KNOWLEDGE CLASSIFICATION FOR SUPPORTING EFFORT ESTIMATION IN GLOBAL SOFTWARE ENGINEERING PROJECTS

Ricardo Britto

Blekinge Institute of Technology
Licentiate Dissertation Series No. 2015:04
Department of Software Engineering
Knowledge Classification for Supporting Effort Estimation in Global Software Engineering Projects

Ricardo Britto
Knowledge Classification for Supporting Effort Estimation in Global Software Engineering Projects

Ricardo Britto

Licentiate Dissertation in Software Engineering

Department of Software Engineering
Blekinge Institute of Technology
SWEDEN
To my wife, daughter, parents and siblings.

“I fear not the man who has practiced 10,000 kicks once, but I fear the man who has practiced one kick 10,000 times.”

–Bruce Lee
ABSTRACT

**Background:** Global Software Engineering (GSE) has become a widely applied operational model for the development of software systems; it can increase profits and decrease time-to-market. However, there are many challenges associated with the development of software in a globally distributed fashion. There is evidence that these challenges affect many processes related to software development, such as effort estimation. To the best of our knowledge, there are no empirical studies to gather evidence on effort estimation in the GSE context. In addition, there is no common terminology for classifying GSE scenarios focusing on effort estimation.

**Objective:** The main objective of this thesis is to support effort estimation in the GSE context by classifying the existing knowledge in this field.

**Method:** In this thesis, we employed systematic literature review, survey, systematic mapping and literature survey as research methods.

**Results:** Our research on the state of the art and the state of the practice on effort estimation in the GSE context shows that the effort estimation techniques employed in GSE projects do not differ from the techniques employed in collocated projects. It was also identified that global aspects, e.g. time, geographical and socio-cultural distances, are accounted for as cost drivers, although it is not clear how they are measured. Our systematic mapping study on software engineering taxonomies showed that most taxonomies are designed in an ad-hoc way; hence, an existing method was revised to support the development of SE taxonomies in a more systematic way. The aforementioned results were combined to specialize an existing GSE taxonomy, with focus on effort estimation. The usage of the specialized taxonomy was illustrated by classifying 8 completed GSE projects. The results show that the specialized taxonomy is comprehensive enough to classify effort estimation aspects of GSE projects.

**Conclusions:** We believe that the taxonomy presented in this thesis can help researchers and practitioners to report new research on effort estimation in the GSE context; researchers and practitioners will be able to gather evidence, compare new studies and find new gaps in an easier way. Finally, further research should be conducted to complement the findings of this thesis, e.g. both the way cost drivers impact the effort of GSE projects and the way they are measured are not fully clear.
ACKNOWLEDGMENTS

First, I would like to express my gratitude to my supervisors Prof. Jürgen Börstler and Prof. Darja Šmite and also to Prof. Emilia Mendes and Prof. Claes Wohlin for their support, guidance and fruitful collaboration that were mandatory for the success of this thesis.

I also would like to thank all my friends and colleagues at SERL Sweden, who also contributed to this thesis by providing a pleasant work environment. In particular, I would like to express my gratitude to Muhammad Usman for his collaboration on papers and for all the great times after work.

Last but not least, I am sincerely grateful to my wife Annyele Moura, who is always there, specially on the hardest moments. I also would like to thank my parents José Antônio and Teresa, my siblings Juliana, Hugo Leonardo and David and all my friends from Brazil, who from even very far have kept giving me encouragement to achieve my goals.

This work was funded jointly by Federal University of Piauí (UFPI), National Council for Scientific and Technological Development (CNPq), National Institute of Science and Technology for Software Engineering (INES) and Blekinge Institute of Technology (BTH).

Karlskrona, October 5, 2015
PUBLICATIONS

This compilation thesis includes the following 4 papers:


**CONTRIBUTION STATEMENT**

Ricardo Britto was the lead author of S1, S2 and S4. The responsibility as the lead author was to design, conduct and write each study. The co-authors provided support by giving feedback on key steps of each investigation, specially regarding the data collection/analysis and writing of the manuscripts.

S3 was led by Ricardo Britto and Muhammad Usman, i.e. both authors equally shared all the responsibilities during the design, conduct and writing of such study. Like for S1, S2 and S4, the other co-authors provided feedback during the data collection/analysis and writing of the manuscripts.
RELATED PAPERS NOT INCLUDED IN THIS THESIS


OTHER PAPERS NOT INCLUDED IN THIS THESIS


## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>vii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>ix</td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1 Problem outline</td>
<td>1</td>
</tr>
<tr>
<td>2 Research context</td>
<td>2</td>
</tr>
<tr>
<td>3 Background</td>
<td>3</td>
</tr>
<tr>
<td>3.1 GSE challenges</td>
<td>3</td>
</tr>
<tr>
<td>3.2 Sourcing strategies</td>
<td>4</td>
</tr>
<tr>
<td>3.3 Software effort estimation</td>
<td>6</td>
</tr>
<tr>
<td>4 Research gaps</td>
<td>9</td>
</tr>
<tr>
<td>5 Research design and methodology</td>
<td>10</td>
</tr>
<tr>
<td>5.1 Research questions</td>
<td>11</td>
</tr>
<tr>
<td>5.2 Research design</td>
<td>11</td>
</tr>
<tr>
<td>5.3 Research methods</td>
<td>11</td>
</tr>
<tr>
<td>6 Summary of the studies</td>
<td>13</td>
</tr>
<tr>
<td>6.1 S1 - Effort Estimation in Global Software Development: A Systematic Literature Review</td>
<td>13</td>
</tr>
<tr>
<td>6.2 S2 - An Empirical Investigation on Effort Estimation in Agile Global Software Development</td>
<td>14</td>
</tr>
<tr>
<td>6.3 S3 - Taxonomies in Software Engineering: A Systematic Mapping Study</td>
<td>14</td>
</tr>
<tr>
<td>6.4 S4 - Extending and Specializing a Global Software Engineering Taxonomy</td>
<td>15</td>
</tr>
<tr>
<td>7 Contributions</td>
<td>15</td>
</tr>
<tr>
<td>8 Limitations</td>
<td>16</td>
</tr>
<tr>
<td>9 Conclusions and future work</td>
<td>17</td>
</tr>
<tr>
<td>2 EFFORT ESTIMATION IN GLOBAL SOFTWARE DEVELOPMENT: A SYSTEMATIC LITERATURE REVIEW</td>
<td>19</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>19</td>
</tr>
<tr>
<td>2 Systematic literature review</td>
<td>20</td>
</tr>
<tr>
<td>2.1 Research questions</td>
<td>20</td>
</tr>
<tr>
<td>2.2 Search strategy</td>
<td>21</td>
</tr>
<tr>
<td>2.3 Search process</td>
<td>22</td>
</tr>
<tr>
<td>2.4 Study selection</td>
<td>23</td>
</tr>
<tr>
<td>2.5 Study quality assessment</td>
<td>25</td>
</tr>
<tr>
<td>2.6 Data extraction</td>
<td>27</td>
</tr>
<tr>
<td>3 Results</td>
<td>27</td>
</tr>
<tr>
<td>3.1 Question 1</td>
<td>27</td>
</tr>
<tr>
<td>3.2</td>
<td>Question 2</td>
</tr>
<tr>
<td>3.3</td>
<td>Question 3</td>
</tr>
<tr>
<td>3.4</td>
<td>Question 4</td>
</tr>
<tr>
<td>3.5</td>
<td>Question 5</td>
</tr>
<tr>
<td>4</td>
<td>Discussion</td>
</tr>
<tr>
<td>5</td>
<td>Threats to validity</td>
</tr>
<tr>
<td>6</td>
<td>Conclusions</td>
</tr>
</tbody>
</table>

3 AN EMPIRICAL INVESTIGATION ON EFFORT ESTIMATION IN AGILE GLOBAL SOFTWARE DEVELOPMENT

1 Introduction | 41 |
2 Related work | 42 |
3 Research Methodology | 43 |
3.1 Research Questions | 43 |
3.2 Survey Design | 43 |
3.3 Survey Execution | 45 |

4 Results and Analysis | 45 |
4.1 Demographics Questions | 47 |
4.2 AGSD Questions | 49 |
4.3 Research Question 1 | 50 |
4.4 Research Question 2 | 51 |
4.5 Research Question 3 | 52 |
4.6 Research Question 4 | 52 |

5 Discussion | 54 |
5.1 RQ1 - Which techniques have been used to estimate effort in AGSD? | 54 |
5.2 RQ2 - What is the impact of the applied sourcing strategy on the selection of an effort estimation method in AGSD? | 54 |
5.3 RQ3 - Which effort predictors have been used to estimate effort in AGSD? | 55 |
5.4 RQ4 - What are the challenges that impact the accuracy of the effort estimates in AGSD? | 55 |

6 Threats to validity | 56 |
7 Conclusions | 56 |

4 TAXONOMIES IN SOFTWARE ENGINEERING: A SYSTEMATIC MAPPING STUDY

1 Introduction | 59 |
2 Background | 60 |
2.1 Taxonomy definition and purpose | 60 |
2.2 Subject matter | 61 |
2.3 Descriptive bases / terminology | 61 |
2.4 Classification procedure | 61 |
2.5 Classification structure | 61 |
LIST OF FIGURES

Figure 1.1 Thesis overview. .................................................. 2
Figure 1.2 The GSE taxonomy (adapted from Smite et al. [149]). 7
Figure 1.3 Expert-based effort estimation process (adapted from Mendes [91]). 8
Figure 1.4 Algorithm-based effort estimation process (adapted from Mendes [91]). 9
Figure 4.1 Employed systematic mapping process. .................. 64
Figure 4.2 Search process. .................................................. 65
Figure 4.3 Selection process. .............................................. 67
Figure 4.4 Extraction process. ............................................. 69
Figure 4.5 Analysis process. ............................................... 73
Figure 4.6 Year and venue wise distributions. ......................... 74
Figure 4.7 Yearly distribution of primary studies by knowledge areas. Horizontal axes represent the years (starting 1987), while vertical axes denote the number of taxonomies. ................................................. 76
Figure 4.8 Systematic map – knowledge area vs research type. . 77
Figure 4.9 Systematic map – knowledge area vs presentation type. ................................................................. 79
Figure 4.10 Distribution of average number of citations per year. The x-axis shows ranges of average citations per year. The y-axis shows the number of primary studies. 86
Figure 5.1 The GSE taxonomy (Adapted from Smite et al. [149]). 98
Figure 5.2 Process for extending and specializing the GSE taxonomy. ................................................................. 103
Figure 5.3 Extended GSE taxonomy. .................................... 106
Figure 5.4 Relationship between PDI and UAI dimensions (Adapted from Hofstede et al. [62]). 112
Figure 5.5 Specialized GSE effort estimation taxonomy. ........ 118
Figure 5.6 Client-server model (Adapted from Coulouris et al. [31]). ................................................................. 121
Figure 5.7 Peer-to-peer model (Adapted from Coulouris et al. [31]). ................................................................. 122
Figure 5.8 Centralized model. ............................................. 123
Figure 5.9 Distributed model. ............................................ 123
Figure 5.10 Semi-distributed model. ................................... 124
Figure 5.11 Classification process. ....................................... 125
Figure 5.12 The collaboration diagram notation. .................... 126
| Figure 5.13 | Collaboration diagram of project 1 | 128 |
| Figure 5.14 | Collaboration diagram of project 2 | 130 |
| Figure 5.15 | Collaboration diagram of project 3 | 134 |
| Figure 5.16 | Collaboration diagram of project 4 | 135 |
| Figure 5.17 | Collaboration diagram of project 5 | 136 |
| Figure 5.18 | Collaboration diagram of project 6 | 138 |
| Figure 5.19 | Collaboration diagram of project 7 | 139 |
| Figure 5.20 | Collaboration diagram of project 8 | 140 |
LIST OF TABLES

Table 1.1  Global distances impact on organizational unit processes (adapted from Ågerfalk et al. [4]) .......................... 5
Table 1.2  Studies’ Id, design type, methods, research questions and chapters. ......................................................... 10
Table 2.1  Summary of search results ........................................... 23
Table 2.2  Final scores for the evaluated works .............................. 26
Table 2.3  Included and excluded works after the study selection and quality assessment ................................. 26
Table 2.4  Identified effort estimation methods ............................ 28
Table 2.5  Identified accuracy metrics for the observed effort estimation methods .................................................. 28
Table 2.6  Identified accuracy levels for the observed effort estimation methods .................................................. 30
Table 2.7  Identified cost drivers .................................................. 31
Table 2.8  Identified size metrics .................................................. 32
Table 2.9  Identified dataset domains ........................................... 32
Table 2.10  Identified dataset types ............................................. 33
Table 2.11  Identified application types ....................................... 33
Table 2.12  Identified sourcing strategies .................................... 34
Table 2.13  Identified number of involved countries ....................... 34
Table 2.14  Identified countries .................................................. 35
Table 2.15  Overall number of sites ............................................. 35
Table 3.1  Survey Categories ..................................................... 46
Table 3.2  Countries in which the respondents work most of the time ................................................................. 48
Table 3.3  Percentages of the selected effort estimation techniques considering the way they were reported .... 51
Table 3.4  Selected technique by sourcing strategy ........................ 52
Table 3.5  Distribution of the responses by type of error ............... 53
Table 4.1  Summary of search results ......................................... 66
Table 4.2  Rationale for excluded studies .................................... 68
Table 4.3  Data extraction form .................................................. 71
Table 4.4  Publication venues with more than two taxonomy papers. ................................................................. 75
Table 4.5  Taxonomy purpose. .................................................... 81
Table 4.6  Validation methods. .................................................... 82
Table 4.7  Classification structure. .............................................. 83
Table 4.8  Descriptive bases. ....................................................... 83
Table 4.9  Classification procedure types. .................. 84
Table 4.10 Classification procedure descriptions. .......... 84
Table 4.11 Number of primary studies (Frequency) by number of citations. ................................. 85
Table 4.12 Number of citations (mean and median) for primary papers by knowledge area. ...................... 87
Table 4.13 The revised taxonomy design method - part 1. ... 88
Table 4.14 The revised taxonomy design method - part 2. ... 89
Table 5.1  PDI and UAI scores and PD and UA classifications. 113
Table 5.2  Language scores and distances ................... 116
INTRODUCTION

1 PROBLEM OUTLINE

Global Software Engineering (GSE) [149] has become a prominent operational model for the development of software systems. Year after year more companies employ GSE in search of increasing profits and decreasing project cycle-times [114, 29, 18, 58]. However, developing software in a globally distributed manner also implies in challenges [58]. Geographical, temporal and socio-cultural distances impact the communication and coordination between the stakeholders, which can lead to longer times to carry out software development related tasks [59] and higher number of software defects [41]. Another problem is the fact that GSE projects are frequently delayed [57, 145]; the actual effort to conduct a GSE project is often bigger than the effort estimated in the beginning of the project. There are many reasons that lead to delays in GSE projects, such as:

- Distributed tasks involve more people than collocated tasks [57].
- There are many risk factors in GSE projects that are frequently overlooked, e.g. organizational differences and terminology differences [145].
- The communication overhead in GSE projects is frequently underestimated [23].
- GSE is not an operational model that fits all kinds of projects. Evidence suggests that projects that do not demand great expertise and are not domain specific are more appropriate to be carried out in a globally distributed fashion [129].

The considerable number of delayed projects reported in the literature indicates that practitioners have fallen short of providing accurate and reliable effort estimates in both collocated and globally distributed projects. In addition, effort estimation\(^1\) seems to be more challenging in globally distributed projects [113], since uncertainty is exacerbated in GSE projects by the aforementioned issues.

\(^1\) i.e. the process used to predict the effort needed to fulfill a given task in a software project
The main goal of this thesis is to support effort estimation in GSE projects through knowledge classification. This thesis summarizes about 2 years of research and shows the progression from identifying and analyzing the states of the art and practice to designing a classification scheme and terminology for classifying GSE projects, with focus on effort estimation.

From 2013 to 2015 4 studies have been conducted, as follows:

- The first study was conducted to identify and analyze the existing evidence on effort estimation in the GSE context (S1).
- The second study was conducted to identify and analyze the state of the practice on effort estimation in the GSE context (S2).
- The third study was conducted to identify and analyze the state of the art of knowledge classification in Software Engineering (S3).
- Finally, the fourth study was conducted to design a taxonomy to classify GSE projects, with focus on effort estimation (S4).

Figure 1.1 shows an overview of the work presented in this thesis, specifically the relationship between the identified gaps, the designed research questions, the employed research methods and the main contributions associated with this thesis.

Figure 1.1: Thesis overview.
The remainder of this chapter is organized as follows: Section 3 presents some background on GSE and effort estimation. Section 4 presents the research gaps identified and addressed herein. Section 5 presents the research design and methodology employed in this thesis. Section 6 presents a summary for each study included herein. Section 7 outlines the main contributions of this thesis. Section 8 presents some limitations associated with the research reported herein. Finally, Section 9 presents conclusions and future directions.

3 BACKGROUND

This section provides a brief description of aspects related to GSE and software effort estimation.

3.1 GSE challenges

The main motivation of organizations for employing GSE is to gain or maintain a competitive advantage in terms of cost, quality, flexibility, productivity and risk dilution in software development [124]. However, GSE has been associated with challenges that are mainly related to the following three organizational processes [25]:

- **Coordination** - It “is the act of integrating each task with each organizational unit, so the unit contributes to the overall objective” (ibid).
- **Control** - It is “the process of adhering to goals, policies, standards or quality levels” (ibid).
- **Communication** - It is “a mediating factor affecting both coordination and control. It is the exchange of complete and unambiguous information” (ibid).

