3.1). Therefore, the goal at this stage was to look for significant correlations between two variables: the use of team building criteria and project success.

The first variable was operationalized using the eight criteria constructed in the qualitative phase. Each criterion was measured in a three point ordinal scale representing the degree of formalization in the use of the criteria by the project manager to select team members, represented by: Low Formalization, Medium Formalization, and High Formalization. The precise meaning of "formalization of use" is presented in Section 4.4.4, where the survey instruments are presented.

To create an operational definition for the second variable (project success), a multidimensional measure of project success was used integrating several theoretical models of teamwork effectiveness [21,46,50,78]. From those models, we selected a set of six success goals: G_1 – Costs, G_2 – Time, G_3 – Scope, G_4 – Team Satisfaction, G_5 – Client Satisfaction, and G_6 – Project Manager Satisfaction. Each success goal was measured in a three point ordinal scale: Not Satisfied, Partially Satisfied, and Completely Satisfied.

4.4.2. Unit of analysis and sampling technique

The unit of analysis in this survey was the software development team associated with a particular software project. That is, we were interested in investigating how members of a software team were selected to work on a particular project and the results of this project. Two sampling procedures had to be performed, the first one to select the companies that would participate in the data collection and the second procedure to select the projects and teams which would be analyzed.

The software companies that participated in the survey were chosen from lists of companies provided by associations of information technology firms located in nine different cities and states of Brazil. We excluded the companies that participated in the Phase I. Invitations were sent to all companies in the lists and data was collect from those that chose to participate. The threats to validity related to self-selection have not been evaluated.

To select the teams and projects, each participant company was asked to select and provided information about two projects: one successful and one non successful. We used this sampling technique to obtain data from contrasting project and therefore to allow more variability in the data.

4.4.3. Participants

Information was collected from 48 projects developed by 24 Brazilian software companies, characterized in Fig. 2. The project manager or another professional (team leader, supervisor, business owner, etc.) that in the company was the person responsible for selecting professionals to build the software teams answered the survey instruments described below.

4.4.4. Instrumentation

The data collection instrument used in the survey was a self-applicable questionnaire. The respondents answered the questionnaire sent to them by e-mail after the company had agreed to participate in the study. The survey questionnaire had three parts: (1) general information about the company and the respondent, (2) level of formalization in the use of the team building criteria (hereafter referred to as the C Questionnaire), and (3) project success goals (hereafter referred to as the G Questionnaire). Parts 2 and 3 were duplicated to allow separate answers for the successful and the unsuccessful projects. The G Questionnaire was based on the work of Haggerty [49] and on the teamwork theories discussed above. In Tables 4 and 5, we show some items of Parts 2 and 3 of the survey instrument, respectively.

At the end of Part 2, we added an open-ended question asking the participant for any other criterion he or she used for team building that were not covered in the previous questions. This was done to increase our confidence on the category saturation. Participants did not add any new criterion in Phase II.

A glossary was provided with two parts: (1) the definition of the terms used both for the team building criteria and the success goals and (2) the definition of the meaning of "formal use" of a criterion. Participants were instructed to read the glossary before answering the questionnaire, but we cannot guarantee that they did as instructed. The purpose of the glossary was to decrease threats to construct validity due to different interpretations of the constructs among participants. This was a particular concern for the definition of "formal use". Therefore, its meaning was defined as follows:

Low Formalization: the criterion was not used to select the team member or it was considered but no assessment of individuals with respect to the criterion was carried out (e.g., productivity was thought to be important in the project but the project manager could not evaluate potential team members with respect to individual productivity).

Medium Formalization: the criterion was considered but no formal evaluation instrument was used and the assessment of the individuals with respect to the criterion was subjective (e.g., productivity was used as a criterion but the company did not have a record of past productivity and the project manager relied on his subjective assessment of the individuals).

High Formalization: the criterion was considered and formal evaluation instrument was used to provide an objective assessment of the individual with respect to the criteria (e.g., productivity was used and the company's records about employee productivity in past projects were used to assess the individuals).

The questionnaires were pre-tested by five project managers (that did not participate in the final sample). Improvements were made with respect to making the definitions clearer. In particular, the use of productivity as an example in the definition of the levels of formalization was added as a suggestion from one of the participants in the pre-test.

4.4.5. Statistics

Cronbach's alpha coefficient [40] was calculated to check the internal reliability of the measures of project success and use of team building criteria. We used the commonly accepted interpretations of the magnitudes of the Cronbach's alpha coefficient to assess internal consistency presented in Table 6.

We used principal component analysis [39] to identify the structure of the team building criteria expressed in the quantitative data and compare this component structure with the conceptual structure constructed in Section 4.2.2. Finally, the correlation between the use of selection criteria and project success was calculated using Spearman's rho correlation coefficient [40]. We used SPSS™ 17[3] software package to performed quantitative data analysis.

4.5. Results of Phase II

4.5.1. Testing reliability

First, we tested the reliability of project success measure using the six items G_1–G_6 discussed in Section 4.2.2 and found results to be reliable ($\alpha = 0.898$). This is an indication that the six items are related to a same latent construct, in this case project success.

We then wanted to check whether the use of the team building criteria, evaluated using a questionnaire with eight questions, would be internally consistent. In other words, we wanted to check whether the consistent (or inconsistent) use of the criteria was uniform by the project managers (all or most of the criteria used at the same level of consistency) or there were large variations in this use (some criteria being consistently used and some not). We then tested the results for the eight criteria C_1–C_8 and found the results to be reliable ($\alpha = 0.853$). This is an indication that managers

Table 4
Examples of Questions in the C Questionnaire (Phase II).

	Low	Medium	High
Variable: level of formal use of C_1			
1. How formal was the use of individual "Technical Profile" considered in the individual selection to compose the software team of this project?	( )	( )	( )
...			
Variable: level of formal use of C_3			
3. How formal was the use of individual "Productivity" considered in the individual selection to compose the software team of this project?	( )	( )	( )
...			

Table 5
Examples of questions in the G questionnaire (Phase II).

How were the goals listed below satisfied by the project A (*successful*)?	Not Satisfied	Partially Satisfied	Completely Satisfied
G_1. Costs	( )	( )	( )
G_2. Time (delivery date)	( )	( )	( )
...			

[3] http://www.ibm.com/software/analytics/spss.

Table 6
Interpreation of the Cronbach's alpha coefficient.

Cronbach's alpha	Internal consistency
$\alpha \geqslant .9$	Excellent
$.9 > \alpha \geqslant .8$	Good
$.8 > \alpha \geqslant .7$	Acceptable
$.7 > \alpha \geqslant .6$	Questionable
$.6 > \alpha \geqslant .5$	Poor
$.5 > \alpha$	Unacceptable

Table 7
Accumulated Scores for Each Team Building Criterion.

Criteria	All (N = 48)	% All	Upper (N = 24)	% Upper	Lower (N = 24)	% Lower
C_1 – Technical Profile	74	77%	47	49%	27	28%
C_7 – Project Importance	66	69%	40	42%	26	27%
C_8 – Customer Importance	62	65%	38	40%	24	25%
C_2 – Individual Cost	58	60%	36	38%	22	23%
C_6 – Behavior	56	58%	39	41%	17	18%
C_5 – Personality	55	57%	38	40%	17	18%
C_3 – Productivity	53	55%	40	42%	13	14%
C_4 – Availability	51	53%	33	34%	18	19%

tended to apply all or most of the criteria with the same level of consistency or formalization in a given project.

4.5.2. The use of the team building criteria

To be able to compare the use of the each criterion in our sample, we added the scores of a criterion C_i ($1 \leqslant i \geqslant 8$) for all projects P_j ($1 \leqslant j \geqslant 48$). In Table 7, we show the set of criteria ordered by this accumulated value for the entire set of 48 projects (column All). In this case, C_1 – Technical Profile was the criterion most consistently used, whereas C_4 – Availability was the least consistently used. The column % All show the percentage that the accumulated score represents of the highest possible accumulated score, which is 96 (48 projects multiplied by 2, which is the highest value in the scale).

We then examined the difference in the use of the criteria with respect to the scores in the project success items. To do this, we ordered the projects using the sum of all G_k ($1 \leqslant k \leqslant 6$) of each project and split the entire set of projects into the projects above the median of the accumulated G scores (Upper set) and those below the median (Lower set). We have a perfect separation of the two sets with a median 0 (zero) and 24 projects in the Upper and 24 in the Lower sets, because we had a balanced distribution of projects with respect to the success goals (which was intentionally constructed by our sampling technique).

The values in the columns Upper and Lower of Table 7 clearly indicates that the criteria are more consistently applied to build the teams of the more successful projects (Upper set) than of the less successful ones (Lower set), indicating the possibility of correlations between the two variables.

4.5.3. Analyzing the factor structure of the team building criteria

We performed a principal component analysis (PCA) [39] on the team building criteria using the scores of the individual criteria for all projects. Our goal was to analyze how the factors arising from the statistical data related to the grouping of categories performed in Phase I. In Table 8, we show the structure of the four components that explains 83% of the data variance. Non-significant cells are blank.

Table 8
Factor structure of team building criteria – original study.

	Component			
	Innate Characteristics	Technical Aspects	Strategic	Operational
C_6 – Behavior	.919			
C_5 – Personality	.910			
C_1 – Technical profile		.820		
C_2 – Individual Costs		.730		
C_3 – Productivity		.617		
C_8 – Customer Importance			.847	
C_7 – Project Importance			.733	
C_4 – Availability				.950

The structure of the four components has a good fit with the conceptual grouping of the criteria we built based on their theoretical meaning (Table 3, Section 4.2.2). We used the same sub-group names to name the components, to illustrate the consistent structure. This indicates that the grouping build from a conceptual perspective was confirmed by the survey data. In other words, the practical use of the criteria is consistent with the conceptual structure of the criteria.

4.5.4. Investigating correlations among variables

We tested the correlation between the use of the team building criteria and the project success goals, using Spearman's rho rank correlation. The correlation matrix has only seven cells with no significant correlation, indicated by the blank cells in Table 9. Goal G_1 – Costs does not correlate with four of the team building criteria (C_1– Technical Profile, C_4 – Availability, C_7 – Project Importance, and C_8 – Customer Importance), being the goal with the least number of significant correlations. Goal G_2 – Implementation date, G_3 – Functionality/scope, and G_5 – User satisfaction correlate significantly with all eight team-building criteria.

The criterion C_3 – Productivity presents the strongest correlations. Among the sub-groups, Technical Aspects followed by Innate Characteristics present the strongest correlations, which seem to indicate that the Individual Factors is the group more strongly correlated with project success. Criterion C_8 – Customer Importance correlates with only three goals, and C_7 – Project Importance with four. These are the criteria with the smallest number of significant correlations and also the weakest correlations.

4.6. Limitations of Phase II

Three main limitations represented threats to validity of the results of Phase II. First, because the G Questionnaire only had one question per G item, it was not possible to calculate internal consistency of the questionnaire on an item-by-item basis. Although the internal consistency of all items was high (α = 0.898), we believe the G Questionnaire could be improved by using more than one question per item.