These processes are specially affected by three aspects, namely temporal, geographical and socio-cultural distances [4].

**Temporal distance** measures the time difference between the actors of two different organizational units. Temporal distance is specially caused by time zone differences.

**Geographical distance** measures the effort to enable the actors of two different organizational units to interact personally (on site). It is better measured by considering the effort to travel between organizational units than by considering the distance in kilometers that separates the organizational units.

**Socio-cultural distance** measures the effort to enable the actors of two different organizational units to gain mutual understanding regarding their respective values and practices. It encompasses organizational and
national culture, as well as language, politics, individual motivations and work ethics.

It was identified that such aspects, also known as global distances, have a negative impact on GSE projects’ performance [59, 112] and quality [41, 112]. Ågerfalk et al. [4] have identified that temporal, geographical and socio-cultural distances challenge the coordination, control and communication processes in the GSE context in many different ways, as displayed in Table 1.1.

3.2 Sourcing strategies

The aforementioned challenges are intrinsically related to the way organizations source the software development work. Recently, Smite et al. have proposed a taxonomy of sourcing strategies [149] (see Figure 1.2). Their taxonomy provides a common terminology and allows for the classification of GSE projects with focus on the sourcing strategy aspects. The taxonomy dimensions are defined as follows:

- **GSE** - This dimension contains the root of the taxonomy, called sourcing. In this context, sourcing means some form of external software development.
- **Location** - A sourcing can be delegated to a site in the same country, i.e. onshore, or to a site in another country, i.e. offshore.
- **Legal entity** - Independently from the location, a sourcing can be transferred to a different branch (site) of the requester company, i.e. insourcing, or subcontracted from a different legal entity (company), i.e. outsourcing.
- **Geographical distance** - There are four different kinds of geographical distances, which depend on the location dimension:
  - **Close** - In onshore projects, the geographical distance is considered close when it is possible to have relatively frequent face-to-face meetings, since no flights are required to go from one site to the other.
  - **Distant** - In onshore projects, the geographical distance is considered distant when at least one flight is required to have face-to-face meetings, which yields time and cost increases.
  - **Near** - In offshore projects, the geographical distance is considered near when the required flying time is less than two hours; i.e. it is possible to hold a meeting of three or four hours and travel back and forth within the same day.
  - **Far** - In offshore projects, the geographical distance is considered far when the flying time is longer than two hours and staying overnight is usually required.
Table 1.1: Global distances impact on organizational unit processes (adapted from Ågerfalk et al. [4])

<table>
<thead>
<tr>
<th></th>
<th>Temporal Distance</th>
<th>Geographical distance</th>
<th>Socio-cultural distance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Coordination</strong></td>
<td>Coordination costs typically increase with distance.</td>
<td>Reduced informal contact can lead to reduced trust and a lack of critical task awareness.</td>
<td>Inconsistency in work practices can impinge on effective coordination, as can reduced cooperation through misunderstandings.</td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td>Management of project artifacts may be subject to delays.</td>
<td>Difficult to convey vision and strategy.</td>
<td>Perceived threat from training low-cost “rivals”. Different perceptions of authority/hierarchy can undermine morale. Managers must adapt to local regulations.</td>
</tr>
<tr>
<td><strong>Communication</strong></td>
<td>Reduced opportunities for synchronous communication, introducing delayed feedback.</td>
<td>Increased cost and logistics of holding face-to-face meetings</td>
<td>Potential for stimulating innovation and sharing best practice, but also for misunderstandings.</td>
</tr>
</tbody>
</table>
• **Temporal distance** - There are four different types of temporal distance, which depend on the location dimension:
  
  – **Similar** - In onshore projects, the temporal distance is considered similar when there is a time difference of one hour or less. It allows an almost complete overlap in work hours between two different sites.
  
  – **Different** - In onshore projects, the temporal distance is considered different when the time difference between two sites is longer than one hour.
  
  – **Small** - In offshore projects, the temporal distance is considered small when there is a time distance between sites of four hours or less. In that situation there is overlap of at least half of a workday between two sites.
  
  – **Large** - In offshore projects, the temporal distance is considered large when there is a time distance between two sites of more than four hours.

Smite et al.’s taxonomy has a facet-based classification structure [79] and its aforementioned dimensions relate to each other as follows:

- The dimension “GSE” is parent of all other dimensions.
- The classification by means of the dimension “geographic distance” depends on the categories of the dimension “location”.
- The classification by means of the dimension “temporal distance” depends on the categories of the dimension “geographic distance”.

3.3 **Software effort estimation**

Software effort estimation is the process used to predict the effort (time) required to develop a software system [91]. A generic effort estimation process encompasses four different activities, as follows [91]:

- **Activity 1** - Collect data and/or knowledge from similar completed projects.
- **Activity 2** - Build an estimation model based on the collected data/knowledge.
- **Activity 3** - Estimate the size and determine the cost drivers’ values of the new software project.
- **Activity 4** - Estimate the effort using the estimated size and determined cost drivers’ values as inputs.

Many different effort estimation techniques have been proposed over the years, e.g. algorithmic and parametric models, expert judgment, formal and informal reasoning by analogy, hybrid approaches and artificial
Figure 1.2: The GSE taxonomy (adapted from Smite et al. [149]).
intelligence based effort estimation [20]. The sequence in which the aforementioned effort estimation process activities are executed depends on the type of the effort estimation technique. For example, the effort estimation process employed in expert-based effort estimation is different from the one employed in algorithmic-based effort estimation.

**Expert-based** effort estimation techniques, such as planning poker [28], yield subjective effort estimates, which are based on historical data or expert knowledge from similar past projects. The effort estimates’ accuracy depends on the experience of the people involved [98, 91].

An expert-based effort estimation process is described in Figure 1.3. First, an expert (or group of experts) looks at the estimated size and cost drivers of the new project (equivalent to activity 3 of the aforementioned generic process). Second, they recall or retrieve historical data or knowledge from similar finished projects (equivalent to activity 1 of the aforementioned generic process). Finally, an effort estimate is “calculated” (equivalent to activity 4 of the aforementioned generic process). Note that activity 2 of the aforementioned generic process is not explicitly executed by expert-based effort estimation techniques, i.e. it produces subjective effort estimates [28, 91].

![Figure 1.3: Expert-based effort estimation process (adapted from Mendes [91]).](image)

**Algorithmic-based** effort estimation techniques, such as COCOMO [17], rely on models that are built based on historical data or expert knowledge from similar finished projects. These models represent, in a precise fashion, the relationship between predictors (size metrics and cost drivers) and the effort required to perform a given project [91].

The effort estimation process based on this category of techniques can be described by the diagram in Figure 1.4. First, historical data from similar finished projects that have real effort documented are retrieved (equivalent to activity 1 of the aforementioned generic process). Second, an algorithmic
model is built using the retrieved historical data (equivalent to activity 2 of the aforementioned generic process). Third, values for the estimated size and cost drivers related to the new project are obtained (equivalent to activity 3 of the aforementioned generic process). Finally, an effort estimate is calculated by using the values obtained in the previous step as inputs to the algorithmic model (equivalent to activity 4 of the aforementioned generic process).

Figure 1.4: Algorithm-based effort estimation process (adapted from Mendes [91]).

4 RESEARCH GAPS

Many practices have been employed over the years by organizations to deal with the GSE challenges [48, 36, 125, 121, 32]. However, Schneider et al. [121] identified that the research in GSE is unbalanced, i.e. there is a lot of literature (solutions) related to some topics (e.g. communication, coordination and distributed project management), while other topics are poorly covered by existing literature, such as effort estimation. In the case of effort estimation studies, their findings are also reported in an ad-hoc way, which hinders the comparison and aggregation of existing knowledge in the field. Smite et al.’s taxonomy has many dimensions that can help researchers and practitioners to report GSE-related research. However, it lacks some dimensions related to the effort estimation process, e.g. how the effort estimation process is divided between sites.

Furthermore, despite the fact that effort estimation is not one of the most frequent topics in the GSE researchers’ agenda, the remarkable number of delayed GSE projects [57, 145] indicates that there is need to investigate further how GSE specific aspects impact the reliability and accuracy of the effort estimates [113].
Thus, to provide a better understanding about effort estimation in the GSE context, the following research gaps are addressed in this thesis:

1. Absence of a systematic study on the state of the art of effort estimation in the GSE context.
2. Absence of a systematic study on the state of the practice of effort estimation in the GSE context.
3. Absence of a classification scheme and terminology to allow researchers for reporting GSE-related effort estimation studies in a standardized fashion.

5 RESEARCH DESIGN AND METHODOLOGY

Table 1.2 shows for each study included in this thesis i) an Id, ii) the research design type, iii) the methods employed, iv) the associated research questions and v) the chapter of this thesis where the study is detailed.

In Subsection 5.1 we present the research question that were addressed in this thesis. Subsection 5.2 elaborates more on the research design type employed in each study. Finally, in Subsection 5.3 the applied research methods are discussed.

Table 1.2: Studies’ Id, design type, methods, research questions and chapters.

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Design type</th>
<th>Method</th>
<th>Research question</th>
<th>Chapter</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>Flexible</td>
<td>Systematic literature review</td>
<td>RQ1 (RQ1a)</td>
<td>Chapter 2</td>
</tr>
<tr>
<td>S2</td>
<td>Flexible</td>
<td>Survey</td>
<td>RQ1 (RQ1b)</td>
<td>Chapter 3</td>
</tr>
<tr>
<td>S3</td>
<td>Flexible</td>
<td>Systematic mapping study</td>
<td>RQ2 (RQ2a)</td>
<td>Chapter 4</td>
</tr>
<tr>
<td>S4</td>
<td>Flexible</td>
<td>Literature survey, taxonomy</td>
<td>RQ2 (RQ2b)</td>
<td>Chapter 5</td>
</tr>
</tbody>
</table>

2 Peixoto et al. conducted before a study to fill this gap [105], but they surveyed just one company.
5.1 Research questions

The following research questions were answered in this thesis:

- **RQ1** - How differs effort estimation in the GSE context from effort estimation in the collocated context?
  - **RQ1a** - What is the state of the art of effort estimation in the GSE context?
  - **RQ1b** - What is the state of the practice of effort estimation in the GSE context?

- **RQ2** - How can the body of knowledge on effort estimation in GSE be organized?
  - **RQ2a** - How is knowledge classified in SE?
  - **RQ2b** - How is it possible to classify the existing GSE knowledge towards supporting the effort estimation process in the GSE context?

5.2 Research design

A research design can be fixed or flexible [116]. In fixed design research, one specifies the design early in the research process. In flexible design research, the design can evolve during the research. In both cases, it is possible to use qualitative and quantitative data.

Kampenes et al. [71] suggest that one must select between fixed or flexible research design accounting for the maturity, the purpose, the setting and the time schedule of the research. In this thesis, all the studies were conducted using a flexible design, due to the uncertainty in their planning phase; they were exploratory in nature. For this reason, it was required a degree of flexibility in their design.

5.3 Research methods

Four research methods were used to carry out the research reported in this thesis: Systematic literature review (SLR), survey, systematic mapping (SMS) and literature survey.

In S1, it was required to identify and analyze the state of the art of effort estimation in the GSE context. Hence, a SLR was selected as the method to conduct S1, because this is the most adequate method when the scope of the study is narrow and there is need to evaluate the strength of the existing evidence [74] (which was the case of S1). The conducted review was based on the guidelines by Kitchenham and Charters [74].
SLR is a research method that allows for systematically identifying, evaluating and interpreting the relevant research on a given research topic [74]. This method encompasses three main phases [74]:

- **Planning** - In this phase, the review protocol is designed. The review protocol contains the research questions, the search strategy to collect relevant studies, the criteria to select primary studies and the criteria to evaluate the quality of selected primary studies.

- **Conducting** - In this phase, related studies are identified, selected, evaluated and synthesized based on the review protocol.

- **Reporting** - The last phase of the SLR method consists on communicating the achieved results by means of a written document.

In S2, it was required to complement the evidence gathered in S1 with the state of the practice on effort estimation in the GSE context. Thus, a survey was selected as the method to conduct S2, because this is the most adequate method to collect qualitative data from a wide population of subjects [44] (which was the case of S2). The survey conducted was based on the guidelines by Fink [44].

Survey is a retrospective form of investigation that gathers data from a wide population [44]. To apply such research method, it is mandatory to define its purpose, the analysis and sample units to be used, along with a representative sample of the aimed population and the questions to guide the investigation; all these aspects were defined herein based on the findings of S1. It is critical to define questions that are unambiguous and enable for the collection of accurate and consistent information [44]. The data collection is often conducted using questionnaires or interviews.

In S3, we decided to identify and analyze how Software Engineering knowledge is classified. Hence, a SMS was selected as method to conduct S3, because this is the most adequate method to get an overview about the literature in a field when the scope is very wide and it is not possible to perform a more complex data analysis [74] (which was the case of S3). This study was based on the guidelines by Kitchenham and Charters [74] and partly employed the mapping process proposed by Petersen et al. [106].

SMS is a research method that enables the identification and quantification of the evidence of wide and poorly defined research fields [74]. The outcomes of mapping studies can be used as input to other research methods, e.g. SLR.

The SMS method encompasses the same phases of the SLR method. However, the main difference lies on the fact that the scope of mapping studies is broader, which leads to broader research questions and less focused search strings.

In S4 the findings of studies S1, S2 and S3 were combined in the following way:
• The findings of S1 and S2 were used as motivation and input to design the classification scheme and terminology presented in S4.
• The revised method presented in S3 was used to design the dimensions of the classification scheme presented in S4.

In addition, a literature survey was conducted to complement the findings of S1, S2 and S3. The main difference between literature survey and SLR and SMS is the absence of a review protocol when using literature survey. The review protocol formalizes all the steps to be carried out during the conduction of the research.

In S4, we fetched additional literature about some of the factors included into the classification scheme presented in such a study, i.e. cultural and language differences, software process models and effort estimation process architectural model (see Subsection 6.4 of this chapter and Chapter 5 for more details). Therefore, a formal method (e.g. SLR and SMS) to identify additional literature was not considered necessary.

6 SUMMARY OF THE STUDIES

In this section, we summarize the studies included in this thesis. Note that the employed terminology evolved during the conduction of this work. For this reason, in the initial studies the “Global Software Development” was used to refer to the area rather than “Global Software Engineering”.

6.1 S1 - Effort Estimation in Global Software Development: A Systematic Literature Review

S1 [21] presents the results of a SLR on effort estimation in the context of GSE, which aimed at helping both researchers and practitioners to gain a holistic view over the current state of the art regarding effort estimation in the context of GSE.

Only few studies (5) were found with empirical evidence on effort estimation in the GSE context. These 5 studies report the usage of effort estimation techniques that are well-known in the context of collocated projects. In addition, a wide range of cost drivers were identified, e.g. socio-cultural, geographical and time distances. Nothing was identified about the impact of sourcing strategies on the effort estimates’ accuracy. None of the included primary studies provided information about the way the effort estimation process is performed in the GSE context.
6.2 S2 - An Empirical Investigation on Effort Estimation in Agile Global Software Development

S2 [23] aimed to complement the evidence of S1 with the state of the practice of effort estimation in the agile GSE context. To do so, a survey was carried out using as instrument an on-line questionnaire. The population analyzed in S2 was composed by software practitioners experienced in effort estimation within agile GSE projects.

Results show that the effort estimation techniques used within the agile GSE and collocated are the same, with planning poker being the one employed the most. Sourcing strategies were found to have no or a small influence upon the choice of estimation techniques.

With regard to effort predictors, aspects that affect negatively communication, coordination and control processes in GSE projects, e.g. socio-cultural and time distances, were reported. In addition, factors that equally impact effort in collocated projects, e.g. team experience, were identified.

Many challenges that impact the accuracy of the effort estimates were also reported by the respondents, such as problems with software requirements and the fact that the communication overhead between sites is not properly accounted for.

The findings of this study indicate that the effort estimation process in both collocated and globally distributed agile projects is very similar, although the cost drivers depend on the operational model (collocated or distributed).

6.3 S3 - Taxonomies in Software Engineering: A Systematic Mapping Study

S3 [140] presents the results of a SMS that aimed at identify and analyze the state of the art of Software Engineering taxonomies.

We identified 244 relevant primary studies reporting SE taxonomies. Results show that a remarkable number of different subjects (240) have been classified in all SE knowledge areas. We also observed that many taxonomies are cited many times by other studies, which suggests that the SE taxonomies are considered relevant.

Results also show that most Software Engineering taxonomies are designed in an ad-hoc way, which negatively affects the clarity and usefulness of these taxonomies. To address this gap, we revised an existing taxonomy design method. To do so, we used the results of S3 and additional literature on taxonomy design from Psychology, which is a knowledge field that is more mature than Software Engineering.
A GSE taxonomy was recently proposed by Smite et al. [149] to address the need for knowledge classification and definition of a common terminology in the context of GSE. However, their taxonomy does not include dimensions related to the effort estimation process, which reduces the taxonomy’s usefulness to report effort estimation studies in the GSE context.

Therefore, the main goal of S4 [24] was to present a specialized taxonomy to classify GSE projects, which can help researchers and practitioners to report GSE effort estimation studies in a standardized way. To do so, the specialized taxonomy was designed as follows:

- A process to update Smite et al.’s taxonomy was designed. In this process, 2 types of dimensions were incorporated into the original taxonomy, named extended dimension (generic factors) and specialized dimension (specific factor) respectively.
- The cost drivers reported the most in S1 and S2 were incorporated into Smite et al.’s taxonomy as extended dimensions (setting, software process model, language distance, power distance and uncertainty avoidance).
- Aspects related to the effort estimation process were incorporated into Smite et al.’s taxonomy as specialized dimensions (effort estimation role, effort estimation stage and effort estimation process architectural model), because they are only useful to classify GSE projects with focus on effort estimation.
- Both extension and specialization were performed following the method proposed in S3.