Second, the sampling technique employed to select the projects required each company to internally select one successful and one non-successful project, as perceived by the company. We then compared the results of the G scores for successful and unsuccessful projects and successful projects scored significantly higher than the non-successful ones. This indicates that the sampling technique yielded consistent results regarding the G scores. This technique was chosen because we wanted to maximize the variety in the data related to project success and we achieved this objective. However, it is possible that this technique biased the answers of the C Questionnaire, with participants assigning more extreme

Table 9
Spearman's rho Correlation Between TBC and Project Goals – Original Study.

Group	Sub-group	Criteria	G_1 – Costs	G2 – Time	G3 – Scope	G4 – Team Satisfaction	G5 – Client Satisfaction	G6 – PM Satisfaction
Individual Factors	Innate Characteristics	C_5 – Personality	**.393****	**.493****	**.473****	**.465****	**.449****	.327*
		C_6 – Behavior	**.418****	**.523****	**.434****	**.494****	**.455****	.348*
	Technical Aspects	C_1 – Technical Profile		**.423****	**.607****	**.559****	**.476****	**.506****
		C_3 – Productivity	**.379****	**.627****	**.510****	**.634****	**.496****	**.585****
Organizational Factors	Internal (or Operational)	C_2 – Individual Cost	.289*	.329*	**.408****	.319*	**.375****	.324*
		C_4 – Availability		.348*	**.524****	**.400****	**.410****	**.426****
	External (or Strategic)	C_7 – Project Importance		.318*	.303*	.291*	**.421****	
		C_8 – Customer Importance		**.444****	.359*		**.513****	

* Significant at $p < .05$, 1-tailed.
** Significant at $p < .001$.

values (0s or 2s) than they would if they had not *a priori* labelled the project as successful or non-successful.

Third, the attempt to collect data from companies geographically dispersed in various cities and states in a large country like Brazil proved to be not adequate. The effort to contact companies and making sure that they respond the questionnaires was too big and the result too small regarding the number of participating companies. Another approach would be to carry out several surveys in more localized contexts and then compare and integrate the results, as it is usually performed in cross-case analysis in multiple case studies [79].

5. Iteration 2: Systematic mapping study and replicated survey research

In this section, we describe the design and results of the studies conducted in Phase III (Sections 5.1 and 5.2) and Phase IV (Section 5.3 and 5.4) of Iteration 2.

5.1. Study design of Phase III: Systematic mapping study

In this section and in Section 5.2, we briefly summarize the protocol and the results of systematic mapping study performed between 2009 and 2010 about software teams. This brief presentation omits many details that would be found in a full description of a systematic mapping study. Our goal here is to use some results of the study to corroborate and extend the set of team building criteria originated from the qualitative study of Phase I.

5.1.1. Goals and research questions

Our goal in this literature review was to build a mapping of the empirical and theoretical research about software development teams in software engineering. We were mainly interested in the empirical studies addressing the practice of software development in industry, although we also analyzed studies performed in academic settings with teams of students and theoretical propositions that had not been empirically tested. Two broad research questions and sub-questions guided our mapping study [55]:

MS-RQ1: What is a software development team?
MS-RQ1.1: What are all types of software development teams?
MS-RQ1.2: What are the properties or characteristics of each type of software development team?
MS-RQ1.3: How a software development team differs from other types of work teams?
MS-RQ2: What are the criteria used to build software development teams?
MS-RQ3: What are the antecedents of team effectiveness for software development teams?

In this phase of our investigation, we were interested in the answers to MS-RQ2, which were used to complement and verify the set of team building criteria found by the qualitative research in Phase I.

5.1.2. Data sources and search strategy

In this review, we performed automatic search in the following three search engines and indexing systems:

- ACM Digital Library – http://portal.acm.org.
- IEEEXplore Digital Library – http://www.ieeexplore.ieee.org/Xplore.
- Scopus – http://www.scopus.com.

Automatic searches were performed on the entire paper on all engines but Scopus, which did not perform full-text search. For this engine, the search was performed on Title and Abstract. For the automatic search, we used a search string composed of terms that are synonyms for "software development team". We performed two pilot tests until reaching the final string, presented in Fig. 5.

The results from the automatic search ($n = 2215$) were analyzed by looking at the title and abstract, and excluding the papers that were clearly not relevant. The search process finished with a set of 729 potentially relevant papers.

5.1.3. Study selection

A set of inclusion and exclusion criteria were applied to the 729 articles resulting from the search process, resulting in a final set of

"software team" OR "software project team" OR "project management team" OR "software development team" OR "software engineering team" OR "software system team" OR "software system development team" OR "software programming team" OR "software developers" OR "software developing team" OR "software development team" OR "program development team" OR "program developing team" OR "programming team" OR "agile team" OR "agile development team" OR "agile system team" OR "agile system development team" OR "agile programming team" OR "agile developing team" OR "open source team" OR "open source development team" OR "open source system development team" OR "open source programming team" OR "open source developing team"

Fig. 5. Search string.

studies that provided answers to at least one of the research questions or sub-questions presented in Section 5.1.1. More specifically, we included articles that presented empirical or theoretical research about (inclusion criteria):

(1) Definitions, types, classifications, or taxonomies of software teams (related to MS-RQ1).
(2) Selection criteria used to build software teams (MS-RQ2).
(3) Structure and characteristics of software teams (MS-RQ3).

We excluded papers that met at least one of the following five exclusion criteria:

(1) Written in any language but English.
(2) Not accessible on the Web.
(3) Invited papers, keynote speeches, workshop reports, and books.
(4) Incomplete documents, drafts, slides of presentations, and extended abstracts.
(5) Addressing areas of computer science that are clearly not software engineering (e.g., database systems, human–computer interaction, computer networks, etc.).

When a study had been published in more than one article, all versions were reviewed for the purpose of data extraction. In this case, we used the first publication in all time-based analyses to track the distribution of the studies over time.

Study selection was performed by one researcher (the sixth author) and verified by a second researcher (the first author) in independent processes. Initially, the sixth author selected 35 articles for inclusion, from the initial set of 729 articles. The first author then looked the excluded articles to check for the possibility of relevant studies being missed in the first selection. This process resulted in a list of six articles that could be considered borderline with respect to exclusion criteria number 5. In a face-to-face meeting, the two researchers decided to include two articles from the list of six borderline articles that were from information systems area but addressed software development teams. The remaining four borderline articles were finally excluded because they did not address software teams.

5.1.4. Data extraction and quality assessment

In the data extraction process, we adapted the data extraction form used by Dyba and Dingsoyr [34]. We extracted information organized in four perspectives: Identification of the Study, Study Description and Methodology, Results of the Study, and Answers to the Research Questions. We used the same criteria used by Dyba and Dingsoyr [34] to assess the quality of the selected studies.

5.1.5. Data synthesis

A coding process similar to the one described in Section 4.1.5 for the qualitative study was used to build categories for the information related to the research questions. The evidence extracted from the studies in the data extraction process was coded and categories were built by constant comparison of the codes. In this way, in the synthesis process for the answers for MS-RQ2, we collected a set of team building criteria presented in the literature and used this set to complement the criteria found in Phase I.

5.2. Results of Phase III

In our mapping study, 19 articles described empirical or theoretical studies about software teams that provided answers to MS-RQ2. We present a summary of these articles in Table 10 and the complete reference list in Appendix B. The results described in Section 5.2.1 were extracted from these 19 articles.

Table 10
Articles in the mapping study describing team building criteria.

Study Id	Year of publication	Quartile of quality score	Development method	Type of publication
S012	1986	1	Traditional	Conference
S018	2004	4	Not informed	Journal
S019	2010	1	Agile	Conference
S026	1997	1	Traditional	Workshop
S028	2006	1	Agile	Conference
S030	2006	3	Agile	Conference
S032	2009	1	Agile	Conference
S035	2001	1	Not informed	Journal
S041	2010	4	Not informed	Journal
S048	2009	3	Agile	Journal
S052	2008	1	Not informed	Conference
S053	1997	2	Not informed	Conference
S061	2009	4	Agile	Workshop
S064	2000	2	Not informed	Conference
S066	2010	1	Not informed	Conference
S077	2010	1	Not informed	Conference
S079	2008	1	Not informed	Conference
S081	2009	3	Traditional	Others
S094	2005	3	Not informed	Journal

5.2.1. Categories describing the team building criteria

The team building criteria were extracted from the 19 articles of Table 10 and the coding and synthesis process resulted in six categories, four of which have been found in Phase I. Therefore, the mapping study added two new team building criteria and confirmed the use of four of the eight criteria found in Phase I. This result is summarized in Table 11, with the references to the studies in which the criteria were found.

As in Phase I, C_1 – Technical Profile was the two most prevalent criteria, with C_1 cited in 18 studies. We did not find evidence for criteria C_2 – Individual Cost, C_3 – Productivity, C_7 – Project Importance, and C_8 – Customer Importance. Two new criteria were found C_9 – Task Preference and C_{10} – Peer Indication. Although these criteria were mentioned by few studies, we decided to include in our set for completeness and to assess their use in practice in the next phase of the research.

5.2.2. Meanings of the team building criteria

In Table 12, we present the meaning of the two new team building criteria found in the mapping study (C_9 and C_{10}). Similarly to Phase I, the meaning of each criterion was defined using the information presented in the articles, together with formal definitions of the criterion found in the literature.

5.2.3. Grouping the criteria

We then included the two new criteria in the grouping structure constructed in Section 4.2.2 (Table 13). For this classification, we used the information provided in the studies analyzed in the mapping study and the literature on organization psychology [6,7]. We include C_{10} – Task Preference in the Innate Characteristics sub-group of the Individual Factors, because these preferences tend to be more stable and related to personality and behavior, as we demonstrated in another study [26]. The criterion C_9 – Peer Indication was classified as an Internal Organizational Factor, and this was consistent with the factor analysis we performed in Phase IV (Table 17).

5.3. Study design of Phase IV: Replicated survey research

In Phase IV, we performed a non-identical, internal replication of the survey research conducted in Phase II. We used the limitations and threats to validity of Phase II to guide the design of Phase IV. We observed the findings and recommendations of Juristo and

Table 11
Team building criteria extracted from the literature.

Team building criteria	Empirical studies	Theoretical studies (not empirically verified)	Total
C_1– Technical Profile	S012, S019, S030, S032, S041, S053, S064, S094	S018, S026, S028, S048, S052, S061, S077, S079, S066, S081	18
C_6– Behavior	S019, S030, S032, S053, S064, S094	S066	7
C_5– Personality	S041, S094	S061, S066, S081	5
C_9 – Task Preference	S032, S064		2
C_4– Availability	S032		1
C_{10} – Peer Indication	S035		1

Table 12
Meaning of each team building criterion.

Criterion	Meaning
C_9 – Peer Indication	*Peer Indication* is an indication or a referal for an individual provided by trustworth source (a reference provided by a person to whom the project manager trusts).
C_{10} – Task Preference	*Task Preference* is used when the project manager tries to match the personal preferences for taks or team roles of a given individual to the tasks and roles of the project.

Table 13
Grouping the extended set of criteria.

Group	Sub-group	Criteria
Individual Factors	Innate Characteristics	C_5 – Personality
		C_6 – Behavior
		C_{10} – Task Preference
	Technical Aspects	C_1 – Technical Profile
		C_3 – Productivity
Organizational Factors	Internal	C_2 – Individual Cost
		C_4 – Availability
		C_9 – Peer Indication
	External	C_7 – Project Importance
		C_8 – Customer Importance

Vegas [52] for the inclusion of variations in the survey design and to compare the results with Phase II. We also followed the guidelines of Carver [20] to write the report of the replication.

5.3.1. Goals of the replication

In this replication, we were interested in improving the study design used in Phase II and compare the findings to check how the variations between the designs influenced the results. In particular, we aimed at: (1) improving the G Questionnaire to increase its internal consistency; (2) designing another sampling technique that would reduced the risk of bias in the answers to the C Questionnaire; and (3) to verify under which conditions the results of Phase II would still hold if we changed the context which participants were sampled from.

5.3.2. Variables

The same variables studied in Phase II were studied in Phase IV. The same operational definition for the use of the team building criteria was used. In the improvement of the G Questionnaire, we changed the operationalization of the project success criteria to a five point ordinal scale, as explained in Section 5.3.5.

5.3.3. Unit of analysis and sampling technique

The same unit of analysis used in Phase II was used in Phase IV. We changed the sampling technique to avoid the potential bias of the project selection technique used in Phase II, discussed in Section 4.6. We randomly selected 10 companies from the list software firms of the Porto Digital Science Park[4] in Recife. To select the projects, we asked the companies to provide a list of projects that fulfilled a simple list of criteria:

- Software development projects involving at least the activities of software building (coding) and software testing as part of the software development life cycle. Projects involving only testing or maintenance tasks (or both, but no coding of new requirements) should not be included.
- On-going projects with at least one release delivered and accepted by the client or customer, or projects finished in the last 12 months.
- Projects for which it was possible to have access to the Project Manager or other person responsible for selecting the members of the project team.

We expected to receive a large enough list of project from the companies and then to randomly select our unities of analysis from this large list to achieve a balance between successful and unsuccessful projects, as we had in Phase II. However, the lists returned did not have enough projects to allow a second selection and we used all projects in the data collection. We will discuss how this new sampling technique influenced the results of Phase IV and comparability of with the results of Phases II in Section 6.1.1.

5.3.4. Participants

Information was collected from 40 projects developed by 10 Brazilian software companies located at the Porto Digital Science Park. The characteristic the sample regarding the size of the companies was presented in Section 3.4. As in Phase II, the project manager or another professional (team leader, supervisor, business owner, etc.) responsible for selecting software engineers to build the software teams answered the survey instrument.

5.3.5. Instrumentation

As in Phase II, the data collection instrument used in the survey was a self-applicable questionnaire applied to the participants. The questionnaires were sent by e-mail after the company had agreed to participate in the research. The answers were collected in person by one of the members of the research team.

One of the goals of the replication was to improve the reliability of the instruments and in particular the G Questionnaire, as explained before. Therefore, we performed some changes in the instruments as follows:

- A section was added to collect information about the technical aspects of the project, such as type of development method

[4] http://www.portodigital.org.br.

used, and about the team structure used in the project, for instance command and control (hierarchical) or self-managed (flat).
- We changed the G Questionnaire to allow the evaluation of internal reliability on an item-by-item basis. We added an extra question for each item, such that each success goal G_1–G_6 was assessed by two questions. We kept the original questions in the positive form and added the second question in the negative form. We then added two questions to assess the success of the project in general terms. We mixed the questions such that those related to the same item were as far apart as possible. Finally, we changed the three-point scale to a five point Likert type scale varying from −2 (strongly disagree) to +2 (strongly agree), rephrasing the questions accordingly.
- We added two questions in the C Questionnaire to assess the use of the two new criteria found in Phase III. We also added glossary terms to explain these new criteria.

The final instrument was, therefore, composed of four parts: (1) general information about the company and the respondent, (2) technical information about the project and project management style, (3) the use of the team building criteria (C Questionnaire), and (4) project success goals (G Questionnaire). The four parts of the survey instrument are presented in Appendix C.

5.3.6. Statistics

As in Phase II, Cronbach's alpha coefficient was calculated to check the internal reliability of the measures of project success. We also used this coefficient to assess the consistency in the use of team building criteria, as we did in the original study. Finally, the correlation between the use of selection criteria and project success was calculated using Spearman's rho correlation coefficient. Data analysis was performed with support of IBM® SPSS® Statistics Version 20. Although we used a different version of the SPPS software in Phase II, there is no reason to believe that this would introduce problems in the comparison of the results of both phases.

5.3.7. Summary of design variations

In Table 14, we summarize the main variations in the study design between the original (Phase II) and the replicated surveys (Phase IV). Besides these intentional design variations, it is important to remember that the studies were conducted with a four years interval, with the data collection of Phase II happening in 2007–2008 and in Phase two happening in 2011–2012. The implications of these variations on the results of Phase IV and on the comparability of these results with Phase II are discussed in Section 6.1.

5.4. Results of Phase IV

5.4.1. Testing reliability

We tested the reliability of the new version of the G Questionnaire using Cronbach's alpha coefficient, for each pair of questions related to the six original success goals and the extra item related to overall project success (Table 15). The internal consistency for all items was good or acceptable, except for G_4 – Team Satisfaction, which was questionable. We also calculated the alpha coefficient for all the 14 questions in the G Questionnaire and obtained an excellent level of internal consistency ($\alpha = 0.900$). The new version of the G Questionnaire with two questions for each item can be considered internally consistent. We, therefore, achieved our first goal of improving the instrument with acceptable level of reliability.

We then checked whether the use of the team building criteria was consistent. As in Phase II, we wanted to check whether use of the criteria was uniform by the project managers (all or most of the criteria used at the same level of formality) or there were large variations in this use (some criteria being formality used and some not). We tested the internal consistency for the eight criteria C_1–C_8 and found the results to be reliable ($\alpha = 0.818$). This is an indication that managers tended to apply all or most of the criteria with the same level of formalization in a given project, which confirms the results of Phase II.

5.4.2. The use of the team building criteria

Similarly to what was presented in Section 4.5.2, we explored the use of each criterion in the new sample of projects (Table 16). In Phase IV, the total possible accumulated value for each criterion was 80 because we had 40 projects in our sample. Therefore, when comparing Tables 7 and 16, it is important to remember to use the percentages and not the absolute numbers.

In this phase, we also separate the full set of projects in the Upper and Lower sets, as performed in Phase II (Section 4.5.2). This

Table 15
Cronbach's alpha coefficient for the items of the G questionnaire.

Items	α	Internal consistency
G_1 – Costs	0.884	Good
G_2 – Implementation date	0.793	Acceptable
G_3 – Functionality/scope	0.750	Acceptable
G_4 – Team Satisfaction	0.649	Questionable
G_5 – User satisfaction	0.761	Acceptable
G_6– Project manager satisfaction	0.712	Acceptable
C7 – Overall Success	0.855	Good

Table 14
Main design varitions between original and replicated survey.

	Original (Phase II)		Replication (Phase IV)		Limitations addressed (Section 4.6)
Population	Projects and teams from software companies located in nine Brazilian States.		Projects and teams from software companies located in one Science Park, in one Brazilian State.		3rd
Sampling Technique	Companies	Projects and Teams	Companies	Projects and Teams	2nd
	Randomly sampled from the lists provided by software company associations in each State.	Each company internally selected two projects: one successful and one not successful	Randomly sampled from the list of software companies located at Porto Digital Science Park.	Each company internally selected projects that satisfied a set of criteria defined by the researchers (Section 5.3.3).	
G Questionnaire (questions)	Six questions, each one related to one success criteria.		14 questions, two related to each of the six success criteria and two for general success, each pair of questions phrased in positive and negative form.		1st
G Questionnaire (scale)	Three point ordinal scale		Likert type 5-point ordinal scale		

Table 16
Use of Team Building Criteria in the Original and Replicated Studies.

Criteria	All ($N = 40$)	%	Upper ($N = 22$)	%	Lower ($N = 18$)	%
C_1 – Technical Profile	63	79%	35	44%	28	35%
C_4 – Availability	54	68%	25	31%	19	24%
C_7 – Project Importance	47	59%	22	28%	17	21%
C_2 – Individual Cost	44	55%	30	38%	24	30%
C_8 – Customer Importance	42	53%	14	18%	12	15%
C_3 – Productivity	39	49%	16	20%	10	13%
C_{10} – Task Preference	28	35%	30	38%	17	21%
C_6 – Behavior	26	33%	27	34%	15	19%
C_5 – Personality	26	33%	14	18%	11	14%
C_9 – Peer Indication	25	31%	16	20%	12	15%

Table 17
Factor structure of team building criteria – replicated study.

	Component				
	External (or Strategic)	Innate Characteristics	Economic Aspects	Technical Aspects	Internal (or Operational)
C_8 – Customer Importance	.882				
C_7 – Project Importance	.845				
C_5 – Personality		.846			
C_6 – Behavior		.778			
C_3 – Productivity			.927		
C_2 – Individual Cost			.682		
C_{10} – Task Preference				.832	
C_1 – Technical Profile				.822	
C_4 – Availability					.438
C_9 – Peer Indication					.928

separation in the current sample produced a median value of 8 (for a range between -12 and 12 in the accumulated score), with 22 projects above the median (Upper set) and 18 below the median (Lower set). The median is well above the medium point in the scale (zero) and this indicates that the sample is skewed towards higher scores for the G items, as opposed to Phase II in which the sample was balance around the medium point of the scale. We discuss the implications of this difference in the two samples in Section 6.1.1.

As in Phase II (Table 7), C_1 – Technical Profile was the most consistently used criteria for the entire set of projects and for the Upper and Lower sets as well. One difference was that C_4 – Availability was the least consistently used criteria in Phase II and the second most consistently used in Phase IV.

Comparing the percentages of each score in Tables 7 and 16 we noticed no general trend in improvement or worsening in the consistent use of the criteria in the entire sets of projects. Some criteria were more consistently used by projects in Phase II and others more consistently used by projects in Phase IV. Table 16 also shows that in Phase IV the projects in the Upper set use all criteria more consistently than the ones in the Lower set (as in Phase II), but in Phase IV the difference is much smaller. In other words, in Phase IV the projects in the Lower set used the criteria almost at the same level of consistency as the projects in the Upper set. We discuss this result in Section 6.1.1.

5.4.3. Analyzing the factor structure of the team building criteria

As in Phase II, we performed a principal component analysis (PCA) [39] on the team building criteria using the data of Phase IV projects. Our goal was to analyze how the factors arising from the statistical data related to the grouping of categories performed in Phases I and III. In Table 17, we show the structure of the five components that explains 84% of the data variance.