To illustrate the specialized taxonomy, we classified 8 completed GSE projects, which were extracted from one of the primary studies included in S1. Results showed that the specialized taxonomy can be used to classify GSE projects with clarity and simplicity. For this reason, we believe that it can be widely adopted by the GSE community, easing the analysis, comparison and aggregation of the findings of new GSE effort estimation studies.

7 CONTRIBUTIONS

The main contributions of this thesis are:

- A description of the state of the art on effort estimation in the GSE context (S1 - Chapter 2).
- A description of the state of the practice on effort estimation in the agile GSE context (S2 - Chapter 3).
• A description of the state of the art on taxonomies in Software Engineering and a revised method to design SE taxonomies (S3 - Chapter 4).

• A specialized taxonomy (classification scheme) with focus on effort estimation in the GSE context (S4 - Chapter 5).

8 LIMITATIONS

There are different classifications for validity threats in literature. According to Wohlin et al. [153], there are two main types of validity:

• **Internal validity** – It relates to the extent to which the findings of a study result only from the variables tracked by the researcher.

• **External validity** – It relates to the capacity of generalizing the findings of a study.

Since the validity threats associated with this thesis (along with the respective applied mitigation actions) are minutely discussed in Chapters 2, 3, 4 and 5, in this section we present only additional limitations that affect the overall thesis. Note that it was not possible to apply mitigation actions regarding the limitations listed below.

In S1 we identified just few studies on effort estimation in the GSE context [21]. It means that the existing evidence on effort estimation in GSE, which was used as one of the bases of the specialized taxonomy presented in S4 (Chapter 5), is not very strong (external validity threat).

S2 (Chapter 3) was conducted only in the context of companies that employ agile GSE. Therefore, it cannot be argued that this thesis covers the entire state of the practice on effort estimation in GSE, since companies that only rely on prescriptive software development approaches did not participated in the conducted survey (external validity threat).

In S4 (Chapter 5), the dimensions added to Smite et al.’s GSE taxonomy were deemed as necessary to classify GSE scenarios focusing on effort estimation, but they do not represent an exhaustive list. Therefore, it is very likely that different SE perspectives can demand additional dimensions (external validity threat).

Although experts on effort estimation and GSE have contributed to S4, we believe that more feedback from other experts can contribute further to the correctness and usefulness of the specialized taxonomy presented in this thesis (external validity threat).

In S4, it was not possible to illustrate a class of the specialized dimension “effort estimation process architectural model”, because the available data used to illustrate the usage of the extended and specialized taxonomy did not cover such a class (external validity threat).
The results of this thesis show that effort estimation in GSE is poorly covered by existing literature. We believe that the following reasons contribute to such a situation:

1. Researchers and practitioners maybe believe that there is no dramatic difference between effort estimation in collocated and globally distributed projects, hence there would not be a need for specific research in the GSE context.
2. The absence of a classification scheme and common terminology regarding effort estimation in GSE might be hindering new GSE effort estimation research endeavors.

Considering all the negative effects of GSE issues on coordination, communication and control of GSE projects, we believe that effort estimation in such a context is more challenging than in the collocated context. However, there is still need to clarify whether there are unique aspects in the GSE context that would justify specific effort estimation related research in this context.

The specialized taxonomy presented in this thesis can foster further research related to effort of GSE projects and effort estimation in GSE projects. The classification scheme and terminology presented herein can be used by researchers and practitioners to report new effort-related studies in the GSE context, facilitating the analysis, comparison and aggregation of results.

Some issues are not addressed by this thesis and should be further investigated to provide a better understanding about the effort of GSE projects:

- The state of the practice should be complemented with data from companies that not necessarily apply agile practices to develop software in a globally distributed way.
- The way cost drivers are measured and incorporated into the effort estimation process by practitioners should be clarified.
- Further investigation should be conducted to clarify the impact of different factors on the effort of GSE projects (e.g. sourcing strategies, uneven experience between sites and uneven experience between team members).

Therefore, the following research questions should be answered in future research:

1. How do different cost drivers impact the effort of GSE projects?
2. How do practitioners measure cost drivers that impact the effort of GSE projects?
Effort Estimation in Global Software Development: A Systematic Literature Review

1 Introduction

Many software development companies have tried to increase profits by improving the time-to-market of their products, reducing costs by hiring people from countries with cheaper work-hours and defying the "clock" by running the projects during 24 hours. As a result, we can see a great number of software development projects performed globally, distributed in many different sites and normally located in different countries. This distribute setting of managing a software project is called Global Software Development (GSD).

The advantages related to GSD can only be achieved if projects are well managed. There are many challenges related to managing GSD projects. We could remark the following key challenges \[56, 59\]:

- **Communication** – Different languages used by each site of a GSD project hinder the execution of the work.
- **Trust** – Cultural variation complicates the process of trusting between each site.
- **Coordination over distance** – Different time zones makes harder the process of hand off between sites.

Those challenges affect directly the effort estimation of a project, which is one of the critical tasks related to the management of a software development project \[42\].

There are many studies related to effort estimation methods/techniques for co-located projects \[92, 94, 93\]. Nevertheless, existing evidence suggests that effort estimation methods/techniques and cost drivers used in the context of co-located projects may not be readily applicable to global software development projects \[113, 102, 105, 81\].

There are many systematic literature reviews (SLR) regarding GSD. Kroll \textit{et al.} \[77\] identified areas within Software Engineering have been investigated regarding GSD. Schneider \textit{et al.} \[121\] performed a SLR to categorize solutions in the GSD context by using process areas. Jalali \textit{et al.} \[69\]
and Hossain et al. [63] performed systematic literature reviews looking at the use of agile methods in GSD. Da Silva et al. performed two SLRs on project management in the context of GSD [33, 34]. Prikladnicki et al. [111] identified papers that describe process models for GSD in the context of overseas outsourcing. Mishra et al. developed a SLR to find the current research and practice of quality management in GSD projects [96]. Smite et al. [148] investigated empirical evidence in GSD-related research literature. Nurdiani et al. [104] discuss risks and strategies to mitigate those risks in the context of GSD projects. Monasor et al. [99] present a SLR regarding methodologies for training and teaching students and developers to the needs of GSD projects. Nidhra et al. [103] present the state of the art and the state of the practice regarding knowledge transfer in GSD. Matusse et al. [90] investigated evidences about the existence of metrics and indicators that are specific to GSD. Many of these previous SLRs focused on the challenges and solutions regarding communication, coordination and cooperation problems in GSD projects [132, 51, 8, 131].

Nevertheless, to the best of our knowledge, there is no SLR regarding effort estimation in the context of GSD. Since nowadays GSD plays an important role to the software industry, a SLR on effort estimation in the context of GSD is deemed quite important as its results on the state of the art in that field can inform both researchers and practitioners. Therefore, the goal and main contribution of this paper is to present a SLR of effort estimation in the context of GSD. In doing so, research gaps that can be the focus of future work were also identified.

The remainder of this paper is organized as follows: Section 2 details the SLR, followed by the presentation and discussion of its results in Sections 3, 4 and 5 respectively. Finally, Section 6 presents the conclusions and view on future work.

2 SYSTEMATIC LITERATURE REVIEW

This section details the SLR, which includes the research questions that guided this SLR as well as the procedures employed to perform the entire SLR process. The SLR detailed herein followed the guidelines defined by Kitcheham and Charters [74]. The protocol was developed jointly by the first, third and forth authors.

2.1 Research questions

The research questions were formulated by using the PICOC criteria [107]. Note that we did not include the comparison attribute because our focus is not to compare interventions. The PIOC is as follows:

- **Population** - Global software development projects.
• **Intervention** - Effort estimation methods/techniques/size metrics/cost drivers.
• **Outcomes** - The accuracy of the effort estimation methods/techniques.
• **Context** - Any possible study, as long as it is an empirical study within the context of GSD will be considered.

Therefore, the addressed research questions are:

• **Question 1** - What methods/techniques have been used to estimate effort in GSD?
  – 1a - What metrics have been used to measure the accuracy of effort estimation methods/techniques in GSD projects?
  – 1b - What are the accuracy levels for the observed estimation methods?

• **Question 2** - What effort predictors (cost drivers/size metrics) have been used to estimate effort in GSD?

• **Question 3** - What are the characteristics of the datasets used for effort estimation in GSD?
  – 3a - What are the domains represented in the dataset (academic/industry projects)?
  – 3b - What are the types represented in the dataset (single-company/cross-company)?
  – 3c - What are the application types represented in the dataset (web-based/traditional)?

• **Question 4** - What are the used sourcing strategies (offshore outsourcing/offshore insourcing)?
  – 4a - What are the countries involved?
  – 4b - How many sites are involved?
  – 4c - How many sites per country are involved?
  – 4d - Which topologies (centralized/distributed) have been used regarding the effort estimation process?
  – 4e - How did each of the sites participate in the effort estimation process (knowledge/data/both)?

• **Question 5** - Which activities were considered in the effort estimation process?

2.2 **Search strategy**

After defining the research questions, we set up a strategy to define a search string and, consequently, identifying the primary studies. To avoid researcher bias, we used the following procedure to define the search string used in this paper:
1. Analyze the questions and identifying the main words in terms of population, intervention and outcome;
2. Analyze collected relevant papers and checking the keywords;
3. Point out alternative spellings and synonyms for major terms, if existing;
4. Connecting alternative spellings and synonyms by using the Boolean OR;
5. Linking the main terms from population, intervention and outcome by using the Boolean AND;
6. Evaluate the resultant search string and perform adjustments by using the quasi-gold standard [154].

The steps one to five were performed by the first author and the last one was performed jointly by the first and forth authors.
As a result, we obtained the following search string:

\[
(\text{effort OR cost OR resource OR size OR metric OR measure OR measurement}) \text{ AND (estimation OR estimating OR estimate OR prediction OR predicting OR predict OR assessment OR forecasting OR forecast OR calculation OR calculate OR calculating OR calculation OR sizing OR measure OR measuring)} \text{ AND ("global software development" OR "global software engineering" OR "globally distributed development" OR "globally distributed work" OR "distributed software development" OR "distributed software engineering" OR "distributed development" OR "multi-site development" OR "multisite development" OR "multi site development" OR "geographically distributed software" OR "collaborative software engineering" OR "collaborative software development" OR "dispersed software development" OR "global software teams" OR "distributed teams" OR "spread teams" OR "dispersed teams" OR "global teams" OR "virtual teams" OR "offshore outsourcing" OR "offshore software development")}
\]

2.3 Search process

Once the search string was defined, the search process took place and was divided in two phases: initial search phase and secondary search phase.
2.3.1 The initial search phase

The initial phase involved identifying primary sources and searching primary studies through those sources by using the search string defined in Subsection 2.2.

The used primary sources are listed in Table 2.1, as well as the number of search results and number of relevant papers. We selected those primary sources based on venues where the literature related to GSD and effort estimation is mostly published.

The search process was performed using titles and abstracts. In addition, results were limited to peer-reviewed conference papers and journal articles, published between 2001 and 2013. The starting year was set to 2001 because that was the year when the term Global Software Development was coined [58].

Table 2.1: Summary of search results

<table>
<thead>
<tr>
<th>Database name / Search engine</th>
<th>Search results</th>
<th>Selected articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scopus</td>
<td>299</td>
<td>17</td>
</tr>
<tr>
<td>IEEExplore</td>
<td>69</td>
<td>9</td>
</tr>
<tr>
<td>ACM Digital Library</td>
<td>59</td>
<td>7</td>
</tr>
<tr>
<td>ScienceDirect</td>
<td>26</td>
<td>0</td>
</tr>
<tr>
<td>Compendex</td>
<td>85</td>
<td>13</td>
</tr>
<tr>
<td>Inspec</td>
<td>108</td>
<td>10</td>
</tr>
<tr>
<td>Web of Science</td>
<td>63</td>
<td>9</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>709</strong></td>
<td><strong>65</strong></td>
</tr>
<tr>
<td><strong>Without duplicates</strong></td>
<td><strong>379</strong></td>
<td><strong>24</strong></td>
</tr>
</tbody>
</table>

2.3.2 The secondary search phase

The secondary search phase was carried out in order to maximize the chances that all important studies were included in this SLR. As part of this phase, all the references cited in the papers gathered in the first phase were checked. This phase was performed jointly with the study selection procedure.

2.4 Study selection

The study selection was carried out by applying the following inclusion and exclusion criteria:
• Inclusion criteria:
  1. Studies that present effort estimation models/methods/metrics for GSD AND;
  2. Studies that have empirical evidence AND;
  3. Studies described in English AND;
  4. Studies reported in peer reviewed workshop OR conference OR journal OR are reported in a technical report OR thesis found as a result of the secondary search phase.

• Exclusion criteria:
  1. Studies that do not present effort estimation models/methods/metrics for GSD OR;
  2. Studies without empirical evidence OR;
  3. Studies that are not written in English;

The screening process, which was used to select primary studies, was conducted in two stages. The first stage, performed by the first three authors, comprised reading studies’ titles and abstracts. A total of 24 studies were judged as relevant (Table 2.1). All the works selected in that stage were identified by a letter (S) and a number ranging from 1 to 24.

We downloaded the full-text of the studies selected in the first stage, in order to perform the second stage of the studies’ selection process. The first author was responsible to download the full-texts of the studies retrieved from Scopus, IEEExplore, Web of Science and Inspec. The second author downloaded the full-texts of the studies retrieved from Compendex, and ACM digital library. No full-text was downloaded from ScienceDirect, since no study from that database was considered relevant.

In the second stage, we read the full-text of the 24 papers selected in the previous stage. The secondary search phase (2.3) was performed concurrently. The final list of the first three authors were compared and discussed until consensus was reached.

As a result of the secondary search phase, the first author found an additional study. After a discussion with the other authors, that study was included, thus making it a total of 25 studies to be screened. This 25th study was identified as D1, to be differentiated from the other 24 studies retrieved in the initial search phase.

As a result of the second stage of the study selection, the authors consensually decided to exclude 17 studies, due to the following reasons:

• Nine studies were excluded because they did not focus on effort estimation models, methods, size metrics or cost drivers in the GSD context (Exclusion criterion 1).
• Eight studies were excluded because they did not provide empirical evidences to support their findings (Exclusion criterion 2).
Thus, by the end of the study selection phase, 8 papers passed our inclusion criteria and were assessed in terms of quality (Subsection 2.5).

2.5 Study quality assessment

In order to evaluate the quality of the selected studies, we developed a questionnaire with 13 questions. The quality score for a paper could range from 0 to 13. The higher the score, the better the quality of the evidence provided by the paper. The questionnaire was developed by using the guidelines defined by Kitcheham and Charters [74].

Each question could be answered using YES (1.0), NO (0.0) or PARTIALLY (0.5). If a question was answered with PARTIALLY, it means that the work addresses the question in a correct way, but the answer could not be confirmed by using the content of the paper. Any study which scored 3.25 or below (first quartile) was excluded.

The questionnaire had the following questions:

- Are the research aims clearly specified?
- Was the study designed to achieve these aims?
- Are the prediction techniques used clearly described and their selection justified?
- Are the variables considered by the study suitably measured?
- Are the data collection methods adequately detailed?
- Is the data collected adequately described?
- Is the purpose of the data analysis clear?
- Are the statistical techniques used to analyze the data adequately described and their use justified?
- Are the results discussed, rather than only presented?
- Do the researchers discuss any problems with the validity/reliability of their results?
- Are all research questions answered adequately?
- How clear are the links between data, interpretation and conclusions?
- Are the findings based on multiple projects?

The first and the second authors of this work have performed the quality assessment for the 8 selected primary studies. The third author evaluated one of the 8 selected studies. The results were discussed until consensus was reached. The final scores can be seen in Table 2.2. Three studies were excluded due to their low quality score.

The lists with the included and excluded studies after the study selection and quality assessment are shown in Table 2.3.
Table 2.2: Final scores for the evaluated works

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>S7</td>
<td>10.5</td>
</tr>
<tr>
<td>S10</td>
<td>7</td>
</tr>
<tr>
<td>S11</td>
<td>10.5</td>
</tr>
<tr>
<td>S13</td>
<td>3</td>
</tr>
<tr>
<td>S15</td>
<td>13</td>
</tr>
<tr>
<td>S20</td>
<td>5</td>
</tr>
<tr>
<td>S24</td>
<td>3</td>
</tr>
<tr>
<td>D1</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 2.3: Included and excluded works after the study selection and quality assessment

<table>
<thead>
<tr>
<th>Included works</th>
<th>Excluded works</th>
</tr>
</thead>
<tbody>
<tr>
<td>S7 [105], S10 [15], S11 [100], S15 [113], S20 [102]</td>
<td>S1 [85], S2 [72], S3 [88], S4 [14], S5 [87], S6 [6], S8 [81], S9 [53], S12 [54], S13 [78], S14 [80], S16 [39], S17 [9], S18 [133], S19 [73], S21 [82], S22 [109], S23 [146], S24 [35], D1 [13]</td>
</tr>
</tbody>
</table>
2.6 Data extraction

The first and the second authors of this paper have performed the data extraction for all the five final selected studies. The third author performed the extraction procedure for one of those five studies, thus to compare the quality of the extractions.

All the extracted values were compared and discussed until consensus was reached. The performed comparison was very important since some values were noticed just by one of the authors.

3 Results

After the data extraction, the extracted data were consolidated in 12 tables so to make it easier to understand the relationship between data and the research questions.