As in Phase II, the structure of the five components showed a good fit with the conceptual grouping of the criteria we built based on their theoretical meaning (Table 13, Section 5.2.3). We used mostly the same sub-group names to name the components, to illustrate the consistent structure. In this case, we created a new component name Economic Aspects to group C_3 – Productivity and C_2 – Individual Cost. It is likely that this new structure will be used in our future research, but we have not checked its conceptual consistency with the literature yet. This result indicates once more that the grouping built from a conceptual perspective was well represented by the survey data, increasing our confidence that our conceptual structure was mostly consistent with the practice, with minor refinements to be performed as part of our future investigations.

5.4.4. Investigating correlations among variables

As in Phase II, we tested the correlation between the use of the team building criteria and the project success goals using Spearman's rho rank correlation. In Phase IV, the correlation matrix (Table 18) only showed three significant correlations, a significantly different result with respect to Phase II (Table 9). We discuss possible explanations for this result in Section 6.1.

6. Discussions

In this section, we discuss the results of all four phases, emphasizing the explanation of the differences between the results of Phase II and Phase IV (Section 6.1). We also discuss the implications of our findings for research (Section 6.2) and for industrial practice (Section 6.3).

6.1. Using variations to explain the results

The replicated survey (Phase IV) used three main design variations, summarized in Table 14, Section 5.3.7. In this section, we investigate how these variations could explain the difference between Phase II and IV with respect to the correlations between team building criteria and project success goals. We looked at the effects of the new sampling technique (intended variation) in Section 6.1.1. We also examined how the change in the context from which companies were sampled (intended variation) together

Table 18
Spearman's rho correlation between TBC and project goals – replicated study.

Group	Sub-group	Criteria	G_1 – Costs	G2 – Time	G3 – Scope	G4 – Team Satisfaction	G5 – Client Satisfaction	G6 – PM Satisfaction
Individual Factors	Innate Characteristics	C_5 – Personality						
		C_6 – Behavior						
		C_9 – Task Preference						
	Technical Aspects	C_1 – Technical Profile						
		C_3 – Productivity						
Organizational Factors	Internal	C_2 – Individual Cost				.355*		
		C_4 – Availability						
		C_{10} – Peer Indication	.	.307*				
	External	C_7 – Project Importance	.454**					
		C_8 – Customer Importance						

* Significant at $p < .05$, 1-tailed.
** Significant at $p < .001$.

with the four year gap between the two studies (unintended and unavoidable variation) might have significantly changed the characteristics of the projects in the two samples with respect to type of development method used (unintended and unexpected variation), and how this could also complementary explain the results, in Section 6.1.2.

6.1.1. Intended variation in the sampling technique

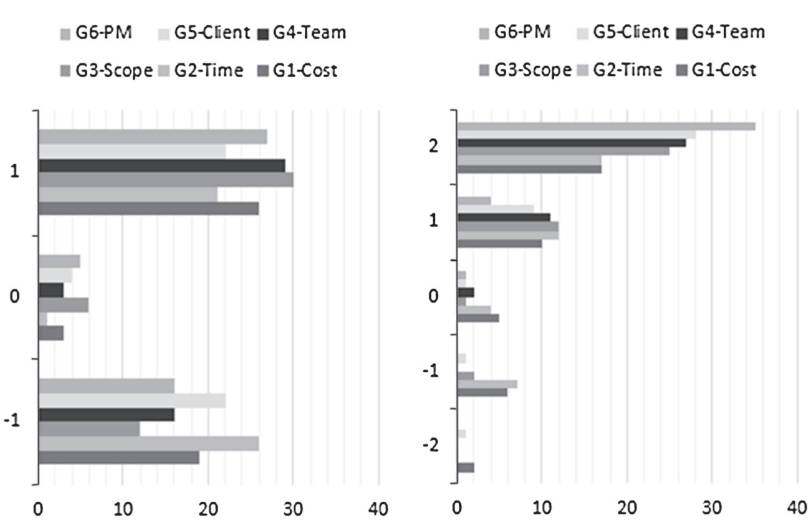
The variation in the sampling technique produced very different distribution of the projects regarding the success goals, as can be seen by comparing the charts in Fig. 6. Fig. 6a shows the scores for the G items for the original studies and Fig. 6b shows the G items for the replication study.

The original sampling technique forced each participant company to choose two projects with contrasting outcomes related to the project success goals, thus yielding a balance distribution between positive and negative scores on the G items. The sampling technique used in the replicated study left the choice of project more open and resulted on a skewed distribution towards positive G scores. To illustrate the difference in the samples and then to explain the difference in the results, we examined the Upper and Lower sets, built as in Section 4.5.2 by ordering the projects using the sum of all G_k $(1 \leqslant k \leqslant 6)$ and then taking the projects above and below the median of the accumulated G scores. Fig. 7 illustrates the difference between the samples in Phases II and IV with respect to the median, minimum and maximum scores, and the distribution and size of the Upper and Lower sets.

Fig. 7 shows that the Upper and Lower sets in Phase IV are likely to have similar characteristics of G scores as the Upper set in Phase II, and that the sample in Phase IV does not have a set of projects with similar G scores as the Lower set in Phase II. In Section 5.4.2, we pointed out that the difference in the use of the team building criteria between project in the Upper and Lower sets was much smaller in the sample of Phase IV than in Phase II, and Fig. 6 also explains why this could be expected.

Therefore, if the projects in the Lower and Upper sets in Phase IV are similar to the projects in the Upper set in Phase II, we would expect to find few correlations in the Upper set of Phase II as well. Besides, it would be reasonable to expect that the Lower set of Phase II would also exhibit few correlations. Table 19 confirms that indeed the Upper set of Phase II exhibit small number of significant correlations, with a similar behavior as the entire set of projects in Phase IV. Although the number of correlations is slightly bigger for the Lower set, it is still significantly different from the full set, as can be seen in Table 9.

Based on these results, we have a consistent explanation for not finding the correlations in Phase IV and furthermore we have support for the contention that if we had a balanced sample in Phase



(a) Original Study – Phase II (b) Replicated Study – Phase IV

Fig. 6. Differences in the scores of G items in Phases II and IV.

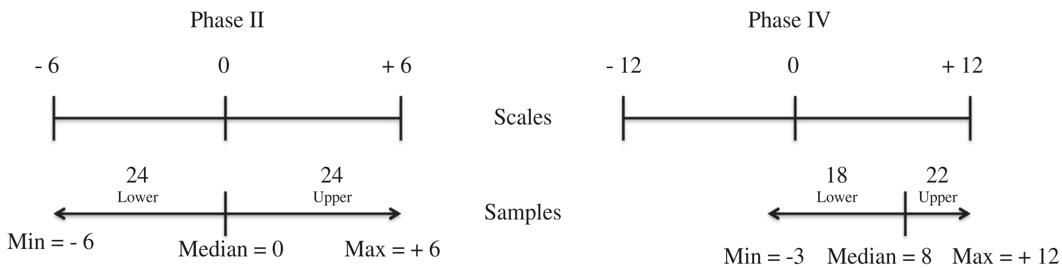


Fig. 7. Differences in the Upper and Lower sets in Phase II and IV.

Table 19
Spearman's rho correlation – Upper and Lower sets in phase II.

Criteria	Upper set ($N = 24$)						Lower set ($N = 24$)					
	G_1 – Costs	G2 – Time	G3 – Scope	G4 – Team Satisfaction	G5 – Client Satisfaction	G6 – PM Satisfaction	G_1 – Costs	G2 – Time	G3 – Scope	G4 – Team Satisfaction	G5 – Client Satisfaction	G6 – PM Satisfaction
C_5 – Personality									.361*			
C_6 – Behavior												
C_1 – Technical Profile									.361*			
C_3 – Productivity			.365*									
C_2 – Individual Cost									.502**			
C_4 – Availability					.344*				.522**		.390*	
C_7 – Customer Importance											.357*	
C_8 – Project Importance			.562**								.566**	

* Significant at $p < .05$, 1-tailed.
** Significant at $p < .001$.

IV we would find similar correlations as in Phase II. Therefore, although we cannot say that Phase IV confirmed the results of Phase II, we can conclude that the results of the two phases are compatible with each other. In fact, we can also conclude that we could use the results of Phase II to predict the results of Phase IV. Therefore, it is reasonable to continue holding onto the conclusions about the existence of correlations in our study.

6.1.2. Unintended variation in the characteristics of the projects

An important unintended variation between the two samples was identified when we were analyzing the contextual data to explain the differences in the results or the correlations. This variation is related to the type of development method used in the projects. As illustrated in Fig. 8, all projects in Phase II used a traditional development method, most of them based on RUP or a mix of RUP and Waterfall. In contrast, over 70% (29/40) of the projects in Phase IV used Agile methods.

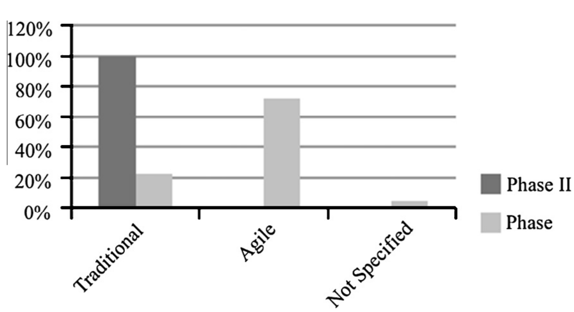


Fig. 8. Type of development method used by projects in Phase II and Phase IV.

If the difference in the development method also explained the difference in the results, this would be important for two reasons. First, we would have a stronger argument for confirming the existence of the correlations. Second, this would give some empirical insights about the effects of Agile methods in software team effectiveness. In Table 20, we show the correlations for the two sets of projects, the first containing the nine projects that used the traditional methods and the second with the 29 projects that used Agile methods. The grey cells in Table 20 indicate where we found the significant correlations for the entire set of 40 projects.

As can be seen in Table 20, the two sets of projects produced very distinct results. Although it is still significantly different from the results of Phase II (explained above by the differences in the samples) the result for the set of traditional projects shows many more correlations than the set of Agile projects, which in turn show the same result as the full set of projects (as expected, because the number of Agile projects dominate in the full set).

Although the number of traditional projects is very small to allow strong statistical conclusions, this result seems to indicate that the difference in the development method offers another explanation for the difference in the correlations between the use of team building criteria and project success goals. In this case, because we have not designed our study to investigate this potential effect, we must be careful with the conclusions we make and we will discuss this problem in more detail in Section 6.2.3.

6.2. Implications for research

In this section, we discuss some implications of our result for the research on teamwork in software engineering and, in general, about study design and replication of empirical research.

Table 20
Spearman's rho correlation – replicated study/traditional method.

Criteria	Traditional projects ($N = 9$)						Agile projects ($N = 29$)					
	G_1 – Costs	G2 – Time	G3 – Scope	G4 – Team Satisfaction	G5 – Client Satisfaction	G6 – PM Satisfaction	G_1 – Costs	G2 – Time	G3 – Scope	G4 – Team Satisfaction	G5 – Client Satisfaction	G6 – PM Satisfaction
C_5 – Personality	.772**	.568*				.621*						
C_6 – Behavior		.873**										
C_9 – Task Preference			-.652*									
C_1 – Technical Profile	.597*					.804**						
C_3 – Productivity		.731*										
C_2 – Individual Cost		.703*								.437**		
C_4 – Availability		.651*										
C_7 – Project Importance	.668*				.620*		.336*					
C_8 – Customer Importance		.557*			.690*							
C_{10} – Peer Indication						.595*		.412*				

* Significant at $p < .05$, 1-tailed.
** Significant at $p < .001$.