The percentage column, which is part of all tables of this section (except for Table 2.6), indicates the percentage of studies out of the total number of studies (five).

3.1 Question 1

Data extraction for Question 1 looked at what methods which have been used for estimating effort in GSD, the accuracy metrics to evaluate the methods and also the calculated accuracy values. Those data were synthesized in Tables 2.4, 2.5 and 2.6.

Table 2.4 shows the identified effort estimation methods/techniques, representing three different approaches: i) expert based approaches (Delphi [17], planning poker [28], expert judgment [91] and expert judgment based on ISBSG database [67]); ii) algorithmic based approaches (COCOMO II [91] and SLIM [1]); and iii) an artificial intelligence approach (Case-based reasoning [91]).

One of the primary studies did not investigate directly any effort estimation approach, but rather focused on the cost drivers. Another primary study used function points and use case points as surrogates for effort estimation techniques. However, both are size metrics. The primary study S20 presented an approach based on linear regression but that approach was not empirically evaluated. So, from S20 we only extracted cost drivers as they were empirically validated.

Table 2.5 shows the accuracy metrics used in the primary studies. As we can see, just one of the primary studies presented the accuracy of the proposed effort estimation approach based on the most used accuracy metrics MMRE, MdMRE and Pred(25) [91].
Table 2.4: Identified effort estimation methods

<table>
<thead>
<tr>
<th>Estimation method</th>
<th>Approach type</th>
<th>Study ID</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>COCOMO II</td>
<td>Algorithmic</td>
<td>S11</td>
<td>20</td>
</tr>
<tr>
<td>SLIM</td>
<td>Algorithmic</td>
<td>S11</td>
<td>20</td>
</tr>
<tr>
<td>Case-based reasoning</td>
<td>Artificial reasoning</td>
<td>S15</td>
<td>20</td>
</tr>
<tr>
<td>Expert judgment</td>
<td>Expert</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>ISBSG based expert judgment</td>
<td>Expert</td>
<td>S11</td>
<td>20</td>
</tr>
<tr>
<td>Delphi</td>
<td>Expert</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>Planning poker</td>
<td>Expert</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>Function point count</td>
<td>-</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>Use case point count</td>
<td>-</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>No estimation approach</td>
<td>-</td>
<td>S10</td>
<td>20</td>
</tr>
</tbody>
</table>

The MRE (Magnitude of Relative Error) is calculated by considering the absolute difference between the actual and estimated effort, relative to the actual effort. The MMRE and MdMRE are respectively the mean and median MRE. Pred\((25)\) is the percentage of estimates with a MRE of 25% or less.

The primary study S11 used just the difference between the actual and the estimated efforts (deviation). The primary study S10 did not use any accuracy metric, since that work just focused on the cost drivers. The studies S7 and S20 did not evaluate the considered effort estimation approaches in terms of accuracy metrics.

Table 2.5: Identified accuracy metrics for the observed effort estimation methods

<table>
<thead>
<tr>
<th>Accuracy metric</th>
<th>Study ID</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviation</td>
<td>S11</td>
<td>20</td>
</tr>
<tr>
<td>MMRE</td>
<td>S15</td>
<td>20</td>
</tr>
<tr>
<td>MdMRE</td>
<td>S15</td>
<td>20</td>
</tr>
<tr>
<td>Pred((25))</td>
<td>S15</td>
<td>20</td>
</tr>
<tr>
<td>No accuracy metrics</td>
<td>S7, S10, S20</td>
<td>60</td>
</tr>
</tbody>
</table>

Table 2.6 shows the measured accuracy values for each observed effort estimation technique/method. As above-mentioned, only study S15 used MMRE, MdMRE and Pred\((25)\) to evaluate the accuracy of the presented
effort estimation approach, thus making it difficult to assess how good were the estimation techniques put forward in studies S7 and S11.

In S11, three projects were used to evaluate three different approaches. The deviation between the actual and the estimated effort was calculated for each approach in the context of each project.

The ISBSG database [67], used in S11, provides three different estimation values: the lower time to perform a project; the time in which a project is more likely to be performed; and the maximum time to finish a project. So, in S11, for each project three deviations were calculated using the actual effort and the three above-mentioned values: Lower (L), Expected (E) and Upper (M).

When applying the thresholds suggested by Conte et al. [30] (good MMRE if calculated value $\leq 25\%$, good MdMRE if calculated value $\leq 25\%$ and good Pred(25) if calculated value $\geq 75\%$) to assess the effort prediction accuracy detailed in S15, results suggest that the prediction technique proposed in S15 provided good estimates.

It is impossible to compare the estimation accuracy between the effort estimation approaches identified in our SLR, since most of the primary studies did not use any prediction accuracy metric. In addition, the two studies that employed some accuracy measure did not use the same one.

3.2 Question 2

Data collection for Question 2 looked for data regarding the used cost drivers and size metrics. The extracted data is presented in Tables 2.7 and 2.8, where all the predictors that do not represent size measures were grouped as cost drivers. Note that whenever the same cost factor presented different names in primary studies, we chose the name we believed to be the most suitable and added it to Tables 2.7 and 2.8.

Table 2.7 shows that there is a wide range of cost drivers in the selected primary studies, where only a single study (S7) did not show any cost driver.

Most of the cost drivers that seem to impact GSD projects are related to the specific factors of globally performed software projects, like differences in culture, language and time zone. The cost drivers that were used the most were "Time zone" (80% of the primary studies), "Language and cultural differences" (60% of the primary studies), "Communication" (40% of the primary studies) and "Process model" (40% of the primary studies).

Table 2.8 lists the size metrics employed in the primary studies, showing that function points and lines of code are the most used size metrics in the context of effort estimation for GSD. One of the studies also employed use case points and user story points as possible size metrics. The primary
<table>
<thead>
<tr>
<th>Estimation method</th>
<th>Study ID</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>COCOMO II</td>
<td>S11</td>
<td>Project 1: 2.26; Project 2: 1.35; Project 3: 1.53</td>
</tr>
<tr>
<td>SLIM</td>
<td>S11</td>
<td>Project 1: 0.36; Project 2: 1.55; Project 3: 2.03</td>
</tr>
<tr>
<td>ISBSG</td>
<td>S11</td>
<td>Project 1: 0.68 (L), 3.44 (E), 12.65 (U); Project 2: 1.96 (L), 2.42 (E), 12.24 (U); Project 3: 0.3 (L), 4.23 (E), 14.53 (U).</td>
</tr>
<tr>
<td>Case-based reasoning</td>
<td>S15</td>
<td>MMRE: 15.99%;MdMRE: 11.67%;Pred(25): 84.12%</td>
</tr>
<tr>
<td>Expert judgment</td>
<td>S7</td>
<td>no accuracy calculated</td>
</tr>
<tr>
<td>Planning poker</td>
<td>S7</td>
<td>no accuracy calculated</td>
</tr>
<tr>
<td>Delphi</td>
<td>S7</td>
<td>no accuracy calculated</td>
</tr>
<tr>
<td>Function point count</td>
<td>S7</td>
<td>no accuracy calculated</td>
</tr>
<tr>
<td>Use case point count</td>
<td>S7</td>
<td>no accuracy calculated</td>
</tr>
</tbody>
</table>
Table 2.7: Identified cost drivers

<table>
<thead>
<tr>
<th>Cost driver</th>
<th>Study ID</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Zone</td>
<td>S10, S11, S15, S20</td>
<td>80</td>
</tr>
<tr>
<td>Language And Cultural Differences</td>
<td>S10, S20, S15</td>
<td>60</td>
</tr>
<tr>
<td>Communication</td>
<td>S11, S20</td>
<td>40</td>
</tr>
<tr>
<td>Process Model</td>
<td>S15, S20</td>
<td>40</td>
</tr>
<tr>
<td>Communication Infrastructure</td>
<td>S10</td>
<td>20</td>
</tr>
<tr>
<td>Communication Process</td>
<td>S10</td>
<td>20</td>
</tr>
<tr>
<td>Travel</td>
<td>S20</td>
<td>20</td>
</tr>
<tr>
<td>Competence level</td>
<td>S10</td>
<td>20</td>
</tr>
<tr>
<td>Requirements Legibility</td>
<td>S10</td>
<td>20</td>
</tr>
<tr>
<td>Process Compliance</td>
<td>S10</td>
<td>20</td>
</tr>
<tr>
<td>Response Delay</td>
<td>S11</td>
<td>20</td>
</tr>
<tr>
<td>Unrealistic Milestones</td>
<td>S11</td>
<td>20</td>
</tr>
<tr>
<td>People Interesting</td>
<td>S11</td>
<td>20</td>
</tr>
<tr>
<td>Trust</td>
<td>S11</td>
<td>20</td>
</tr>
<tr>
<td>Clients Unawareness</td>
<td>S11</td>
<td>20</td>
</tr>
<tr>
<td>Shared Resources</td>
<td>S11</td>
<td>20</td>
</tr>
<tr>
<td>Team Structure</td>
<td>S11</td>
<td>20</td>
</tr>
<tr>
<td>Work Pressure</td>
<td>S11</td>
<td>20</td>
</tr>
<tr>
<td>Work Dispersion</td>
<td>S15</td>
<td>20</td>
</tr>
<tr>
<td>Range of Parallel-sequential Work Handover</td>
<td>S15</td>
<td>20</td>
</tr>
<tr>
<td>Client-specific Knowledge</td>
<td>S15</td>
<td>20</td>
</tr>
<tr>
<td>Client Involvement</td>
<td>S15</td>
<td>20</td>
</tr>
<tr>
<td>Design and Technology Newness</td>
<td>S15</td>
<td>20</td>
</tr>
<tr>
<td>Team Size</td>
<td>S15</td>
<td>20</td>
</tr>
<tr>
<td>Project Effort</td>
<td>S15</td>
<td>20</td>
</tr>
<tr>
<td>Development Productivity</td>
<td>S15</td>
<td>20</td>
</tr>
<tr>
<td>Defect Density</td>
<td>S15</td>
<td>20</td>
</tr>
<tr>
<td>Rework</td>
<td>S15</td>
<td>20</td>
</tr>
<tr>
<td>Reuse</td>
<td>S15</td>
<td>20</td>
</tr>
<tr>
<td>Project Management Effort</td>
<td>S15</td>
<td>20</td>
</tr>
</tbody>
</table>
study S10 did not show any size metrics because it focused solely on cost drivers.

The use of lines of code as size metric seemed somewhat surprising to us given that this metric is really very difficult to forecast early in the development life cycle [5]. Such difficulties were one of the main motivations as to why function points were proposed so many years ago.

It is interesting to notice that just length metrics (lines of code) and functionality metrics (function points, use case points and story points [122]) were identified in the selected primary studies, despite other possible types of size metrics (e.g. complexity metrics, like cyclomatic complexity [40]).

Table 2.8: Identified size metrics

<table>
<thead>
<tr>
<th>Size metric</th>
<th>Study ID</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Function points</td>
<td>S7, S11, S20</td>
<td>60</td>
</tr>
<tr>
<td>Lines of code</td>
<td>S11, S15, S20</td>
<td>60</td>
</tr>
<tr>
<td>Use case points</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>Story points</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>No size metric used</td>
<td>S10</td>
<td>20</td>
</tr>
</tbody>
</table>

3.3 Question 3

Data collection for Question 3 looked at domains and types regarding the data used to evaluate the proposals in each primary study. That question also looked at the evaluated application types. The data were arranged in Tables 2.9, 2.10 and 2.11.

Table 2.9 presents the domain relating to the type of dataset used in the primary studies’ investigation. It shows that all the selected primary studies employed industrial data to evaluate and/or making conclusions about the effort estimation approaches and/or predictors (size metrics and cost drivers).

Table 2.9: Identified dataset domains

<table>
<thead>
<tr>
<th>Domain</th>
<th>Study ID</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industry</td>
<td>S7, S10, S11, S15, S20</td>
<td>100</td>
</tr>
<tr>
<td>Academia</td>
<td>none</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 2.10 presents the identified dataset types in the primary studies. Two of the primary studies (S10 and S15) used single-company data (data
that comes from a single company). Study S11 used cross-company data, since the ISBSG database was used to estimate effort in that study. Studies S7 and S20 did not state the type of the dataset used.

Table 2.10: Identified dataset types

<table>
<thead>
<tr>
<th>Type</th>
<th>Study ID</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-company</td>
<td>S10, S15</td>
<td>40</td>
</tr>
<tr>
<td>Cross-company</td>
<td>S11</td>
<td>20</td>
</tr>
<tr>
<td>Not stated</td>
<td>S7, S20</td>
<td>40</td>
</tr>
</tbody>
</table>

Table 2.11 presents the identified application types used in the primary studies. We considered only two types of applications during the data extraction - traditional and Web-based applications. In this work, a traditional application means an application which does not require Web infrastructure. Only one study (S11) stated the types of the application used in the study, which included both traditional and web-based applications.

Table 2.11: Identified application types

<table>
<thead>
<tr>
<th>Types</th>
<th>Study ID</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web-based</td>
<td>S11</td>
<td>20</td>
</tr>
<tr>
<td>Traditional</td>
<td>S11</td>
<td>20</td>
</tr>
<tr>
<td>Not stated</td>
<td>S7, S10, S15, S20</td>
<td>80</td>
</tr>
</tbody>
</table>

3.4 Question 4

Data collection for Question 4 looked at specific aspects of Global Software Development. The extracted data is presented in Tables 2.12, 2.13, 2.14 and 2.15.

Table 2.12 presents the identified sourcing strategies in the primary studies. During the data extraction, we considered three options to address the question about sourcing strategies:

- **Offshore outsourcing** - A company moves the software development to an external third party abroad [149].
- **Offshore insourcing** - A company moves the software development to a branch established abroad [149].
- A combination of both above-mentioned strategies.
Table 2.12 shows that just the offshore insourcing strategy was considered by the selected primary studies (S7, S10, S11, S15). One of the works did not inform which kind of sourcing strategy was used (S20).

Table 2.12: Identified sourcing strategies

<table>
<thead>
<tr>
<th>Sourcing strategy</th>
<th>Study ID</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Offshore insourcing</td>
<td>S7, S10, S11, S15</td>
<td>80</td>
</tr>
<tr>
<td>Offshore outsourcing</td>
<td>none</td>
<td>0</td>
</tr>
<tr>
<td>Both</td>
<td>none</td>
<td>0</td>
</tr>
<tr>
<td>Not stated</td>
<td>S20</td>
<td>20</td>
</tr>
</tbody>
</table>

Table 2.13 presents the identified number of involved countries in the collected works. In the selected primary studies, at least three countries were identified as involved in the context of a GSD project.

Table 2.13: Identified number of involved countries

<table>
<thead>
<tr>
<th>Number</th>
<th>Study ID</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>S11</td>
<td>20</td>
</tr>
<tr>
<td>7</td>
<td>S10</td>
<td>20</td>
</tr>
<tr>
<td>10</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>Not stated</td>
<td>S15, S20</td>
<td>40</td>
</tr>
</tbody>
</table>

Table 2.14 presents the name of the identified countries, thus showing a wide range of countries. The USA, China, India and UK are the four countries mostly involved in GSD projects within the context of this SLR. All the selected studies that presented this information had at least a site in a developed country and a site in developing country. Since one of the main reasons for using GSD is the cost-reduction, it is not surprising that those countries were identified. USA and UK have higher costs, specially regarding human resources, compared to countries like India [119].

Table 2.15 presents the identified number of sites in the primary studies. A primary study should have at least two participating sites, each placed in a different country, to be characterized as within the context of GSD.

We observed a maximum number of four countries in a GSD project. Study S7 did not state the exact number of sites, and studies S15 and S20 did not report the number of sites involved.

Question 4d deals with the effort estimation process topologies. In this work, the topology of the effort estimation process is the way the work
Table 2.14: Identified countries

<table>
<thead>
<tr>
<th>Name</th>
<th>Study ID</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>S7, S10, S11</td>
<td>60</td>
</tr>
<tr>
<td>UK</td>
<td>S7, S11</td>
<td>40</td>
</tr>
<tr>
<td>India</td>
<td>S7, S10</td>
<td>40</td>
</tr>
<tr>
<td>China</td>
<td>S7, S10</td>
<td>40</td>
</tr>
<tr>
<td>Malaysia</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>Japan</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>Taiwan</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>Ireland</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>Brazil</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>Slovak Republic</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>Finland</td>
<td>S10</td>
<td>20</td>
</tr>
<tr>
<td>Germany</td>
<td>S10</td>
<td>20</td>
</tr>
<tr>
<td>Norway</td>
<td>S10</td>
<td>20</td>
</tr>
<tr>
<td>Sweden</td>
<td>S10</td>
<td>20</td>
</tr>
<tr>
<td>Pakistan</td>
<td>S11</td>
<td>20</td>
</tr>
<tr>
<td>Not stated</td>
<td>S15, S20</td>
<td>40</td>
</tr>
</tbody>
</table>

Table 2.15: Overall number of sites

<table>
<thead>
<tr>
<th>Number of sites</th>
<th>Study ID</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>S10, S11</td>
<td>40</td>
</tr>
<tr>
<td>more than 2</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>S10</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>S10</td>
<td>20</td>
</tr>
<tr>
<td>Not stated</td>
<td>S15, S20</td>
<td>40</td>
</tr>
</tbody>
</table>
regarding the estimation process is performed by the involved sites. There are three possible topologies:

- **Centralized** - All the inherent effort estimation calculations are centralized in just one site and the remaining sites just supply knowledge and/or data to the central site.
- **Distributed** - Each site of a GSD project performs their own effort estimation process independently. The estimates only relate to what the site is doing as part of the project.
- **Hybrid** - All the early effort estimation is performed as in a centralized topology, but late effort estimates can be obtained in a distributed fashion. We consider early effort estimation the one performed just after the requirement analysis. The late effort estimation process is the one performed after an implementation iteration.