6.2.1. Survey design and choice of sampling technique

The choice of the sampling technique is a delicate and complex part of the design of surveys (and other types of empirical studies as well) due to its potential to introduce bias and other threats to validity. This study shows a case in which an attempt to reduce a potential bias through the use of an alternative sampling technique produced a sample of projects not totally adequate to the study. This seems to be an instance of a general problem that shall be found whenever one needs to select a sample with enough variation in one factor and knowing this selection *a priori* can potentially bias the measures of other factors. The sampling technique used in Phase II guaranteed the balanced distribution of projects, with respect to project success goals, but could have introduced bias in the answers about the use of team building criteria. The technique employed in Phase IV reduced or even eliminated the potential bias, but did not produce a balanced sample. In this case, we could argue that participants would have a natural tendency to select only successful projects (indeed, the most successful ones in the company) and therefore would not select unsuccessful projects unless they were forced to do so as part of the sampling technique, which could in turn introduce bias.

We can think of at least two possible solutions to avoid this problem, each one with its own shortcomings. The first one would be to sample the projects in two stages. In a first stage, we would use a sampling technique similar to Phase IV and collect a large set of projects, for instance, by asking the companies in the study to provide information about all their projects. Then, in a second stage, we could randomly sample over the result of the first stage using blocking to get the wanted distribution of some of the factors. One of the problems with this technique is that it is likely to require a much large number of units of analysis for relatively few useful data points.

The second solution would be to use a blinding technique, similar to what is used in medical trials. In our case, one person in the company would select successful and unsuccessful projects and a second person would answer the C and G questionnaires without knowing the results of the initial selection (blind to the selection process). This technique would increase the complexity of the data collection process, thus increasing the possibility of inconsistencies in the data set. Besides, it would be difficult to guarantee that the second person was completely "blind". In some cases, mainly in small companies, it would also be difficult to find two participants with enough knowledge of the projects to make this strategy feasible. Finally, it is very likely that the person answering the questionnaires would easily identify that criterion used to select the projects, which could again bias the results.

As a general conclusion, this is a difficult problem that does not seem to have a simple and general solution in practice. Researchers facing similar situations should understand the potential problems that any choice of sampling technique will introduce in their research and work to reduce these problems.

6.2.2. Generating knowledge from the replication

In the design of replications of empirical studies, we are always faced with a number of potential variations between the replication and the previous studies (original and previous replications) [52]. Some of these variations are intentional, introduced to improve the previous design or to investigate other hypothesis in different contexts. Other variations are not intended and some of these are unavoidable. In fact, it is possible that the researchers will not be aware of these unintended variations until late in the research development. In our study, each of these types of variations had different impacts in the result of the replication.

Our goal with the survey replication was not only to verify the results of the original survey but also to improve the original design and then identify contextual factors that could have an effect on the results of the study. To achieve these goals we introduced two intended variations in the design of the replication. First, we changed the sampling technique and the impact of this change has been extensively discussed in the previous section. Second, we sampled the companies located in a single science park instead of from different locations. This second variation allowed faster and more reliable data collection, because the researchers collecting data could directly interact with the participants. Therefore, this variation brought benefits for the development of the study.

A third, unintended and unavoidable variation is that the data for the replication was collected four years after the data collection of the original study. The difference in time can introduce significant changes in studies in software engineering due to the fast changing pace of the technologies in our field.

Temporal variation (unintended) and the change in the location of the companies (intended) could, together, account for the difference in the use of Agile methods between the projects in the original study and the replication. On the one hand, it is likely that when the original study (Phase II) was performed, Agile methods were not widely spread in the Brazilian software industry. On the other hand, it is possible that the companies in the Porto Digital science park had distinct characteristics compared to companies located elsewhere due to the strong links among the companies and between the companies and the local universities and research institutions. Despite the cause, the fact remains that the two sets of projects were very different regarding the use of Agile methods. The impact of this difference in the results of the replication (discussed in Section 6.1.2) prompted us to consider that the investigation of team building criteria and project success must be performed using different approaches in the context of Agile projects. Therefore, these two variations in the replication were important to stimulate a new research agenda for our study of software teams.

6.2.3. Some considerations about agile projects and teams

Our results seem to indicate two important distinctions between Agile and traditional projects. The first related to the perception of success and the second about team behavior. We present this discussion as a set of propositions for further investigation because we do not have evidence to corroborate our conclusions.

From the results illustrated in Fig. 6b we can see that the success goals associated with satisfaction (in particular G_6 – Project Manager Satisfaction) scored significantly higher than performance goals (in particular C_1 – Cost and C_2 – Time). This trend could indicate that satisfaction goals were being prioritized over the performance goals. Because our sample is clearly dominated by Agile projects, it would be possible to hypothesize that this trend could be a characteristic of the Agile development "culture". However, given the small number of traditional projects, our data neither supports nor refutes this hypothesis. This opens the opportunity for research to investigate if this trend actually happens in industry and, if it does, whether it is actually related to Agile methods and practices, and, finally, what are the long-term impacts of this prioritization on the individuals and organizations developing Agile projects.

As another point for discussion and future investigation, we have showed indications that the correlations between the use of team building criteria and the success goals of the project could be different depending on the use of Agile or Traditional development methods. In fact, Table 20 shows that the correlations between use of the criteria and the success goals are present in the sample of traditional projects but cannot be found in the sample of Agile projects (it is important to remember that the non existence of correlations is explained also by the characteristics of the sample). Looking at this result, one could be tempted to conclude that the use of team building criteria was less relevant for the Agile projects. However, our data does not allow this conclusion, because most of the Agile projects were successful and used the criteria, and we do not have unsuccessful projects to compare with. What the data seems to indicate only is that the poor use of the criteria affects the traditional projects.

However, this results stimulated the formulation of the hypothesis that Agile (self-managing) teams, being more flexible and adaptable than traditionally managed teams, are more capable of compensating for a poor initial team design throughout the development of the project whereas the Traditional teams would struggle more to compensate the poor team design. Therefore, the consistent use of the team building criteria at the beginning of the project would be relevant for all types of projects and would be especially important for the Traditional ones because they are less flexible.

This hypothesis cannot be tested in cross-sectional studies because it requires investigating team behavior and performance along the development of the project. This prompts for the use of longitudinal studies. Following the line of investigation of Robinson, Segal, and Sharp [71], we contend that such studies should be ethnographically informed and combine qualitative and quantitative methods, comparing the behavior of software teams with high and low performance (as performed by Yeatts and Hyten [78]), and using Traditional and Agile methods.

6.2.4. Cultural issues in personnel selection

Aycan and Kanungo [8] argued that "culture influences the process of recruitment and selection in many ways, such as attitudes towards selection and testing, the purposes the staffing serves, and the perceived fairness and appropriateness of criteria and methods that are used in the process". For instance, while in North American countries selection is naturally seen as a process whereby individuals are assessed under certain criteria used as predictors of performance in the job (see Section 2.1.3), in European countries, such as Italy and France, testing is perceived as invasion of privacy and, therefore, receive a negative connotation.

Culture differences may account for distinct uses of selection criteria in different countries and even for the different levels of formalization in the use of the criteria. Considering that Phases I, II and IV were carried out in a single country, cultural issues may have biased the findings. We are not aware of studies about the influences of culture on personnel selection carried out in the Brazilian context and, thus, cannot assess the extent of this potential limitation. Our goal with the mapping study performed in Phase III was to overcome some of these biases by analyzing studies from different countries and in different organizational contexts.

6.3. Implications for industrial practice

Our results indicated that carefully selecting team members for software teams is likely to influence the projects in which these teams participate. Besides, it seems that the type of development method used can moderate (increase or decrease) this influence. The results found in Phase II indicate that companies can create better conditions for software project success by adopting more comprehensive set of team building criteria, and making them explicit and more formal inside the company. We did not found the same correlations in the replication of the study in Phase IV due to the reasons discussed in Section 6.1. However, as explained above, we have enough empirical evidence to believe in the validity of the results of Phase II.

6.3.1. A set of selection criteria, not a team building model

It is important to remind the practitioner that our focus was in the identification of selection criteria used in practice and the study of the relationships between the level of formalization in the use of the criteria and team effectiveness. We did not address how managers used the criteria to create specific team compositions [62]. In practice, a given criteria can be used with different strategies by different managers to create specific team compositions. For instance, personality can be used to create teams with low or high diversity, and a given composition may be better suited for certain tasks/projects than others [14,38,64,77]. Therefore, we are not proposing a team-building model, but our findings form an important step towards such proposition.

Nevertheless, we grouped the team building criteria into four categories to facilitate their use in practice. These categories are related to the level of direct influence managers and team leaders can have when selecting team members and, most importantly, when managing a team. This grouping was based on theoretical studies and corroborated by the empirical data (through principal component analysis), meaning that it is conceptually sound and also reflects the use of the criteria in the industrial practice. In the grouping of the criteria, innate characteristics (personality and behavior) are relatively stable and difficult to change. Managers can use these criteria in team selection and expect them to remain constant throughout the duration of the teamwork. On the other hand, technical aspects can be subject to improvement, e.g., through training. On the organizational level, internal criteria are, up to certain extent, under the control of the organization, while external criteria is directed by broad competitive strategies and are, consequently, more difficult to manage.

6.3.2. Agile and traditional teams

The difference in the results between Agile and traditional projects are also relevant in practice. Although this still requires further studies to produce empirical evidences, it seems that the type of development method could have a moderating effect on the use of the team building criteria. This can be explained by differences on how software teams behave in the Agile and Traditional contexts. Our hypothesis, discussed above, is that the flexible style of management with high levels of autonomy for the team to define its structure and task assignments may compensate for a less consistent use of team building criteria or increase the effectiveness of a carefully built team. Another hypothesis is that Traditional teams would not exhibit the same flexibility and would be less adaptable to compensate for a poorly designed team. In other words, because a traditional team has less flexibility to adjust to individual characteristics, it requires a more careful use of the team building criteria up front.

6.3.3. Multi-dimensional definition of effectiveness

Use a multi-dimensional measure of team effectiveness is important because it provides a more holistic view of effectiveness, as proposed by several theories about teamwork [21,22,47,78]. A carefully chosen set of effectiveness criteria could be an important managerial tool. The six success goals used in our research provide a view of effectiveness that combines performance (cost, time, and scope) and satisfaction (team, client, and project manager). We showed that instruments to assess these success goals can be designed with high reliability and internal consistency, and these instruments could be useful tools in practice.

Finally, human factors must be taken into account when assembling a software team, as previously emphasized in the literature (Section 2). However, in several interviews in Phase I, project managers declared that they had little formal knowledge or training on methods and tools to evaluate individual and social factors. The managers acknowledge the importance of human factors and that they lack of knowledge can jeopardize their ability to act in certain situations that affect project performance. Companies should address this limitation by not only having human resources personnel skilled in dealing with such situations, but also training project managers on some of these human factors skills. If our hypothesis about Agile teams adaptation is true, then in an Agile context the skills on human factors would be required in all or most of the members of the software team.