Question 4e deals with the roles for sites in the effort estimation process. As above-mentioned, a site can be involved in the effort estimation process as a knowledge/data supplier and/or as the responsible for performing the inherent calculations.

We identified that none of the selected primary studies documented data about the aspects topology and the roles for a site in the effort estimation process. For this reason, we did not address Questions 4d and 4e.

3.5 Question 5

Data extraction for Question 5 looked at the activities of a GSD project considered in the effort estimation process. Herein we considered generic activities embraced by most of the well-known life cycles, e.g. Requirements and Testing.

The main goal in extracting this data was to identify the life cycle activities considered the most relevant as far as effort estimation is concerned, within the context of GSD projects. However, it was not possible to address this question, since none of the selected primary studies documented this aspect either.

4 DISCUSSION

In order to facilitate the reading of our observations regarding the extracted data, we divided our comments according to the SLR research questions.

Regarding Question 1 we have identified that:

- The used effort estimation approaches are well-known in the context of co-located projects. Those approaches have been used in the con-
text of globally distributed projects without any kind of adaptation in the mechanisms of each approach.

- There is no standard effort estimation approach, since all the primary studies presented different solutions to the observed problem. However, it seems that expert-based approaches are the ones used the most by the practitioners.
- Software engineering works with imprecise and uncertain knowledge [108]. However, none of the primary studies treated explicitly the uncertainty regarding the data usually used to perform the effort estimation process.
- Since just one of the selected studies employed commonly used accuracy metrics, it was impossible to compare the primary studies in terms of prediction techniques’ estimation accuracy.

Regarding Question 2 we have observed that:

- The primary studies showed a wide range of cost drivers where most of them were related with the GSD specific issues, like cultural, language and time zone differences.
- None of the studies discussed deeply about the listed cost drivers.
- The observed primary studies show a trend in applying original co-located effort estimation approaches in GSD projects with cost drivers which reflect the specific problems of globally distributed projects.
- Solely length and functionality size metrics were identified in the studies, where the metrics used the most are lines of code and function points.

Regarding Question 3 we have identified that:

- The extracted data showed that only industrial data has been considered to evaluate the proposed effort estimation approaches.
- Single-company data appears to be the most used type.
- Most of the primary studies did not explicitly specify the type of the estimated application.

Regarding Question 4 we have observed that:

- None of the primary studies considered the particularities of each sourcing strategy.
- We did not identify any study considering offshore outsourcing as sourcing strategy. Therefore, there is a lack of research about effort estimation in offshore outsourced projects. Consequently, all the findings of this SLR are related to offshore insourced projects.
- The primary studies showed a trend in which there was always a site located in a country where salaries for software developers are
quite high, and another site located in a country where salaries for developers are very low.

• The selected primary studies did not document the topology and the role of each site in the effort estimation process.

Finally, regarding Question 5, none of the selected studies explained how the data used to estimate the effort were extracted during the evaluated projects.

5 Threats to Validity

In terms of threats to the validity of this SLR, the major issue is whether we have failed to find all the relevant primary studies. In order to mitigate that threat, we performed a very deep search strategy. We used keywords found in the most relevant works in effort estimation and Global Software Development.

We cannot claim that we retrieved all the available literature. However, we can state that we retrieved as many studies as possible, considering the restrictions that we have applied to our search string to reduce the number of irrelevant papers.

We would like to remark that one important task performed during the search process, the secondary search phase, was performed solely to the primary studies selected in the initial search phase, it was just a small portion of the complete set of studies retrieved in the search process. However, since the inclusion and exclusion criteria of this SLR are very straight, we believe that most relevant works regarding the main topic of this work were evaluated.

It is important to note that only some of the initially retrieved studies complied with our inclusion criteria. Seven studies were excluded because they did not show any empirical evidence to support their proposals. Another three studies were excluded because of their quality. So, since we had just few studies to extract data and drawing conclusions about the state of the art of effort estimation in the GSD context, it is very hard to generalize the findings.

6 Conclusions

This paper presented the results of a systematic literature review of effort estimation in the context of Global Software Development. The main goal of this paper was to present the state of the art regarding effort estimation in the context of GSD, in order to inform both research and practice.

The initial search phase returned 379 unique results and just 24 were selected. Out of those 24, 17 were excluded after a further investigation performed by reading the full text of those primary studies. The remaining
7 studies were then used in the secondary search phase, from which 1 more study was retrieved, bringing the total of selected studies to 8. Finally, after the achievement of the quality assessment process, we had a final list of 5 studies.

Only few studies complied with this SLR’s inclusion criteria. Many studies were excluded because either they lack empirical evidence (eight) or have low quality (three).

It is important to note that none of the selected primary studies considered offshore outsourcing as sourcing strategy, it means that all the findings of this SLR are applicable to projects which apply offshore insourcing as sourcing strategy.

The identified effort estimation approaches are well-known in the context of co-located projects. The main difference appears to be in the used cost drivers. There is no standard effort estimation approach and the observed effort estimation approaches did not consider the uncertainty that is inherent to such domain. It was not possible to evaluate the accuracy of the observed approaches because the studies did not use the same accuracy metrics.

This SLR identified a wide range of cost drivers, where most of the primary studies presented cost drivers regarding cultural, language and time zone differences; these are factors directly related to a globally performed software project and they make the fulfillment of those projects more challenging.

The selected primary studies showed that the researchers are using just industrial single-company data; some studies did not document anything regarding the used data, as well as the type of the estimated applications. Most of the questions regarding specific aspects of GSD projects were not addressed. Studies did not document the effects of the used sourcing strategy and the topology of the effort estimation process. On the other hand, we identified that the primary studies showed a trend in which there were always a site located in a developed country, where salaries for software developers are quite high, and a site located in a developing country, where salaries for developers are very low. This was expected, since one of the biggest reasons to perform a software project globally is to reduce costs.

Finally, the selected primary studies did not state which activities of the development process were considered to calculate the effort estimation.

We believe that further investigation regarding effort estimation approaches used in co-located context is needed given the scarcity of primary studies identified in this SLR. We also put forward that the adaptation of those approaches considering the specific aspects of GSD and also the inherent uncertainty of the data [94] may provide more accurate effort estimates.
It seems that the GSD sourcing strategy and the effort estimation process topology can have a significant influence upon the effort estimates. For example, the transition of software or part of software developed by using the offshore outsourcing strategy seems to require more effort compared to the development carried out by using the offshore insourcing strategy. An effort estimation process performed in a distributed way seems to be more flexible and adequate to companies which apply agile methods.

However, since we did not find any evidence to support the above-mentioned assumptions, we believe that more investigation must be performed to confirm so. A single effort estimation technique may not be equally applicable to offshore outsourcing and offshore insourcing projects, thus warranting separate investigations that can later be combined. The same can be applied to the effort estimation process topologies.

Since we did not identified any study considering effort estimation in the context of offshore outsourced projects, we believe researches should explore this existent gap in order to provide solutions which could help practitioners in that kind of scenario.

Regarding the cost drivers, we believe that there is room for assessing their impacts on the effort and developing much more formal ways to measure them, which are directions for our future work.
AN EMPIRICAL INVESTIGATION ON EFFORT ESTIMATION IN AGILE GLOBAL SOFTWARE DEVELOPMENT

1 INTRODUCTION

Effort estimation, i.e. the process used to predict the effort needed to fulfill a given task [91], is a project management activity that is fundamental to manage resources in an effective way [91]. Reliable effort estimates increase the chances of accomplishing the required work related to a given software project without time or cost overruns [42].

Nowadays, software can be developed in many different contexts, by means of different software development approaches, such as Global Software Development (GSD – software development in a globally distributed manner) [59], Agile Software Development (ASD – software development based on agile methods) [123] and Agile Global Software Development (AGSD) [147, 70] (a combination of GSD and ASD). Given such diversity, we put forth that the way that effort is estimated within the context of a project may vary depending on the software development approach being used.

Despite the fact that there are lots of effort estimation techniques designed for the collocated context [93, 92, 94], i.e. software development performed by a team located in a single physical room [135], there is evidence that effort estimation techniques that were originally designed for collocated contexts are not readily applicable in the global context [113, 102, 105, 81]. To make effort estimation even more challenging in the AGSD context, there is also evidence that agile methods are not readily applied in the global context [10, 63, 69].

To better understand the state of the art in effort estimation in those three contexts (GSD, ASD and AGSD), three earlier studies were carried out:

- A systematic literature review (SLR) on effort estimation in GSD [21].
- A SLR on effort estimation in ASD [139].
The results from the two SLR’s above were then combined to identify the differences and commonalities and obtain evidence on effort estimation in AGSD [22].

Given the growing number of AGSD projects [69, 63], it is important to complement the evidence on the state of the art towards effort estimation in AGSD with evidence from the state of the practice, which is the main goal of the present work. To do so, we carried out a survey [44] among software practitioners with experience in effort estimation in the AGSD context.

The present paper details that survey and makes five main contributions within the AGSD context:

• It presents how existing effort estimation techniques have been applied in practice.
• It presents the predictors that have been considered in practice.
• It presents the impact of sourcing strategies on the selection of effort estimation techniques.
• It presents the accuracy of the effort estimates reported by practitioners.
• It presents challenges that impact the accuracy of the effort estimates.

The remainder of this paper is organized as follows: Section 2 presents related work. Section 3 details the employed research methodology, followed by the presentation and discussion of the results of the conducted empirical study in Sections 4 and 5, respectively. Section 6 discusses validity threats. Conclusions and future work are described in Section 7.

2 RELATED WORK

There has been only one previous survey within the context of effort estimation in the GSD domain, by Peixoto et al. [105]. They investigated how effort estimation techniques have been used by practitioners in the GSD context. Their results did not identify clear criteria used by practitioners when choosing an effort estimation technique.

The present survey differs from Peixoto et al.’s survey (S1) in the following ways:

• The context of our study is AGSD, whereas S1 was performed in the GSD context.
• S1 investigated the practices within a single company, whereas our work gathered data from practitioners from many different companies.
• S1 focused solely on effort estimation techniques, whereas ours also focused on the predictors used in effort estimation and the effort estimates’ accuracy.
In the following subsections we describe the research methodology employed in this paper.

3.1 Research Questions

To obtain a detailed understanding on how effort estimation has been performed in the AGSD context, our research questions focused not only on the methods used to estimate effort, but also on the effort predictors used, the accuracy of the effort estimates, and the impact that sourcing strategies may have on the selection of effort estimation methods, as follows:

- **Research question 1 (RQ1)** – Which techniques are used to estimate effort in AGSD?
- **Research question 2 (RQ2)** – What is the impact of the applied sourcing strategy on the selection of an effort estimation technique in AGSD?
- **Research question 3 (RQ3)** – Which effort predictors are used to estimate effort in AGSD?
- **Research question 4 (RQ4)** – What are the challenges that impact the accuracy of the effort estimates in AGSD?

3.2 Survey Design

A survey is a retrospective form of investigation that targets at gathering data from a wider population \([44]\). To perform surveys, it is mandatory to define its purpose, the analysis unit to be used and a representative sample of the population related to the research problem. In the present work, our survey was defined as follows:

- The **purpose** of the survey is to collect data on the state of the practice in effort estimation in the AGSD context.
- The **analysis unit** is the effort estimation process element (technique, cost driver or size metric).
- The **target population** consists of practitioners who have worked with effort estimation in the AGSD context.
- The **sampling unit** is the practitioner responsible for performing effort estimation in a given company.

The survey’s instrument was an on-line semi-structured questionnaire, designed using SurveyMonkey\(^2\) tool. The questionnaire encompassed closed

---

1 SurveyMonkey is a popular tool for on-line survey development and execution, see [http://www.surveymonkey.com](http://www.surveymonkey.com).
and open-ended questions. The motivation for using an on-line questionnaire was to maximize respondents’ participation and coverage.

The questionnaire had three types of questions: demographic questions, AGSD questions and effort estimation questions. Those questions reflected the survey’s research questions (see Subsection 3.1).

Demographic questions were related to the respondents’ country of residence, their job titles and number of years of experience as a professional in software development industry, teams’ sizes and types of applications developed by the respondents’ companies. AGSD questions were related to aspects of agile global software development, such as the agile methods and sourcing strategies employed by the respondents’ companies. Finally, effort estimation questions were related to the effort estimation techniques and predictors (size metrics/cost drivers) employed by the respondents.

It is important to highlight that the survey detailed herein is part of a wider investigation that has been conducted by the authors focusing on the state of the practice in effort estimation within the agile context (both global and collocated). Therefore, only the questions that are relevant for effort estimation in the AGSD context are presented below. Note that the numbers associated with each of the questions presented below may sometimes differ from the number used in the original questionnaire.

- **Demographic Questions:**
  - **Question 1 (SQ1)** – In which country do you work most of the time?
  - **Question 2 (SQ2)** – For how long have you been working as a professional in software development industry?
  - **Question 3 (SQ3)** – What is your job title?
  - **Question 4 (SQ4)** – What is the size of your team?
  - **Question 5 (SQ5)** – Which types of software systems does your company develop?

- **AGSD Questions:**
  - **Question 6 (SQ6)** – Which agile methods are employed in your company?
  - **Question 7 (SQ7)** – Which sourcing strategies are applied by your company?
  - **Question 8 (SQ8)** – What is the average number of offshore insourced sites when considering the AGSD projects performed by your company?
  - **Question 9 (SQ9)** – What is the average number of offshore outsourced sites when considering the AGSD projects performed by your company?
  - **Question 10 (SQ10)** – What is the most common role of your company in AGSD outsourced projects?
• Effort Estimation Questions:
  – **Question 11 (SQ11)** – Which effort estimation techniques for AGSD projects are employed in your company?
  – **Question 12 (SQ12)** – What is the average effort estimation accuracy for AGSD projects in your company?
  – **Question 13 (SQ13)** – Which of the following factors do you believe are fundamental for estimating effort in AGSD projects?
  – **Question 14 (SQ14)** – What are the challenges that impact the accuracy of the effort estimates in AGSD projects?

Questions SQ1–SQ13 used categorical measurement scales (either nominal or ordinal) and were compulsory. Questions SQ3, SQ5, SQ6, SQ8–Q11 and SQ13 also provided free-text space to input an “Other” category answer. Question SQ14 is an open-ended question, so just a free-text space was provided to answer such question. Table 3.1 details the scale type and points/categories used for each question.

3.3 **Survey Execution**

Survey respondents were invited primarily from the authors’ contact networks. Furthermore, the survey was advertised in two conferences², and in many LinkedIn³ communities on agile software development, global software development and software measurement. It was made clear in the invitation that only people who worked with effort estimation in the AGSD context should answer the questionnaire.

The survey’s questionnaire was available on-line from August 12th to October 10th 2014 and was answered by a self-selected sample of 51 respondents. Prior to making the questionnaire available on-line, it was validated by another researcher to ensure that the questions were clear and straightforward to answer.

Although the questionnaire was anonymous, respondents could optionally provide contact details (email) if a follow-up clarification was needed.

4 **RESULTS AND ANALYSIS**

The survey’s results herein presented are organized as follows: Subsection 4.1 details the respondents’ demographics; subsection 4.2 presents the results related to the AGSD questions; subsections 4.3 to 4.6 present the results for research questions RQ1 to RQ4 respectively.

² XP 2014, the 15th International Conference on Agile Software Development and ICGSE 2014, the 9th International Conference on Global Software Engineering 2014
³ See http://www.linkedin.com
Table 3.1: Survey Categories

<table>
<thead>
<tr>
<th>Question ID</th>
<th>Scale Type</th>
<th>Question Type</th>
<th>Categories’ description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQ1</td>
<td>Nominal</td>
<td>Single answer</td>
<td>All the 206 states recognized by United Nations</td>
</tr>
<tr>
<td>SQ2</td>
<td>Ordinal</td>
<td>Single answer</td>
<td>Less than 1 year; More than 1 year, Less than 3 years; More than 3 years, less than 5 years; More than 5 years, less than 10 years; Above 10 years</td>
</tr>
<tr>
<td>SQ3</td>
<td>Nominal</td>
<td>Single answer</td>
<td>Program manager; Product owner; Scrum master; Team leader; QA manager; System analyst; Software architect; Designer; Developer; Tester; Other</td>
</tr>
<tr>
<td>SQ4</td>
<td>Ordinal</td>
<td>Single answer</td>
<td>1–5; 6–10; 11–15; 16–20; 21–50; 51–100; above 100</td>
</tr>
<tr>
<td>SQ5</td>
<td>Nominal</td>
<td>Multiple answers</td>
<td>Telecom/mobile applications; Business applications; Embedded systems; Safety critical systems; E-commerce applications; Financial applications; Data processing applications; Real time systems; System software; Other</td>
</tr>
<tr>
<td>SQ6</td>
<td>Nominal</td>
<td>Multiple answers</td>
<td>Extreme programing (XP); Scrum; Kanban; Feature driven development (FDD); “crystal”; Dynamic system development method (DSDM); Customized XP; Customized Scrum; Combination of XP and Scrum; Other</td>
</tr>
<tr>
<td>SQ7</td>
<td>Nominal</td>
<td>Single answer</td>
<td>insourcing; outsourcing</td>
</tr>
<tr>
<td>SQ8</td>
<td>Ordinal</td>
<td>Single answer</td>
<td>1; 2; 3; 4; Other</td>
</tr>
<tr>
<td>SQ9</td>
<td>Ordinal</td>
<td>Single answer</td>
<td>1; 2; 3; 4; Other</td>
</tr>
<tr>
<td>SQ10</td>
<td>Nominal</td>
<td>Single answer</td>
<td>Client; Vendor; Other</td>
</tr>
</tbody>
</table>
### 4.1 Demographics Questions

SQ1 captured the respondents’ work countries. A total of 12 different countries were represented in the sample. Their distribution by continents is presented in Table 3.2. Most of the respondents work in Europe or Asia, followed by North America. Such a data distribution complies with the countries/regions where AGSD has been reported more frequently to date [21, 22].

SQ2 measured the respondent’s industrial experience. The results show that 76.47% of the respondents have at least 5 years of experience in the software industry, with close to 51% with 10 years or more of experience (25.49% with “more than 5 years, less than 10 years” and 50.98 with “above 10 years”).

SQ3 captured the respondent’s job titles (roles). The results show a variety of roles, which also suggests that, at least amongst the respondents, effort estimation is not centralized within a small subset of roles. The results show that in most of the cases effort is estimated by either “developers” (27.45%) or “other” (21.57%) (“line manager”, “information security manager” and “agile coach”).

Considering the context of our survey (AGSD), it seems reasonable that evidence suggests developers are also responsible for estimating effort, given that within such context teams are cross-functional and most of the

<table>
<thead>
<tr>
<th>SQ11</th>
<th>Nominal</th>
<th>Multiple answers</th>
<th>Planning poker; Use case points; Analogy; Expert judgment; Delphi; COCOMO; Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQ12</td>
<td>Ordinal</td>
<td>Single answer</td>
<td>Underestimated by 50% or more; Underestimated by 25% or more; Spot on (0-5)% variation; Overestimated by 5% or more; Overestimated by 25% or more; Overestimated by 50% or more</td>
</tr>
<tr>
<td>SQ13</td>
<td>Nominal</td>
<td>Multiple answers</td>
<td>Communication model; Time zone difference; Time zone overlap; Cultural difference; Software process model; Language difference; Geographical distance; Team’s prior experience; Team’s expertise; Project domain; Non functional requirements; Number of sites; Customer communication; Story points; Use case points; Function points; Object points; Lines of codes; Task size; Other.</td>
</tr>
<tr>
<td>Continent</td>
<td>Countries</td>
<td>Percentage</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>------------------------------------------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td>Asia</td>
<td>India (13.73%), Pakistan (17.65%),</td>
<td>33.34%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Singapore (1.96%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oceania</td>
<td>Australia</td>
<td>7.84%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>United Kingdom (1.96%), Sweden (29.41%),</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Germany (3.92%), Netherlands (1.96%),</td>
<td>39.21%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Italy (1.96%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>North America</td>
<td>USA</td>
<td>15.69%</td>
<td></td>
</tr>
<tr>
<td>Central and South</td>
<td>Costa Rica (1.96%), Brazil (1.96%)</td>
<td>3.92%</td>
<td></td>
</tr>
<tr>
<td>America</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.2: Countries in which the respondents work most of the time
team members are called “developers” [123]. Note that the role “project manager” was not selected by any of the respondents. However, such a role is not necessarily required in agile teams [123].

SQ4 measured the respondents’ team size. Our results show that more than half of the respondents (56.86%) reported teams with up to 10 members. Such a size range complies with common agile best practices, where agile teams range from 5 to 9 members in general [123].

There is also a large percentage of respondents (33%) who reported team sizes with “more than 16” and even “more than 100 people”. Such numbers do not support the most commonly reported team sizes in the agile literature. However, it is possible to scale the team size up in ASD projects, in order to perform more resource-consuming software projects. Considering the context of this work (AGSD), in our view it is also reasonable to assume that a project carried out globally can demand more human resources and therefore larger teams, when compared to collocated agile projects.

SQ5 captured the software types developed by the respondents. Results show that “Telecom/Mobile” (52.94%) and “Business Applications” (56.86%) were the most frequent answers, followed by “Financial” (39.22%) and “E-Commerce” (37.25%) applications. For the free-text “Other” option (17.65%), “CAD” and “Health Care Applications” were provided most frequently.

4.2 AGSD Questions

SQ6 captured the agile methods employed by the respondents’ companies. Results show that the agile method selected most often was “Scrum” (84.31%), followed by “Kanban” [7] (37.25%) and “Extreme programming (XP)” [12] (23.53%). The free-text “Other” option (7.84%) revealed only one additional answer, model-driven development [130], which is not strictly an agile method. Since most respondents did not provide contact details (optional in the questionnaire), it was not possible to follow-up the “Other” answers for clarification.

Note that respondents could select several agile methods, which explains the high percentage for “Scrum”, in particular. However, it is important to highlight that the fact that “Scrum” was the most frequent answer is not surprising, since there is evidence that such method is the most employed in the software industry [142, 38, 69].

SQ7 asked for the sourcing strategies employed by the respondents’ companies. Results show that “insourcing” was the most frequent answer (47%), followed by “both” (35%) and “outsourced” (18%). This result cor-
robates our earlier results about the state of the art in effort estimation within the context of GSD [21, 22].

Using SQ8, we captured the average number of insourced sites. This question was only answered by respondents who answered “insourcing” or “both” in SQ7. The results show small differences between categories. The two answers with the highest percentages were “2” sites (26%) and “1” site (24%), followed by “4” sites (19%), “other” (17%), and “3” sites (14%). The valid free-text answers for option “Other” were “average of 5 insourced sites” and number of “insourced sites between 1 and 10”. The remaining 5 free-text answers were unclear or empty. Since contact information was not available for those respondents, a follow-up for clarification was not possible.

Using SQ9, we captured the average number of outsourced sites. This question was only answered by respondents who answered “Outsourcing” or “Both” in SQ7. The results for outsourced sites differ from those for insourced sites. The most frequent answer was “3” sites (33%), followed by “2” (30%) sites and “Other” (19%). All “Other” respondents provided unclear/empty free-text answers. Since contact information was not available for those respondents, a follow-up for clarification was not possible.

SQ10 captured the roles of the respondents’ companies in outsourced projects. Results show an even distribution between “Client” (41%) and “Vendor” (44%) companies. Details regarding the four “Other” answers (15%) were not available.

4.3 Research Question 1

RQ1 relates to the effort estimation techniques that are employed by AGSD practitioners. SQ6 was used to answer RQ1. Results show that “Planning poker” (72.55%) was the effort estimation technique selected most frequently, followed by “Expert judgment” (47.06%). “COSMIC” (which is actually a size metric), “Multiple expert judgment” and “Delphi-PERT approach” were the methods reported using the “Other” option (15.69%).

Although “Planning poker” and “Expert judgment” are also frequently used in agile software development [28], it is interesting to note that practitioners did not add any additional technique as a result of a context change from collocated to global software development. Note that more than 40% of the respondents work in teams larger than 10 people, which could perhaps bring some challenges when relying solely on techniques such as “Planning poker” and “Expert judgment”.

Also note that the number of respondents using a single effort estimation technique (51%) was quite similar to the number of respondents using more than one technique (49%).
Table 3.3 shows small differences between the two ways in which “Planning poker” was reported: as a single technique (48.65%), or together with other techniques (51.35%). The other techniques were mostly reported together with other options. COCOMO and Delphi were reported only together with other techniques.

Table 3.3: Percentages of the selected effort estimation techniques considering the way they were reported

<table>
<thead>
<tr>
<th>Technique</th>
<th>Single</th>
<th>Multiple</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning Poker</td>
<td>48.65%</td>
<td>51.35%</td>
</tr>
<tr>
<td>Expert Judgment</td>
<td>20.83%</td>
<td>79.17%</td>
</tr>
<tr>
<td>Analogy</td>
<td>6.67%</td>
<td>93.33%</td>
</tr>
<tr>
<td>Use Case Points</td>
<td>0.0%</td>
<td>100%</td>
</tr>
<tr>
<td>Delphi</td>
<td>0.0%</td>
<td>100%</td>
</tr>
<tr>
<td>COCOMO</td>
<td>0.0%</td>
<td>100%</td>
</tr>
</tbody>
</table>

The fact that a respondent selected more than one effort estimation technique can be interpreted in two different ways. Either the effort estimation techniques are used together in the context of a given project, or individual effort estimation techniques are singly employed in different projects. As part of our future work, we plan to investigate further this issue via interviews with the survey participants who selected more than one technique and who provided their contact details.

4.4 Research Question 2

RQ2 was answered by combining the data gathered from SQ5 and SQ6. Results are displayed in Table 3.4 and show that “planning poker” was the most selected effort estimation technique, regardless of the sourcing strategy used (insourcing 43.24%; outsourcing 35.71%; Both 36.84%).

The second most reported effort estimation technique changes according to the applied sourcing strategy. When “insourcing” was the employed sourcing strategy, “expert judgment” was as reported as “planning poker” (43.24%). On the other hand, when “Outsourcing” was the employed sourcing strategy, “analogy” was the second most reported technique (28.57%).
Table 3.4: Selected technique by sourcing strategy

<table>
<thead>
<tr>
<th>Technique</th>
<th>Insourcing</th>
<th>Outsourcing</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning Poker</td>
<td>43.24%</td>
<td>35.71%</td>
<td>36.84%</td>
</tr>
<tr>
<td>Expert Judgment</td>
<td>43.24%</td>
<td>21.42%</td>
<td>31.57%</td>
</tr>
<tr>
<td>Analogy</td>
<td>18.91%</td>
<td>28.57%</td>
<td>13.15%</td>
</tr>
<tr>
<td>Use Case Points</td>
<td>10.81%</td>
<td>7.14%</td>
<td>7.89%</td>
</tr>
<tr>
<td>Delphi</td>
<td>2.70%</td>
<td>7.14%</td>
<td>5.26%</td>
</tr>
<tr>
<td>COCOMO</td>
<td>0.0%</td>
<td>0.0%</td>
<td>5.26%</td>
</tr>
</tbody>
</table>

4.5 Research Question 3

RQ3 relates to the predictors (size metrics/cost drivers) that have been considered by practitioners in AGSD. Survey question SQ7 was used to gather data to answer RQ3.

Results show that “Team’s skill level” (71%) was the most selected cost driver, followed by “communication model” (65%), “time zone difference” (59%), “team’s prior experience” (55%), “cultural difference” (45%) and “software process model” (41%). “technical dependencies”, “difference in work hours between sites” and “uncertainty level” were reported by means of the option “Other” (14%). The results are in-line with the state of the art [21, 22].

Results also show that “task size” (67%) was the most selected size metric, followed by “story points” (63%) and “use case points” (27.45%). “UML points” and “COSMIC function points” were reported by means of the option “Other” (7.8%). Size metrics that are popular in plan-driven software development, e.g. “Function points” (20%) and “Lines of code” (8%), were selected by fewer respondents.

4.6 Research Question 4

RQ4 relates to the challenges that impact the accuracy of the effort estimates in AGSD. Survey questions SQ12 and SQ14 gathered data to answer RQ4.

SQ12 asked for the average accuracy of the effort estimates, when compared to the actual effort. The results show that “underestimated by 25% or more” was the most common answer (33.33%), followed by “spot on (0–5)% variation” (19.61%) and “underestimated by 5% or more” (17.65%). In 54.90% of the cases the effort estimates were underestimated (see Total in Table 3.5).
Considering the accuracy values reported in the effort estimation literature, the results suggest that the survey respondents are estimating effort with an acceptable level of accuracy [30]. However, it was puzzling to see close to 20% of the respondents reporting that estimates were “spot-on”. We hypothesize that perhaps such respondents work with fixed budget contracts, or projects where effort is estimated and later adjusted to comply with actuals.

Table 3.5: Distribution of the responses by type of error

<table>
<thead>
<tr>
<th>Range</th>
<th>Underestimated</th>
<th>Overestimated</th>
<th>Spot on</th>
</tr>
</thead>
<tbody>
<tr>
<td>By 5% or more</td>
<td>17.65%</td>
<td>9.8%</td>
<td>-</td>
</tr>
<tr>
<td>By 25% or more</td>
<td>33.33%</td>
<td>11.76%</td>
<td>-</td>
</tr>
<tr>
<td>By 50% or more</td>
<td>3.92%</td>
<td>3.92%</td>
<td>-</td>
</tr>
<tr>
<td>Between 0% and 5%</td>
<td>-</td>
<td>-</td>
<td>19.61%</td>
</tr>
<tr>
<td>Total</td>
<td>54.90%</td>
<td>25.48%</td>
<td>19.61%</td>
</tr>
</tbody>
</table>

SQ14 asked for the challenges that impact the accuracy of the effort estimates in the AGSD context. Only 24 respondents (47%) answered this optional open-ended question. The provided answers were categorized as follows:

- **Distribution-related challenges** - Thirteen respondents (25.5%) reported that time, language and cultural differences between sites and the distance to the client can affect the accuracy of the effort estimates. In their opinion, such factors affect the communication between sites and this communication overhead is not accounted when estimating effort in the AGSD context.

- **Requirements-related challenges** - Twenty respondents (39.2%) informed that unclear, unstable and mis-documented requirements affect the accuracy of the effort estimates. In their opinion, those challenges lead to unpredicted activities during the development process, increasing the real effort of software projects.

- **Team-related challenges** - Fourteen respondents (27.5%) reported that the lack of domain knowledge, lack of knowledge on the required technologies, lack of experience on the used effort estimation technique and lack of team cohesion (teams with people that never worked or worked little time together) lead to wrong assumptions, which affect the accuracy of the effort estimates. Another challenge presented by the respondents was the fact that in many cases the effort is not estimated by the same team responsible for constructing the software, which also leads to wrong assumptions, jeopardizing the accuracy of the effort estimates.
• **Client-related challenges** - Two respondents (4%) reported that poor participation and mis-prioritization of the tasks by the clients affect the accuracy of the effort estimates. It means that the effort is estimated considering an active participation of the clients, however development teams find a different scenario when a project starts, which in general leads to bigger effort.

5 Discussion

The results found in this paper are discussed as follows (organized by research questions).

5.1 **RQ1 - Which techniques have been used to estimate effort in AGSD?**

The survey results suggest that the same techniques used in the collocated context have been used in the AGSD context, without any adaptations. Considering that there are few studies about effort estimation in GSD [21] and AGSD [22], the absence of customized effort estimation techniques is not surprising. Nevertheless, this does not mean that such customization would not be important to improve estimation accuracy and even the prediction process. Thus, we believe that more research should be carried out to investigate this issue further.

5.2 **RQ2 - What is the impact of the applied sourcing strategy on the selection of an effort estimation method in AGSD?**

In relation to the impact of the sourcing strategy on the selection of effort estimation techniques, the results suggest the following:

- The applied sourcing strategy has no or small influence on the selection of the technique when just one technique is selected to estimate effort. The same finding appears to be true regarding the selection of the first technique when two or more techniques are selected to estimate effort in AGSD projects. In both cases, “planning poker” was the most frequently reported method.
- The applied sourcing strategy appears to have a bearing on the selection of additional methods, i.e. when two or more methods are employed to estimate effort. In such situations, “expert judgment” appears to be related to an “insourcing” strategy and estimation by “analogy” appears to be related to an “outsourcing” strategy.

We suspect that estimation by “analogy” is the second most frequently selected technique because a company would not necessarily have access to the internal information of outsourced sites. Rather, the client company
would be obliged to compare a given new project with similar past finished projects. This could explain the choice of estimation technique. We also aim to investigate further this issue in collaboration with some of the respondents who answered this survey and provided their contact details.

5.3  RQ3 - Which effort predictors have been used to estimate effort in AGSD?

The results from this survey suggest that most of the well-known GSD challenges (e.g. “communication” and “cultural difference” [59]) have been considered as cost drivers. However, “team’s skill level”, which is a relevant cost driver in both global and collocated context, was the most frequently reported cost driver.

The survey did not ask more detailed questions on how the reported cost drivers were measured by the respondents and used in the effort estimation process. Further work will be carried out to obtain such data.

Regarding the reported size metrics, the results suggest that “task size” and “story points” are the size metric that fit properly for estimating effort in AGSD. Not surprisingly, they are the most frequent size metric in agile software development literature [123, 12].

5.4  RQ4 - What are the challenges that impact the accuracy of the effort estimates in AGSD?

A considerable number of the respondents reported reasonable effort estimates in AGSD (Table 3.5). The effort estimates’ accuracy reported in the literature on effort estimation is slightly worse [76, 113]. Maybe the respondents were not comfortable to report actual accuracy, even when answering an anonymous questionnaire. Maybe the answers are legitimate. Some of our future work aims to investigate further this issue.

According to the participants of our survey, there are many different challenges that impact the accuracy of the effort estimates in the AGSD context. Most of the respondents reported that the main problem lays on the requirements, which is also a challenge in collocated agile software projects.

The participants of the survey also raised the point that the communication overhead, which is bigger in globally distributed projects, is not properly accounted most of the time by practitioners. It affects the accuracy of the effort estimates, leading to underestimated effort. The impact reported for this challenge is in-line with the accuracy values reported in the survey (See Subsection 4.6). We believe that further research should be conducted to identify or develop approaches to calculate the impact of the communication overhead on the overall effort of AGSE projects.
In this section, we describe the threats to validity of this work.

- **Construct validity** is concerned with the relation between the theories behind the research and the observations [153]. This type of threat was mitigated by collecting data from a wide range of respondents, who worked in many different countries and companies.

- **Conclusion validity** is concerned with the possibility of incorrect conclusions about relationships in the observations that could arise from errors in the data sources [153]. To mitigate such kind of threat, the survey was validated by other researchers to ensure that the questions were clear and straightforward to answer. Besides, we made clear for the respondents that just people who had worked with effort estimation should answer the questionnaire. Nonetheless, we cannot claim that all questions were well understood by the respondents.