7. Conclusions

In this article, we reported the results of four studies performed over a five-year period, between 2007 and 2012. Three empirical studies looked at the criteria used by project managers to select software team members and how the use of these criteria related with project success. In these studies, we used data from project managers, teams, and projects from software companies in Brazil. Forty different companies participated in the studies and we analyzed data collected from 88 software projects in total.

We employed a mix-method [35], replication approach [16], based on two iterations each one composed of two studies. In the first iteration, we performed a qualitative study to understand what team building criteria were used by project managers and team leaders in practice. Using coding techniques from qualitative research, we extracted eight criteria from interviews with project managers. This first phase was followed by a cross-sectional survey in which we investigated the correlation between the consistent use of the eight criteria and a multidimensional set of project success goals. The results of this first iteration showed that the consistent use of the team building criteria correlated significantly with project success.

In the first phase of the second iteration, we performed a mapping study to verify and complement the set of team building criteria. From the results of the review, we added two new criteria to the original set. Finally, we replicated the cross-sectional survey with improved instruments and on a different organizational context. The results of the replication partially confirmed the results obtained in the first iteration. Besides, we found that the type of development method (Agile or Traditional) may have a moderator effect on the correlation between the consistent use of team building criteria and the success goals of the projects.

Our results indicated that carefully selecting team member for software teams are likely to influence the projects in which these teams participate. We grouped the team building criteria in four categories to facilitate their use in practice, according to the level of direct influence managers and team leaders can exert on them. This grouping was based on theoretical studies and its structure was corroborated by the empirical data, meaning that it is conceptually sound and also reflects the use of the criteria in the industrial practice.

The knowledge gained from performing a mix-method, replicated study prompted various directions for future work that can be relevant for research and practice, including:

- *Longitudinal study about team behavior*: As mentioned above, some hypothesis about software team behavior cannot be tested adequately with cross-sectional studies. We believe that longitudinal studies based on ethnographic methods would provide adequate research strategy for the investigation of the effects of different development methods on team effectiveness. We expect to use this research strategy to perform a comparative study of the role of team building criteria in the context of Agile and traditional projects.
- *Investigation about team composition and effectiveness*: In this study we did not investigate how the team building criteria were applied. For instance, if personality was assessed, was the personality types used to create variety or homogeneity in the team? In future research, we would like to investigate the internal characteristics of the teams, like personality and technical skills diversity, and relate these characteristics with the use of the team building criteria and with project success.
- *Investigation about the perception of success in Agile projects*: As discussed in Section 6.2.3, our results indicated that there is a potential trend in Agile projects to prioritize satisfaction over performance in project development. The investigation of whether this trend actually exists and is related to an "Agile culture" would have important practical applications.
- *Knowledge, skills, and abilities for teamwork*: The use self-managing team in software development proposed by the Agile methods is challenging. The types of knowledge, skills, and abilities to work in a self-managed team can be quite diverse

then those needed to work under a project manager in a traditional team. Steven and Champion [75] proposed a set of knowledge, skills, and abilities (KSA) necessary for teamwork that explicitly include self-management KSAs. We intend to use this framework in future investigations about self-managing software teams, in particular in the context of Agile development.

- *Study about the relationship of the use of TBC and the maturity level of the company*: It is possible that organizational maturity affects the use of TBC. One hypothesis could be that more mature companies would tend to consistently use more criteria than less mature ones. Future studies could try to investigate the correlations between the use of TBC and maturity levels of software processes or other type of quality and process certifications.
- *Investigations about replication*: In a recently developed systematic mapping study about replication of empirical research in software engineering, we discussed how temporal gap between studies could affect the comparability of results between replications and original studies [28,29]. The replication performed in this case study shows that this temporal gap can cause unexpected variations in replications that could produce unanticipated results. We believe that a deeper understanding of this phenomenon would be important for the empirical research in software engineering. Researchers could use the studies analyzed in the above-cited mapping study to conduct such investigation.

The research presented in this article is an initial attempt to contribute with the problem of software team building by first investigating how project managers select individual members to compose teams in practice and how the selection of team member relate to project success. This is a first necessary step in the study of team composition and its relationship with team processes and, ultimately, with team effectiveness. We understand that the study of team effectiveness is complex, involving many more factors that interact with team composition in complex ways [47,78]. Our long-term research goal is to build and test explanatory theories of software team effectiveness, grounded on empirical data collected through case studies conducted in the practice of software engineering in industry.

It is our contention that reductionist approaches, that tend to isolate a small number of factors to be analyzed in artificial laboratory settings, do not provide adequate research methods for the investigation of this problem in all its complexity. This type of research requires holistic mix-method approaches, based on the long-term involvement and collaboration between industry and academia. It also requires many replicated studies in different contexts and their consistent integration. For this, collaborative research agenda among researchers is needed. We hope this article will stimulate other researchers to join in our effort in understanding the complex phenomena related to the effects of human factors in software engineering.

Acknowledgments

Fabio Q.B. da Silva holds a research grant from the Brazilian National Research Council (CNPq), process #314523/2009-0. This article was written while Prof. Fabio Silva was in a sabbatical leave at the University of Toronto, receiving a CAPES research grant process #6441/10-6. A. César C. França and Cleviton V.F. Monteiro have doctoral scholarships from CNPq, process #141156/2010-4, and CAPES, respectively. We would like to thank Prof. Steve Easterbrook, Elizabeth Patitsas, and Alicia Grubb for discussions, comments, and criticisms that lead to important improvements in the content and structure of this article. We also would like to thank all participants in all phases of this research. Finally, we are grateful for the partial support of the Samsung Institute for Development of Informatics (Samsung/SIDI) for this research.

Appendix A. Phase I: Interview guide

ID	Question	Type [63]
Q1.	Describe, in detail, how you select software engineers to compose a software project team. Probe: after the spontaneous answer, ask explicitly about selection criteria in the case that the interviewee did not mention any. *Could you please be more specific about the selection criteria you use?*	Experience and behavior
Q2.	Regarding the selection criteria mentioned before, could you tell about their level of formalization or documentation?	Knowledge
Q3.	Still regarding the selection criteria, tell me about how you assess the degree to which an individual meets each criterion. Probe: after the spontaneous answer, stimulate the interviewee to provide names of the methods, techniques, or tools deployed in the assessment, including commercial or company made instruments. *Could you please be more specific about the types of tests?* *How were these tests developed? Internally or externally?*	Experience and behavior
Branch	If the interviewee mention any criteria related to personal aspects (personality, behavior, etc.) go to Q4, otherwise go to Q7.	
Q4.	You mentioned selection criteria related to personal aspects (such as ... [as mentioned by the interviewee in Q3]); could you tell me why did you use these criteria?	Experience and behavior
Q5.	How do you observe or assess personal aspects when selecting individual to build software teams?	Experience and behavior
Q6.	Explain how you acquired the skills or knowledge used to assess the personal aspects of prospective team members. Probe: after the spontaneous answer, stimulate the interviewee to provide details about formal training. *Could you please be more specific about any training you received?* *Are they internal or external to the company?* Go to Q10.	Knowledge
Q7.	Do you believe that personal aspects, such as personality or behavior, are important criteria to be considered	Opinion and values

Appendix A. (*continued*)

ID	Question	Type [63]
	when building a software team?	
Q8.	Why, in your opinion, the personal aspects are important/not important?	Opinion and values
Branch	If the answer to Q7 is Yes go to Q9, otherwise go to Q10.	
Q9.	You did not mention any criterion related to personal aspects, although you think they are important. Why don't you use such criteria in selection individuals for your software teams?	Experience and behavior
Q10.	What else would you like to share about your experience in selecting individual and building software teams?	Experience and behavior

Appendix B. Phase III: List of studies

[S012] M. Mcguire, Team software development techniques, in: SIGCPR '86 Proceedings of the Twenty-Second Annual Computer Personnel Research Conference on Computer Personnel Research Conference, 1986.
[S018] J. Duggan, J. Byrne, G.J. Lyons, A Task Allocation Optimizer for Software Construction, vol. 21 (3), IEEE Computer Society Press, May 2004, pp. 76–82.
[S019] J. Srinivasan, K. Lundqvist, Agile in India: challenges and lessons learned, in: ISEC'10 – Proceedings of the 2010 India Software Engineering Conference, 2010, pp. 125–130.
[S026] V. Plekhanova, R. Offen, Managing the human–software environment, in: Proceedings of the 8th International Workshop on Software Technology and Engineering Practice, 1997, pp. 422–432.
[S028] K.M. Lui, K.C.C. Chan, Programming task demands, in: Proceedings of the 5th IEEE International Conference on Cognitive Informatics, ICCI 2006, 2006, pp. 765–770.
[S030] A. Gray et al., Forming Successful eXtreme Programming Teams, in: Proceedings – AGILE Conference, 2006, pp. 390–399.
[S032] C.J. Goebel III, How being agile changed our human resources policies, in: Proceedings – 2009 Agile Conference, 2009, pp. 101–106.
[S035] R. Barrett, Labouring under an illussion? The labour process of software development in the Australian information industry, New Technology, Work and Employment 16 (1) (2001) 18–34.
[S041] R. Feldt et al., Links between the personalities, views and attitudes of software engineers, Information and Software Technology 52 (6) (2010) 611–624.
[S048] N.B. Moe, T. Dingsøyr, T. Dybå, Overcoming Barriers to Self-Management in Software Teams, IEEE Software 26 (6) (2009) 20–26.
[S052] B. Xu, X. Yang, A. Ma, Role based cross-project collaboration in multiple distributed software design projects, in: International Conference on Computer Supported Cooperative Work in Design, 2008, pp. 177–182.
[S053] E. Demirors, G. Sarmagik, O. Demirors, The role of teamwork in software development: microsoft case study, in: Proceedings of the EUROMICRO Conference, 1997, pp. 129–133.
[S061] S. Licorish, A. Philpott, S.G. MacDonell, Supporting agile team composition: a prototype tool for identifying personality (in)compatibilities, in: Workshop on Cooperative and Human Aspects on Software Engineering, 2009, pp. 66–73.
[S064] T.B. Hilburn, Teams need a process!, in: Proceedings of the Conference on Integrating Technology into Computer Science Education, 2000, pp. 53–56.
[S066] Z. Shujuan et al., The balance study of IT project and team member, in: IEEE International Conference on Information Management and Engineering, 2010, pp. 442–445.
[S077] T.A.B. Pereira et al., A recommendation framework for allocating global software teams in software product line projects, in: Proceedings of the 2nd International Workshop on Recommendation Systems for Software Engineering, 2010, pp. 36–40.
[S079] H. Hu, L. Li, B. Xu, A role based human resource organization model in dual-shore software development, in: Conference Proceedings – IEEE International Conference on Systems, Man and Cybernetics, 2008, pp. 3657–3662.
[S081] F.E.D.O.D. Silva, An Approach based on Social Combination to Support Team Building, Master Dissertation in Informatics, Institute of Mathematics, NCE, Federal University of Rio de Janeiro, Rio de Janeiro, 2009, 168p. (in Portuguese).
[S094] M. Rajendran, Analysis of team effectiveness in software development teams working on hardware and software environments using Belbin self-perception inventory, Journal of Management Development 24 (8) (2005) 738–753.