- **Internal validity** is related to issues that may affect the causal relationship between treatment and outcome [153]. To mitigate this threat, the survey was validated by other researchers. Considering that the questionnaire was to be answered just once per respondent, we believe that the possibility of learning effect was removed. Finally, no fatigue effect was expected to affect the respondents, since the time to answer the questionnaire was short (15 minutes).

- **External validity** is concerned with the ability to generalize the findings of a given study beyond its study context [153]. In general, research in empirical software engineering is hard to generalize, because the research outcomes are highly influenced by the research context. In order to improve the representativeness of the findings of this work, we collected data from many respondents, widening the number of contexts for which the findings of this work could be generalized.

7 CONCLUSIONS

This paper presented the results of a survey on effort estimation in the context of Agile Global Software Development. The main goal of this paper was to complement the already existent evidence obtained from the state of the art by obtaining evidence from the state of the practice.

An on-line questionnaire was made accessible from August 12 of 2014 to October 10 of 2014. Respondents from 12 different countries have participated in the survey. A self-selected sample of 51 respondents have fully answered the questionnaire.
The reported effort estimation techniques are well-known in the collocated context. This result was not surprising once it was already known that there have been few studies on effort estimation in AGSD context.

It was surprising the fact that most of the respondents reported accurate effort estimates. They also reported many challenges that impact the accuracy of the effort estimates, such as unclear, unstable and mis-documented requirements. In addition, they reported that it is common to underestimate the impact of the communication overhead between the distributed sites of AGSE projects, which leads to underestimated effort.

Regarding the impact of the sourcing strategies on the selection of effort estimation techniques, the results suggest that when just one method is applied, the impact of the applied sourcing strategy appears to be insignificant. However, when more than one method is applied, the selection of the second method appears to be affected by the sourcing strategy (i.e. “expert judgment” related to “offshore insourcing” and “estimation by analogy” related to “offshore outsourcing”).

Finally, about the predictors, the results suggest that the main GSD challenges have been considered by the practitioners, along with other factors that are also relevant in any other context, such as the experience of the team.

We intend to perform further research in order to confirm and/or clarify some of our findings, such as the accuracy level of the reported effort estimates. Likewise, we intend to understand how the practitioners have been measured and incorporating the considered cost drivers in their effort estimation processes. Finally, additional research should be conducted to identify or develop approaches to calculate the impact of the communication overhead on the overall effort of AGSE projects. We believe that the listed further research could be carried out by means of case studies, which is one of the directions for our future work.
1 INTRODUCTION

In science and engineering, a systematic description and organization of the investigated subjects helps to advance the knowledge in this field [141]. This organization can be achieved through the classification of the existing knowledge. Knowledge classification has supported the maturation of different knowledge fields mainly in four ways [141]:

- Classification of the objects of a knowledge field provides a common terminology, which eases the sharing of knowledge.
- Classification can provide a better understanding of the interrelationships between the objects of a knowledge field.
- Classification can help to identify gaps in a knowledge field.
- Classification can support decision making processes.

Summarizing, classification can support researchers and practitioners in generalizing, communicating and applying the findings of a knowledge field [143].

Software Engineering (SE) is a comprehensive and diverse knowledge field that embraces a myriad of different research subareas. The knowledge within many subareas is already classified, in particular by means of taxonomies [65, 66, 19, 149, 138]. According to the Oxford English Dictionary [37], a taxonomy is “a scheme of classification”. A taxonomy allows for the description of terms and their relationships in the context of a knowledge area. The concept of taxonomy was originally proposed by Carolus Linnaeus [84] to group and classify organisms by using a fixed number of hierarchical levels. Nowadays, different classification structures (e.g. hierarchy, tree and faceted analysis [79]) have been used to construct taxonomies in different knowledge fields, such as Education [16], Psychology [97] and Computer Science [120].

Taxonomies have contributed to mature the SE knowledge field. Nevertheless, likewise the taxonomy proposed by Carolus Linnaeus that keeps
being extended [137], SE taxonomies are expected to evolve over time incor-
porting new knowledge. In addition, due to the wide spectrum of SE
knowledge, there is still a need to classify the knowledge in many SE sub-
areas.

Although many SE taxonomies have been proposed in the literature, it
appears that taxonomies have been designed or evolved without following
particular patterns, guidelines or processes. A better understanding of how
taxonomies have been designed and applied in SE could be very useful for
the development of new taxonomies and the evolution of existing ones.

To the best of our knowledge, no systematic mapping or systematic liter-
ature review has been conducted to identify and analyze the state-of-the-
art of taxonomies in SE. In this paper, we describe a systematic mapping
study [74, 106] aiming to:

- Characterize the state-of-the-art research on SE taxonomies.
- Identify practices to support the design of SE taxonomies.

The remainder of this paper is organized as follows: Section 2 describes
related background. Section 3 discusses two SE taxonomies. Section 4
presents the employed research methodology, followed by the presentation
and discussion of the current state-of-the-art on taxonomies in SE (Section
5). In Section 6, we present a revised method to design SE taxonomies. Va-
lidity threats are discussed in Section 7 and finally conclusions and future
work are given in Section 8.

2 BACKGROUND

In this section, we discuss important aspects related to taxonomy design
that serve as motivation for the research questions described in Section 4.

2.1 Taxonomy definition and purpose

Taxonomy is neither a trivial nor a commonly used term. According to
the most cited English dictionaries, a taxonomy is mainly a classification
mechanism:

- The Cambridge dictionary\(^1\) defines taxonomy as “a system for naming
  and organizing things, especially plants and animals, into groups that share
  similar qualities”.
- The Merriam-Webster dictionary\(^2\) defines taxonomy as “Orderly classi-
  fication of plants and animals according to their presumed natural relation-
  ships”.

\(^1\) www.dictionary.cambridge.org
\(^2\) www.merriam-webster.com
• The Oxford dictionaries\(^3\) define taxonomy as “The classification of something, especially organisms” or “A scheme of classification”.

Since taxonomy is mainly defined as a classification system, one of the main purposes to develop a taxonomy should be to classify something.

2.2 Subject matter

The first step in the design of a new taxonomy is to clearly define the units of classification. In software engineering this could be requirements, design patterns, architectural views, methods and techniques, defects etc. This requires a thorough understanding of the subject matter to be able to define clear taxonomy classes or categories that are commonly accepted within the field \([50, 150]\).

2.3 Descriptive bases / terminology

Once the subject matter is clearly defined or an existing definition is adopted, the descriptive terms, which can be used to describe and differentiate subject matter instances, must also be specified. An appropriate description of these bases for classification is important to perform the comparison of subject matter instances. Descriptive bases can also be viewed as a set of attributes that can be used for the classification of the subject matter instances \([50, 150]\).

2.4 Classification procedure

Classification procedures define how subject matter instances (e.g. defects) are systematically assigned to classes or categories. Taxonomy’s purpose, descriptive bases and classification procedures are related and dependent on each other. Depending upon the measurement system used, the classification procedure can be qualitative or quantitative. **Qualitative** classification procedures are based on nominal scales. In the qualitative classification systems, the relationship between the classes can not be determined. **Quantitative** classification procedures, on the other hand, are based on numerical scales \([150]\).

2.5 Classification structure

As aforementioned, a taxonomy is mainly a classification mechanism. According to Kwasnik \([79]\), there are four main approaches to structure a

---

3 www.oxforddictionaries.com
classification scheme (classification structures): hierarchy, tree, paradigm and faceted analysis.

**Hierarchy** leads to taxonomies with a single top class that "includes" all sub- and sub-sub classes, i.e. a hierarchical relationship with inheritance ("is-a" relationship). Consider, for example, the hierarchy of students in an institution wherein the top class "student" has two sub-classes of "graduate student" and "undergraduate student". These sub-classes can further have sub-sub classes and so forth. A true hierarchy ensures the mutual exclusivity property, i.e an entity can only belong to one class. Mutual exclusivity makes hierarchies easy to represent and understand; however, it cannot represent multiple inheritance relationships though. Hierarchy is also not suitable in situations when researchers have to include multiple and diverse criteria for differentiation. To define a hierarchical classification, it is mandatory to have good knowledge on the subject matter to be classified; the classes and differentiating criteria between classes must be well defined early on.

**Tree** is similar to the hierarchy, however, there is no inheritance relationship between the classes of tree-based taxonomies. In this kind of classification structure, common types of relationships between classes are "part-whole", "cause-effect" and "process-product". For example, a tree representing a whole-part relationship between a country, its provinces and cities. Tree and hierarchy share similar strengths and limitations.

**Paradigm** leads to taxonomies with two-way hierarchical relationships between classes. The classes are described by a combination of two attributes at a time. For example, paradigm would be suitable if we have to also represent gender in the "student" hierarchy example above-mentioned. It can also be viewed as a two-dimensional matrix whose vertical and horizontal axes allow for the inclusion of two attributes of interest. This type of classification structure shares similar strengths and limitations with the hierarchy structure.

**Faceted analysis** leads to taxonomies whose subject matters are classified along different perspectives (facets). Each facet is independent and has its own classes, which enable facet-based taxonomies to be easily adapted so they can evolve smoothly over time. In order to properly classify CASE tools, for example, multiple facets need to be considered. These facets may include supported platform(s), license type, SE area, web support etc. In faceted classifications it is possible to employ the aforementioned classification structures to design each facet. Faceted analysis is suitable for new and evolving fields, since it is not required to have the complete knowledge related to the selected subject matter to design a facet-based taxonomy. However, it can be challenging to define an initial set of facets. In addition, although it is possible to define relationship between facets, in most cases the facets are independent and have no meaningful relationship between each other.
2.6 Validation

To improve reliability, it is important to validate a taxonomy proposed. Validation can be done using different criteria, which can be theoretical (e.g. mutual exclusiveness and exhaustiveness) or practical (e.g. efficiency and utility of the classification) [150]. The validation criteria can be applied to validate a taxonomy using different approaches, e.g. illustration of use, case study or experiment. These approaches offer different degrees of rigor. If the objective is to thoroughly validate a taxonomy, a case study or an experiment may be the appropriate choice. However, if the objective is just to demonstrate the usefulness of a taxonomy, a simple illustration may be enough.

3 Taxonomies in Software Engineering

In this section, we present two SE taxonomies to exemplify how taxonomies can support knowledge classification in SE.

3.1 Smite et al.’s GSE taxonomy [149]

Global Software Engineering (GSE) is a relatively new field where terminology is still used inconsistently. In many cases one term has more than one meaning or several terms are used to refer to the same phenomenon. To address this problem, Smite et al. proposed a taxonomy to unify the terminology and classify the knowledge related to GSE sourcing strategies. Smite et al. conducted a Delphi-inspired study with GSE researchers to develop their taxonomy. This taxonomy was developed to classify the relationship between pairs of sites, although it is equally possible to describe more complex GSE projects, with more than two sites. To illustrate the usage of the proposed GSE taxonomy, Smite et al. applied their taxonomy to classify the sourcing strategies from 68 different studies.


There are many critical factors that determine the success or failure of software process deployment initiatives in an organization. Based on a review of relevant literature, industry feedback and experts’ knowledge, Bayona-Oré et al. propose a taxonomy for critical success factors for software process deployment. The taxonomy provides a common language for deployment initiatives, and is designed to facilitate organizations in identifying relevant critical factors that impact their deployment process. The work also presents a method for designing taxonomies based on a number
of proposals in different areas. This method is discussed further in Section 6.

4 RESEARCH METHODOLOGY

We chose the systematic mapping study method (SMS) to identify and analyze the state-of-the-art towards taxonomies in SE, because this method works well for broad and weakly defined research areas [74, 106]. We employed the guidelines by Kitchenham and Charters [74] and partly implemented the mapping process provided by Petersen et al. [106]. The employed mapping process is summarized in Figure 4.1 and described further in Subsections 4.1–4.5.

Figure 4.1: Employed systematic mapping process.

4.1 Research questions

The following research questions were formulated to guide this SMS:

- **Question 1 (RQ1)** – What taxonomy definitions and purposes are provided by publications on SE taxonomies?
- **Question 2 (RQ2)** – Which subject matters are classified in SE taxonomies?
- **Question 3 (RQ3)** – How are SE taxonomies validated?
- **Question 4 (RQ4)** – How are SE taxonomies structured?
- **Question 5 (RQ5)** – To what extent are SE taxonomies used?

The main idea behind RQ1 is to identify how and why the term “taxonomy” is used in primary studies that claim to present a taxonomy. RQ2 focuses on identifying the subject matters classified by means of taxonomies in SE. RQ3 focuses on identifying the approaches used to validate taxonomies. With RQ4 we intend to identify the classification structures, related descriptive bases and classification procedures employed to design SE taxonomies. Finally, RQ5 focuses on the extent to which proposed SE taxonomies are used.
4.2 Search process

The search process employed in this work is displayed in Figure 4.2 and has 6 activities.

First, we defined the terms to be included in our search string. We selected all SWEBOK knowledge areas [19] to be included as terms, except for the three knowledge areas on related disciplines (Computing Foundations, Mathematical Foundations and Engineering Foundations). We also included the term “Software Engineering”, to augment the comprehensiveness of the search string. Finally, to reduce the scope of the search string to studies that report SE taxonomies, we included the term “taxonomy”.

Since some of the knowledge areas are referred by the SE community through of other terms (synonyms), we also included their synonyms. Specifically, the following synonyms were included into the search string:

- **Requirements** – requirements engineering.
- **Construction** – software development.
- **Design** – software architecture.
- **Management** – software project management, software management.
- **Process** – software process, software life cycle.
- **Models and methods** – software model, software methods.
- **Economics** – software economics.

The selected SWEBOK knowledge areas and the term “Software Engineering” were all linked using the operator OR. The term “taxonomy” was linked with the other terms using the operator AND. The final search string is shown below.
Although SE knowledge classification could be named in different ways, e.g. taxonomy, ontology [144] and classification scheme [141], we limited the scope of this paper to taxonomies. Extending our search string to include the terms “ontology” and “classification scheme” would have led to an excessive number of search results that would have been infeasible to handle4.

Once the search string was designed, we selected the primary sources to search for relevant studies. Scopus5, Compendex/Inspec6 and Web of Science7 were selected because they cover most of the important SE databases, such as IEEE, Springer, ACM and Elsevier. In addition, the selected primary sources are able to handle advanced queries. The search string was applied on meta data (i.e. title, abstract and author keywords) in August 2014 on the selected data sources. Table 4.1 presents the number of search results for each data source.

Table 4.1: Summary of search results

<table>
<thead>
<tr>
<th>Database/Search Engine</th>
<th>Search Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scopus</td>
<td>833</td>
</tr>
<tr>
<td>Compendex and Inspec</td>
<td>611</td>
</tr>
<tr>
<td>Web of Science</td>
<td>287</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1731</strong></td>
</tr>
<tr>
<td><strong>Total without duplicates</strong></td>
<td><strong>1371</strong></td>
</tr>
</tbody>
</table>

---

4 Inclusion of the terms “ontology”, “ontologies” and “classification scheme” returned 9603 “hits” in total for Scopus, Compendex/Inspec and Web of Science.
5 www.scopus.com
6 www.engineeringvillage.com
7 apps.webofknowledge.com
4.3 Study selection process

The selection process employed in this work is displayed in Figure 4.3 and detailed as follows.

First, the following inclusion and exclusion criteria were defined:

- **Inclusion criteria**
  1. Studies that propose or extend a taxonomy AND
  2. Studies that are within Software Engineering (SE), according to SWEBOK’s KAs (see Subsection 4.2).

- **Exclusion criteria**
  1. Studies where the full-text is not accessible OR;
  2. Studies that do not propose or extend a SE taxonomy OR;
  3. Studies that are not written in English OR;
  4. Studies that are not reported in a peer-reviewed workshop, conference, or journal.

The selection of primary studies was conducted using a two-stage screening procedure. In the first stage, only the abstracts and titles of the studies were considered. In the second stage, the full texts were read. Note that we used in both stages an inclusive approach to avoid premature exclusion of studies, i.e. if there was doubt about a study, such a study was to be included.

For the first stage (level-1 screening), the total number of 1371 studies were equally divided between the two first authors. As a result, 471 studies were judged as potentially relevant.
Table 4.2: Rationale for excluded studies

<table>
<thead>
<tr>
<th>Reason</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not proposing or evolving a SE taxonomy</td>
<td>1048</td>
</tr>
<tr>
<td>Duplicate</td>
<td>360</td>
</tr>
<tr>
<td>Full-text not accessible</td>
<td>49</td>
</tr>
<tr>
<td>Non-peer reviewed</td>
<td>21</td>
</tr>
<tr>
<td>Not written in English</td>
<td>9</td>
</tr>
<tr>
<td>Total retrieved</td>
<td>1731</td>
</tr>
<tr>
<td>Total excluded</td>
<td>1487</td>
</tr>
<tr>
<td>Total included after study selection</td>
<td>254</td>
</tr>
<tr>
<td>Total included after data extraction</td>
<td>244</td>
</tr>
</tbody>
</table>

To increase the reliability of the level-1 screening result, the fourth author screened a random sample of 11.37% (78 studies) from the studies screened by the first author and the third author screened a random sample of 11.38% (78 studies) from the studies screened by the second author. It means that 22.75% of the 1371 studies were screened by two different reviewers. The first and fourth authors had the same judgment for 91% (71) of the studies. The second and third authors had the same judgment for 93.6% (73) of the studies.