Appendix C. Phase IV: Survey instrument (revised and improved from Phase II)

Part I – General information.

Company Identification	
1. Company Name	
2. Address	
Official Contact Regarding this Research	
1. Name	
2. Phone	
3. E-mail	
Information about the Respondent of this Questionnaire	
1. Name	
2. E-mail	
3. Sex	( ) Female ( ) Male
4. Age	( ) 15–20 ( ) 21–25 ( ) 26–30 ( ) 31–35 ( ) 36–40 ( ) 41–45 ( ) 46–50 ( ) 51–55 ( ) above 56
5. Education	Undergraduate Degree: ( ) Completed ( ) In progress Master Degree: ( ) Completed ( ) In progress Doctoral Degree: ( ) Completed ( ) In progress ( ) Other. Specify: ________________
6. Background	( ) Computer Science ( ) Software Engineering ( ) Systems Engineering ( ) Information Systems ( ) Other. Specify: ________________
7. Years of Professional Experience	

(*continued on next page*)

8. Current Role or Position	( ) Project Manager ( ) Product Owner (Scrum) ( ) Scrum Master (Scrum) ( ) Other. Specify: ____________________
9. Years in the Current Job	

Part II – Project characterization.

Please, answer this part of the questionnaire with information about the Project identified below.	
Project Code	PRJ-nnn (internal code used in this study)
Project Name	<as used in the company>
Start Date	
End Date	
Duration in months	
Team Size	
Team Structure	( ) Fully hierarchical (Project manager and development roles) ( ) Scrum type (Product Owner, Scrum Master, Team) ( ) Fully flat (no predefined roles/no hierarchical structure) ( ) Other. Describe:
Type of Development Method	( ) *Waterfall* ( ) Iterative-incremental (RUP style) ( ) Agile (XP, Scrum, TDD, etc.) ( ) Mix-method. Describe: ( ) Other. Describe: ( ) None.

Part III – Use of team building criteria.

Please, carefully read the following definitions before answering the questions below. If you have any questions or doubts that could affect your answers, please get in touch with the Research team.

Criterion	Meaning
Technical Profile	*Profile* is the appearance, character, or the general characteristics of a person. *Technical* means the specialized area of activity or human knowledge in a particular area. Therefore, *technical profile* is connected directly to the technical capacity of the individual in a particular technology, language, platform, etc. This also includes expert knowledge in a module of a system or business process.
Personality	*Personality* is what distinguishes one individual from other individuals, that is, the character, thoughts, and feelings possessed by a person that "uniquely influences his or her cognitions, motivations, and behaviors in various situations".
Behavior	Behavior is the set of actions and reactions observed in individuals in certain circumstances and in relation to the environment.
Customer Importance	*Customer Importance* refers to the strategic, competitive, or financial importance the customer for which the project will be developed has to the company at the time the team is being assembled.
Productivity	*Productivity* of the software engineer is the ratio between what an individual is capable of producing (in terms of lines of code, test cases tested, etc.) by a given amount of time.
Availability	*Availability* refers to the amount of time an individual is available to work in a new project.
Individual Costs	*Individual Cost* refers to the impact on the project costs exerted by adding that individual to the project team.
Project Importance	*Project Importance* refers to the strategic, competitive, or financial importance a project has to the company at the time the team is being assembled.
Peer Indication	*Peer Indication* is an indication or a referral for individual provided by trustworthy source (a reference provided by a person to whom the project manager trusts).
Task Preference	*Task Preference* is used when the project manager tries to match the personal preferences for tasks or team roles of a given individual to the tasks and roles of the project.

Appendix C. (*continued*)

Value	Meaning
Low Formalization	The criterion was not used to select the team member or it was considered but no assessment of individuals with respect to the criterion was carried out (e.g., productivity was thought to be important in the project but the project manager could not evaluate team members with respect to individual productivity).
Medium Formalization	The criterion was considered but no formal evaluation instrument was used and the assessment of the individuals with respect to the criterion was subjective (e.g., productivity was used as a criterion but the company does not have a record of past productivity and the project manager relied on his subjective assessment of the individuals).
High Formalization	The criterion was considered and formal evaluation instrument was used to provide an objective assessment of the individual with respect to the criteria

Appendix C. (*continued*)

Value	Meaning
	(e.g., productivity was used and the company's records about employee productivity in past projects were used to assess the individuals).

On a scale between 0 and 2, with 0 meaning "Low Formalization", 1 "Medium Formalization" and 2 "High Formalization" (according to the definitions above), please mark an "X" on the item that best represents the use of each criterion on the selection of individuals to compose the software team of the Project characterized in Part II.

0	1	2
▲ Low Formalization	▲ Medium Formalization	▲ High Formalization

	Criterion	0	1	2
C_1	Technical Profile	○	○	○
C_2	Individual Costs	○	○	○
C_3	Productivity	○	○	○
C_4	Availability	○	○	○
C_5	Personality	○	○	○
C_6	Behavior	○	○	○
C_7	Project Importance	○	○	○
C_8	Customer Importance	○	○	○
C_9	Peer Indication	○	○	○
C_{10}	Task Preference	○	○	○

Part IV – Project results.

On a scale between −2 a +2, with −2 meaning "Totally Disagree" and +2 meaning "Totally Agree", please mark only one item that best represents your level of agreement with the following affirmatives with respect the performance of the project characterized in Part II.

−2	−1	0	+1	+2
▲ Totally Agree	▲	▲	▲	▲ Totally Disagree

	Affirmatives	−2	−1	0	+1	+2
(1)	The team members worked effectively together.	○	○	○	○	○
(2)	The project failed in achieving its cost goals, as initially planned.	○	○	○	○	○
(3)	The project produced its results in a timely fashion.	○	○	○	○	○
(4)	My experience with this project was rewarding.	○	○	○	○	○
(5)	The project successfully achieved its scope and quality goals.	○	○	○	○	○

Appendix C. (*continued*)

	Affirmatives	−2	−1	0	+1	+2
(6)	The client/customer constantly complained about the results of this project.	○	○	○	○	○
(7)	I do not have any reason to consider this project a success.	○	○	○	○	○
(8)	The project successfully achieved its costs goals.	○	○	○	○	○
(9)	The results of this project were often delivered late.	○	○	○	○	○
(10)	This was the worst project in which I participated.	○	○	○	○	○
(11)	The client/customer was satisfied with the results of the project.	○	○	○	○	○
(12)	The overall effectiveness of the teamwork was highly unsatisfactory.	○	○	○	○	○
(13)	This project was a success.	○	○	○	○	○
(14)	The project failed in producing the requirements expected by the customer.	○	○	○	○	○

References

[1] S.T. Acuña, N. Juristo, Assigning people to roles in software projects, Software: Practice and Experience 34 (2004) 675–696.

[2] S.T. Acuña, N. Juristo, A.M. Moreno, Emphasizing human capabilities in software development, IEEE Software 23 (2) (2006) 94–101.

[3] S.T. Acuña, M. Gómez, N. Juristo, Towards understanding the relationship between team climate and software quality: a quasi-experimental study, Empirical Software Engineering 13 (2008) 401–434.

[4] S.T. Acuña, M. Gómez, N. Juristo, How do personality, team processes and task characteristics relate to job satisfaction and software quality?, Information and Software Technology 51 (3) (2009) 627–639, http://dxdoi.org/10.1016/j.infsof.2008.08.006.

[5] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR), fourth ed., Amer. Psychiatric Pub., 2000.

[6] N. Anderson, D.S. Ones, H.K. Sinangil, C. Viswesvaran (Eds.), Handbook of Industrial, Work and Organizational Psychology, Personal Psychology, vol. 1, Sage Publications Ltd., Thousand Oaks, CA, 2001.

[7] N. Anderson, D.S. Ones, H.K. Sinangil, C. Viswesvaran (Eds.), Handbook of Industrial, Work and Organizational Psychology, Organizational Psychology, vol. 2, Sage Publications Ltd., Thousand Oaks, CA, 2001.

[8] Z. Aycan, R. Kanungo, Cross-cultural industrial and organizational psychology: A critical appraisal of the field and future directions, in: N. Anderson, D.S. Ones, H.K. Sinangil, C. Viswesvaran (Eds.), Handbook of Industrial Work & Organizational Psychology, vol. 1, SAGE, London, 2001.

[9] F. Baker, Chief programmer team management of production programming, IBM Systems Journal 11 (1) (1972) 56–73.

[10] K. Bartol, D. Martin, G. Matthews, M. Tein, Management: A Pacific Rim Focus, The McGraw Hill Companies, Australia, 1998.

[11] S. Beecham, N. Baddoo, T. Hall, H. Robinson, H. Sharp, Motivation in software engineering: a systematic literature review, Information and Software Technology 50 (9–10) (2008) 860–878.

[12] M.R. Belbin, Management Teams: Why they succeed or Fail?, Butterworth-Heinemann Ltd, 1981. ISBN: 0750626763.

[13] B.W. Boehm, Software Engineering Economics, Prentice-Hall, Inc., Englewood Cliffs, 1981.

[14] J.H. Bradley, F.J. Hebert, The effect of personality type on team performance, Journal of Management Development 16 (5) (1997) 337–353.

[15] H.J. Brightman, Group Problem Solving: An Improved Managerial Approach, Georgia State University Business Publishing Division, Atlanta, 1998.

[16] A. Brooks, M. Roper, M. Wood, J. Daly, J. Miller, Replication's role in software engineering, in: F. Shull, J. Singer, D.I.K. Sjøberg (Eds.), Guide to Advanced Empirical Software Engineering, Springer, 2008, pp. 365–379 (Chapter 14).

[17] F. Brooks, The Mythical Man-Month, Addison-Wesley, 1975. ISBN: 201-00650-2.

[18] M.A. Campion, G.J. Medsker, A.C. Higgs, Relations between work group characteristics and effectiveness: implications for designing effective work

groups, Personnel Psychology 46 (4) (1993) 823–847, http://dx.doi.org/10.1111/j.1744-6570.1993.tb01571.x.
[19] A.V. Carron, L.R. Brawley, Cohesion: conceptual and measurement issues, Small Group Research 31 (2000) 89–106.
[20] J.C. Carver, Towards reporting guidelines for experimental replications: a proposal, in: RESER̂2010: Proceedings of the 1st International Workshop on Replication in Empirical Software Engineering Research, Cape Town, South Africa, 2010.
[21] S.G. Cohen, Designing effective self-managing work teams, in: M. Beyerlein (Ed.), Advances in Interdisciplinary Studies of Work Teams, Series of Self-Managed Work Teams, vol. 1, JAI Press, Greenwich, Connecticut, 1993.
[22] S.G. Cohen, D.E. Bailey, What makes team work: group effectiveness research from the shop floor to the executive suite, Journal of Management 23 (1997) 239–290.
[23] L. Constantine, Peopleware Papers: The Notes on the Human Side of Software, Prentice Hall, 2001.
[24] S.J.O. Cruz, F.Q.B. da Silva, C.V.F. Monteiro, P.C.F. Santos, I.R.M. dos Santos, Personality in software engineering: preliminary findings from a systematic review, in: Proceedings of the 15th Annual Conference on Evaluation and Assessment in Software Engineering (EASE 2011), Durham, 2011. doi:10.1049/ic.2011.0001.
[25] B. Curtis, W.E. Hefley, S.A. Miller, People Capability Model (P-CMM), Version 2.0, Tech. Report CMU/SEI-2001-MM-001, Software Eng. Inst., Carnegie Mellon Univ., 2001.
[26] F.Q.B. da Silva, A.C.F. Cesar, An experimental research on the relationships between preferences for technical activities and behavioural profile in software development, in: Proceedings of the XXIII Brazilian Symposium on Software Engineering, 2009, pp. 126–135. doi:10.1109/SBES.2009.16.
[27] F.Q.B. da Silva, A.C.C. França, T.B. Gouveia, C.V.F. Monteiro, E.S.F. Cardozo, M. Suassuna, An empirical study on the use of team building criteria in software projects, in: Proceeding of the International Symposium on Empirical Software Engineering and Measurement (ESEM), 2011, pp. 58–67. doi:10.1109/ESEM.2011.14.
[28] F.Q.B da Silva, M. Suassuna, R.F. Lopes, T.B. Gouveia, A.C.C. França, J.P.N. de Oliveira, L.M.F. de Oliveira, A.L.M. Santos, Replication of empirical studies in software engineering: preliminary findings from a systematic mapping study, in: Proceedings of the 2nd International Workshop on Replication of Empirical Software Engineering Research (RESER), Banff, AL, 2011.
[29] F.Q.B. da Silva, M. Suassuna, A.C.C. França, A.M. Grubb, T. B. Gouveia, C.V.F. Monteiro, I.E. dos Santos, Replication of empirical studies in software engineering research: a systematic mapping study, Empirical Software Engineering Journal, (2012). http://dx.doi.org/10.1007/s10664-012-9227-7.
[30] R.C.G. de Miranda, A Systematic Review about Software Development Teams: Typology, Characteristics, and Building Criteria, Master dissertation, Centre for Informatics, Federal University of Pernambuco, 2011.
[31] A.P.L.F de Vasconcelos, Cohesion in Software Engineering Teams: A Systematic Mapping Study, Master dissertation, Centre for Informatics, Federal University of Pernambuco, 2012.
[32] T. DeMarco, T. Lister, Peopleware: Productive Projects and Teams, Dorset House Publishing Co., New York, 1999.
[33] S. Drescher, G. Burlingame, A. Fuhriman, Cohesion: an odyssey in empirical understanding, Small Group Research 16 (1985) 3–30.
[34] T. Dyba, T. Dingsoyr, Empirical studies of agile software development: a systematic review, Information and Software Technology 50 (9–10) (2008) 833–859.
[35] S.M. Easterbrook, J. Singer, M. Storey, D. Damian, Selecting empirical methods for software engineering research, in: F. Shull, J. Singer, D.I.K. Sjøberg (Eds.), Guide to Advanced Empirical Software Engineering, Springer, 2008, pp. 285–311. Chap. 11.
[36] A.C.C. Franca, T.B. Gouveia, P.C.F Santos, C.A. Santana, F.Q.B da Silva, Motivation in software engineering: a systematic review update, in: Proceedings of 15th Annual Conference on Evaluation & Assessment in Software Engineering (EASE 2011), Durham, UK, 2011, pp. 154–163.
[37] L.L. Cummings, B.M. Staw, in: P.S. Goodman, E. Ravlin, M. Schminke (Eds.), Understanding groups in organizations, Research in Organizational Behavior 9 (1987) 121–173.
[38] N. Gorla, Y.W. Lam, Who should work with whom? Building effective software project teams, Communications of the ACM 47 (6) (2004) 79–82.
[39] R.L. Gorsuch, Factor Analysis, Erlbaum, Hillsdale, NJ, 1983.
[40] F.J. Gravetter, L.B. Forzano, Research Methods for the Behavioral Sciences, fourth ed., Cengage, Learning, 2011.
[41] N. Gross, W.E. Martin, American Journal of Sociology 57 (6) (1952) 546–564.
[42] P.J. Guinan, J.G. Cooprider, S. Faraj, Enabling software development team performance during requirements definition: a behavioral versus technical approach, Information System Research 9 (2) (1998) 101–125.
[43] R.M. Guion, Assessment, Measurement, and Prediction for Personnel Decision, Lawrence Erlbaum Associates, Mahwah, NJ, 1998.
[44] R.A. Guzzo, M.W. Dickson, Teams in organizations: recent research on performance and effectiveness, Annual Review of Psychology 47 (1996) 307–338. doi:0.1146/annurev.psych.47.1.307.
[45] R.A. Guzzo, E. Salas (Eds.), Team Effectiveness and Decision Making in Organizations, Jossey-Bass, San Francisco, 1995.
[46] J.R. Hackman, Groups that Work, Jossey Bass, San Francisco, 1990.
[47] J.R. Hackman, The design of work teams, in: J.W. Lorsch (Ed.), Handbook of Organizational Behavior, Prentice-Hall, Englewood Cliffs, NJ, 1987, pp. 315–342.
[48] J.R. Hackmam, G.R. Oldham, Work Redesign, Addison-Wesley, Reading, MA, 1980.
[49] N. Haggerty, Understanding the link between IT project manager skills and project success research in progress, in: Proceedings of the 2000 ACM SIGCPR Conference on Computer Personnel Research, Chicago, Illinois, United States, 2000, pp. 192–195.
[50] J. Hallows, Information Systems Project Management, AMACOM, New York, 1998.
[51] J. Ivancevich, M. Matteson, Organisational Behaviour and Management, McGraw Hill, Irwin, New York, 2002.
[52] N. Juristo, S. Vegas, Using differences among replications of software engineering experiments to gain knowledge, in: Proceedings of the ACM/IEEE 3rd International Symposium on Empirical Software Engineering and Measurement (ESEM), IEEE Computer Society, Washington, DC, USA, 2009, pp. 356–366.
[53] J. Katzenbach, D. Smith, The Discipline of Teams: A Mindbook–Workbook for Delivering Small Group Performance, John Wiley & Sons, New York, 2001.
[54] B.A. Kitchenham, S.L. Pfleeger, Personal opinion surveys, in: F. Shull, J. Singer, D.I.K. Sjøberg (Eds.), Guide to Advanced Empirical Software Engineering, Springer, 2008, pp. 63–92 (Chapter 3).
[55] B. Kitchenham, S. Charters, Guidelines for Performing Systematic Literature Reviews in Software Engineering, Technical Report EBSE-2007-01, School of Computer Science and Mathematics, Keele University, 2007.
[56] R. Klimoski, L.N. Zukin, Selection and staffing for team effectiveness, in: E. Sundstrom (Ed.), Supporting Work Team Effectiveness, Jossey-Bass, San Francisco, 1999, pp. 63–91.
[57] F. Lettice, M. McCracken, Team performance management: a review and look forward, Team Performance Management 13 (5/6) (2007) 148–159.
[58] J. Mathieu, M.T. Maynard, T. Rapp, L. Gilson, Team effectiveness 1997–2007: a review of recent advancements and a glimpse into the future, Journal of Management 34 (3) (2008) 410–476, http://dx.doi.org/10.1177/0149206308316061.
[59] S. McConnell, Problem programmers, IEEE Software 15 (2) (1998) 126–128.
[60] S. McDonald, H.M. Edwards, Who should test whom? Examining the use and abuse of personality tests in software engineering, Communications of the ACM 50 (1) (2007) 67–71.
[61] J.E. McGrath, Groups: Interaction and Performance, Prentice-Hall, Englewood Cliffs, New Jersey, 1984.
[62] M.L. Maznevski, Understanding our differences: performance in decision-making groups with diverse members, Human Relations 47 (5) (1994) 531–552.
[63] S.B. Merriam, Qualitative Research: A Guide to Design and Implementation, Jossey-Bass, San Francisco, 2009.
[64] J. Miller, Z. Yin, A cognitive-based mechanism for constructing software inspection teams, IEEE Transactions on Software Engineering 30 (11) (2004) 811–825.
[65] N.B. Moe, T. Dingsoyr, Scrum and team effectiveness: theory and practice, in: Proceeding of the 9th Agile Processes in Software Engineering and Extreme Programming International Conference, Limerick, Springer, 2008, pp. 11–20.
[66] I.B. Myers, The Myers–Briggs Type Indicator, Consulting Psychologists Press, Palo Alto, CA, 1962.
[67] M. Petticrew, H. Roberts, Systematic Reviews in the Social Sciences, Blackwell Publishing, 2006.
[68] A. Peslak, The impact of personality on information technology team projects, in: Proceedings of the ACM SIGMIS CPR Conference on Computer Personnel Research, Forty Four Years of Computer Personnel Research: Achievements, Challenges & the Future, ACM, 2006, pp. 273–279.
[69] M. Rajendran, Analysis of team effectiveness in software development teams working on hardware and software environments using Belbin self-perception inventory, Journal of Management Development 24 (8) (2005) 738–753, http://dx.doi.org/10.1108/02621710510613753.
[70] R. Ryckman, Theories of Personality, Thomson/Wadsworth, Belmont, CA, 2004.
[71] H. Robinson, J. Segal, H. Sharp, Ethnographically-informed empirical studies of software practice, Information and Software Technology 49 (6) (2007) 540–551, http://dx.doi.org/10.1016/j.infsof.2007.02.007.
[72] J.F. Salgado, C. Viswesvaran, D.S. Ones, Predictors used for personnel selection: an overview of constructs, methods, and techniques, in: N. Anderson, D.S. Ones, H.K. Sinangil, C. Viswesvaran (Eds.), Handbook of Industrial Work & Organizational Psychology, vol. 1, SAGE, London, 2001.
[73] C.B. Seaman, Qualitative methods, in: F. Shull, J. Singer, D.I.K. Sjøberg (Eds.), Guide to Advanced Empirical Software Engineering, Springer, 2008 (Chapter 2).
[74] B. Shneiderman, Software Psychology: Human Factors in Computer and Information Systems, Winthrop Publishers, 1980.
[75] M.J. Stevens, M.A. Campion, Knowledge, skills, and abilities for teamwork: implications for human resources management, Journal of Management 20 (2) (1994) 503–530.
[76] G.M. Weinberg, The Psychology of Computer Programming, silver anniversary ed, first ed., Dorset House Publishing, 1971.
[77] K. White, A preliminary investigation of information systems team structures, Information & Management 7 (1984) 331–335.
[78] D.E. Yeatts, C. Hyten, High-Performing Self-Managed Work Teams, Sage Publications Inc., 1998.
[79] R.K. Yin, Case study Research: Design and Methods, third ed., Sage, London, 2003.