To evaluate the reliability of the inter-rate agreement between the authors, we calculated the Cohen’s kappa coefficient [45]. The Cohen’s kappa coefficient between the first and fourth authors was statistically significant (significance level = 0.05) and equal to 0.801. The Cohen’s kappa coefficient between the second and third authors was also statistically significant (significance level = 0.05) and equal to 0.857. According to Fleiss et al. [45], Cohen’s kappa coefficient values above 0.75 mean excellent level of agreement.

The level-2 screening (second stage), performed by the first and second authors, consisted on applying the selection criteria on the full-text of the studies selected during the level-1 screening. The total number of 471 studies were equally divided between the first two authors. As a result, 254 studies were judged as relevant.

To increase the reliability of the level-2 screening, a two-step validation was performed, as follows:
1. The first author screened 55% (70) of the studies deemed as relevant by the second author during the level-2 screening (randomly selected) and vice-versa. No disagreements were found between the authors.
2. Seven studies were randomly selected from each of the two sets allocated to the first two authors for further validation. The third author applied the study selection process on these 14 studies (about 5% of 254) for validation purposes. No disagreements were found with respect to the study selection (i.e. include/exclude) decisions.

During the entire screening process (stages 1 and 2), we tracked the reason for each exclusion, as presented in Table 4.2.

4.4 Extraction process

The extraction process employed in this work is summarized in Figure 4.4 and consists of four main steps: Define a classification scheme, define an extraction form, extract data, and validate the extracted data.

The classification scheme was designed following the guidelines by Petersen [106]. It has the following facets:

- **Research type** – This facet is used to distinguish between different types of studies (adapted from Wieringa et al. [151]).
  - *Evaluation research* – A study that reports a taxonomy implemented in practice, i.e. evaluation in a real environment, in general by means of the case study method.
Validation research – A study that reports a taxonomy that was not implemented in practice yet, although it was validated in laboratory environment, in general by means of experiment.

Solution proposal – A study that reports a taxonomy that was neither implemented in practice nor validated although it is supported by a small example (illustration) or a good line of argumentation.

Philosophical paper – A study that reports a new taxonomy that has no type of evaluation, validation or illustration.

- SE knowledge area – This facet is used to distinguish between the SE knowledge areas in which taxonomies have been proposed. The categories of this facet follow the SWEBOK [19]: software requirements, software design, software construction, software testing, software maintenance, software configuration management, software engineering management, software engineering process, software engineering models and methods, software quality, software engineering professional practice and software engineering economics.

- Presentation approach – This facet is used to classify the studies according to the overall approach used to present a taxonomy: textual and graphical, respectively.

For the data extraction, the relevant studies (254) were equally divided between the first and second authors. For each paper, data was collected in a spreadsheet using the data extraction form shown in Table 4.3.

To increase the reliability of the extracted data, a two-step validation was performed, as follows:

1. The first author independently re-extracted the data of 55% (70) of the studies originally extracted by the second author (randomly selected) and vice-versa. Five disagreements were identified and all of them were related to the item “classification structure”.

2. Fourteen studies were randomly selected from the studies originally extracted by the first and second authors (7 studies from each author). Those studies were independently re-extracted by the third author. Twenty one disagreements were identified; 2 on the “taxonomy purpose”, 9 on “classification structure”, 2 on “classification procedure type”, 3 on “classification procedure description” and 5 on “validation approach”.

All disagreements except for “classification structure” were easily resolved. We believe that the high level of disagreement on the item “classification structure” was due to the fact that none of the studies explicitly stated and motivated the employed classification structure, which demanded the inference of such data from the text in each paper.
<table>
<thead>
<tr>
<th>Data item(s)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citation data</td>
<td>Title, author(s), year and publication venue</td>
</tr>
<tr>
<td>Taxonomy definition</td>
<td>Definition of taxonomy that is used or referred to</td>
</tr>
<tr>
<td>Purpose</td>
<td>Text that states the purpose for the taxonomy</td>
</tr>
<tr>
<td>Purpose keyword</td>
<td>Key word used in the paper to describe the purpose (e.g. classify, understand, describe)</td>
</tr>
<tr>
<td>Subject matter</td>
<td>The name of the thing/concept are that is taxonomized</td>
</tr>
<tr>
<td>Descriptive bases</td>
<td>Is the subject matter defined in sufficient detail/clarity to enable classification (Yes/No)</td>
</tr>
<tr>
<td>Classification structure</td>
<td>Hierarchy, tree, paradigm, or faceted analysis, according to Kwasnik [79]</td>
</tr>
<tr>
<td>Classification procedure</td>
<td>The criteria for putting items in different classes (qualitative, quantitative or no details provided)</td>
</tr>
<tr>
<td>Classification procedure</td>
<td>Do the authors explicitly describe the classification procedure (Yes/No)</td>
</tr>
<tr>
<td>Presentation approach</td>
<td>Textual or graphical</td>
</tr>
<tr>
<td>Validation approach</td>
<td>Is the taxonomy validated/tested? If yes, how (e.g. illustration, case study, experiment)?</td>
</tr>
<tr>
<td>Primary knowledge area</td>
<td>Primary knowledge area as per SWEBOK v3 [19]</td>
</tr>
<tr>
<td>Secondary knowledge area</td>
<td>Secondary knowledge area as per SWEBOK v3 (if applicable)</td>
</tr>
<tr>
<td>Number of citations</td>
<td>Number of times a primary study is cited by other studies, as per Google Scholar</td>
</tr>
</tbody>
</table>
To improve the reliability of the extracted data, we decided to re-screen all 254 papers, focusing only on the item “classification structure”. In this process 52 papers were re-screened together by the three first authors and 202 were re-screened together only by the first and second authors.

First, we discussed classification structures in detail (based on Kwasnik [79]) to come to a common understanding of the terms. Thus, three of us did an independent re-assessment of the classification structure of 52 papers. As a result, we reached full agreement on 50 papers (3 identical results) and partial agreement on 2 papers (2 reviewers agreeing). There were no primary studies without full or partial agreement.

After a discussion of the results, the remaining 202 studies were re-assessed by the first and second authors. As a result, the first and second authors reached agreement on 190 papers. The remaining 12 papers were independently re-assessed by the third author, who did not know the results from the other two reviewers. In the end, full agreement was achieved for 50 studies and partial agreement was achieved for 204 studies.

During the re-assessment of the primary studies, 10 studies were excluded because they do not present taxonomies, reducing the final number of primary studies to 244 (see Table 4.2).

4.5 Analysis process

Figure 4.5 presents the analysis process conducted herein. First, we classified the extracted data using the scheme defined in Subsection 4.4. This led to the results detailed in Section 5. We also performed a quantitative analysis of the extracted data to answer the research questions of this paper. Finally, the overall result of the data analysis (see Section 5), along with information from additional literature ([150, 50, 79]), was used to revise an existing method previously proposed to design SE taxonomies [11], as detailed in Section 6.

5 Results and discussion

In this section, we describe the results of the mapping study reported herein, which are based on the data extracted from 244 papers reporting 245 taxonomies (one paper presented two taxonomies). The percentages in Sections 5.1 and 5.7 reflect the number of papers (244). The percentages in all other subsections reflect the number of taxonomies (245).

8 The full list of the included 244 primary studies is available at drive.google.com/file/d/0B2kvKPmJJREDX3FYdzFTdzdTRms/view?usp=sharing.
Figure 4.5: Analysis process.

5.1 General results

Figure 4.6 shows that SE taxonomies have been proposed continuously since 1987, with an increasing number of these published after the year 2000, which suggests higher interest on this research topic.

Table 4.4 displays that 53.7% (131) of the studies were published in relevant conferences in maintenance (International Conference on Software Maintenance), requirements engineering (Requirements’ Engineering Conference) or general SE topics (e.g. International Conference on Software Engineering). Taxonomies were published at 100 unique conferences with 74 featuring only a single SE taxonomy publication. These results further indicate a broad interest in SE taxonomies in a wide range of SE knowledge areas.

Table 4.4 also shows that 33.2% (81) of the primary studies were published as journal articles in 27 unique journals. Taxonomies have been published frequently in relevant SE journals (e.g. IEEE Transactions on Software Engineering and Information and Software Technology). The scope of these journals is not confined to a specific SE knowledge area.

Primary studies were published also in 26 unique workshops (32 – 13.1%). As for journals and conferences, the results indicate an increasing interest in SE taxonomies in a broad range of SE knowledge areas.

Figures 4.7a–h depict the yearly distribution of SE taxonomies by knowledge area for the knowledge areas with 10 or more taxonomies. Note that most knowledge areas follow an increasing trend after 2000, with many taxonomies for construction, design, and quality in the 1980s and 1990s.
Figure 4.6: Year and venue wise distributions.
Table 4.4: Publication venues with more than two taxonomy papers.

<table>
<thead>
<tr>
<th>Publication venue</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>IEEE Intl. Conference on Software Maintenance (ICSM)</td>
<td>9</td>
</tr>
<tr>
<td>Intl. Conference on Requirements Engineering (RE)</td>
<td>5</td>
</tr>
<tr>
<td>Intl. Conference on Software Engineering (ICSE)</td>
<td>5</td>
</tr>
<tr>
<td>Hawaii Intl. Conference on Systems Sciences (HICSS)</td>
<td>4</td>
</tr>
<tr>
<td>Asia Pacific Software Engineering Conference (APSEC)</td>
<td>4</td>
</tr>
<tr>
<td>European Conference on Software Maintenance and Reengineering (CSMR)</td>
<td>4</td>
</tr>
<tr>
<td>Americas Conference on Information Systems (AMCIS)</td>
<td>3</td>
</tr>
<tr>
<td>Intl. Conference on Software Engineering and Knowledge Engineering (SEKE)</td>
<td>3</td>
</tr>
<tr>
<td>Intl. Symposium on Empirical Software Engineering and Measurement (ESEM)</td>
<td>3</td>
</tr>
<tr>
<td>Other Conferences</td>
<td>91</td>
</tr>
<tr>
<td><strong>Conference papers total</strong></td>
<td><strong>131</strong></td>
</tr>
<tr>
<td>IEEE Transactions on Software Engineering (TSE)</td>
<td>11</td>
</tr>
<tr>
<td>Information and Software Technology (IST)</td>
<td>7</td>
</tr>
<tr>
<td>Journal of System and Software (JSS)</td>
<td>6</td>
</tr>
<tr>
<td>ACM Computing Surveys (CSUR)</td>
<td>6</td>
</tr>
<tr>
<td>IEEE Computer</td>
<td>5</td>
</tr>
<tr>
<td>Empirical Software Engineering (ESE)</td>
<td>4</td>
</tr>
<tr>
<td>Journal of Software: Evolution and Process</td>
<td>4</td>
</tr>
<tr>
<td>IEEE Software</td>
<td>3</td>
</tr>
<tr>
<td>Communications of the ACM</td>
<td>3</td>
</tr>
<tr>
<td>Requirements Engineering</td>
<td>3</td>
</tr>
<tr>
<td>Other Journals</td>
<td>29</td>
</tr>
<tr>
<td><strong>Journal papers total</strong></td>
<td><strong>81</strong></td>
</tr>
<tr>
<td><strong>Workshop papers total</strong></td>
<td><strong>32</strong></td>
</tr>
</tbody>
</table>
Figure 4.7: Yearly distribution of primary studies by knowledge areas. Horizontal axes represent the years (starting 1987), while vertical axes denote the number of taxonomies.

5.2 Classification scheme results

In this section, we present the results corresponding to the three facets of the classification scheme described in Section 4, i.e. SE knowledge area (KA), research type and presentation approach.
The vertical axis in Figure 4.8 depicts the SE knowledge areas in which taxonomies have been proposed. Construction and design are the leading SE knowledge areas with 52 (21.22%) and 50 (20.41%) taxonomies, respectively. These are relatively mature SE fields with a large body of knowledge and a high number of subareas.

A high number of taxonomies have also been proposed in the knowledge areas requirements (38 – 15.51%), maintenance (31 – 12.65%) and testing (21 – 8.57%). Few taxonomies have been proposed in economics (3 – 1.14%).
and professional practice \((2 - 0.76\%)\), which are more recent knowledge areas.

The results show that most SE taxonomies \((78\%)\) are proposed in the knowledge areas requirements, design, construction, testing and maintenance, which correspond to the main activities in a typical software development process \([110]\).

The horizontal axis in Figure 4.8 shows the distribution of taxonomies by research types, according to Wieringa et al. \([151]\). Most taxonomies are reported in papers that are classified as “solution proposals” \((120 - 48.98\%)\), wherein the authors propose a taxonomy and explain or apply it with the help of an illustration. Eighty three taxonomies \((33.88\%)\) are reported in “philosophical papers”, wherein authors propose a taxonomy, but do not provide any kind of validation, evaluation or illustration. Relatively fewer taxonomies are reported in “evaluation paper” \((32 - 13.06\%)\) and “validation paper” \((10 - 4.08\%)\).

Figure 4.8 also depicts the classification of the taxonomies using 2 aspects of the classification scheme, i.e. SE knowledge area and research type.

Taxonomies in the knowledge areas construction and design are mostly reported either as solution proposals (construction – 27; design – 31) or philosophical papers (construction – 19; design – 16). There are few taxonomies in design and construction that are reported as evaluation papers. Three taxonomies are reported as validation papers in the KA construction, while no taxonomy is reported as a validation paper.

Taxonomies in the knowledge areas requirements, maintenance and testing are better distributed across different research types, wherein besides the solution proposal and the philosophical research types, a reasonable percentage of taxonomies are reported as evaluation or validation papers.

The horizontal axis in Figure 4.9 shows the distribution of taxonomies by presentation approach. Most taxonomies \((57.96\%)\) are presented purely as text or table, while \(42.04\%\) of the taxonomies are presented through some graphical notation in combination with text.

Figure 4.9 also displays the classification of the identified taxonomies in terms of SE knowledge area and presentation approach. The results show 2 different trends:

- For knowledge areas such as design, quality, models and methods, and process, both textual and graphical approaches are used an almost equal number of times. This suggests that the taxonomies in the KAs that involve a lot of modeling might be better presented using graphical modeling approaches.
- Most taxonomies in construction \((35\) out of \(52\)), maintenance \((22\) out of \(31\)), testing \((17\) out of \(21\)) and software management \((6\) out of \(6\)) are textually presented.
Figure 4.9: Systematic map – knowledge area vs presentation type.
We extracted data about the following 2 aspects to answer RQ1:

- **Taxonomy definition**: We investigated from each study whether or not the authors made any attempt to communicate their understanding about the concept of taxonomy by citing or presenting any definition of it.

- **Taxonomy purpose**: We identified from each study the stated (if any) main reason for designing a taxonomy.

The results show that only 5.7% (14) of the taxonomies were reported with a definition for the term “taxonomy”. Out of these 14 taxonomies, 3 use the Cambridge dictionary’s definition (see Section 2), 7 studies do not provide an explicit source and the remaining 4 have other unique references: The American heritage dictionary, Carl Linnaeus, Whatis and the IEEE standard taxonomy for SE standards. For the remaining 94.3% (231) taxonomies, no definition of “taxonomy” was provided.

To identify the purpose of each taxonomy, we extracted the relevant text, referred here as purpose descriptions, from each of the primary studies. Using a process similar to open coding, we coded these descriptions. These codes are the keywords that the primary studies’ authors themselves used in purpose descriptions to describe the taxonomy’s purpose.

Table 4.5 lists the taxonomy purpose codes. The results show that 48.57% of the taxonomies are designed to classify different subject matters, followed by categorize (18 – 7.35%), characterize (15 – 6.12%) and identify (12 – 4.9%) as purposes. All unique keywords (i.e. with frequency equal to 1) are grouped as “Other” in Table 4.5. Some examples for these unique purpose keywords include support, explore, overview, capture and catalog. We could not identify purpose descriptions associated with 16 taxonomies (6.53%).

**Discussion**

Taxonomy is a non-trivial concept that has a defined meaning. Some of its definitions are listed in section 2. In SE discipline, a large number of taxonomies have been proposed, but in only a handful of the cases (5.7%) the authors make any effort to clarify their understanding and perspective of the term “taxonomy”.

Going beyond the citation or use of some definition for the term “taxonomy”, we also attempted to extract the taxonomy purpose from each primary study. Taxonomy is defined as a classification mechanism; hence,
### Table 4.5: Taxonomy purpose.

<table>
<thead>
<tr>
<th>Taxonomy purpose</th>
<th>F</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classify</td>
<td>119</td>
<td>48.57</td>
</tr>
<tr>
<td>Categorize</td>
<td>18</td>
<td>7.35</td>
</tr>
<tr>
<td>Characterize</td>
<td>15</td>
<td>6.12</td>
</tr>
<tr>
<td>Identify</td>
<td>12</td>
<td>4.90</td>
</tr>
<tr>
<td>Understand</td>
<td>9</td>
<td>3.67</td>
</tr>
<tr>
<td>Describe</td>
<td>7</td>
<td>2.86</td>
</tr>
<tr>
<td>Organize</td>
<td>6</td>
<td>2.45</td>
</tr>
<tr>
<td>Define</td>
<td>4</td>
<td>1.63</td>
</tr>
<tr>
<td>Analyze</td>
<td>3</td>
<td>1.22</td>
</tr>
<tr>
<td>Determine</td>
<td>3</td>
<td>1.22</td>
</tr>
<tr>
<td>Other</td>
<td>33</td>
<td>13.47</td>
</tr>
<tr>
<td>Not possible to identify</td>
<td>16</td>
<td>6.53</td>
</tr>
</tbody>
</table>

The primary purpose of any taxonomy is to classify a subject matter. Nevertheless, we found that about 56% of the taxonomies are designed to classify (48.57%) or categorize (7.35%) a subject matter, while remaining taxonomies (44%) are designed to achieve some other purpose (e.g. define and understand), although they do classify some subject matter, i.e. they employ some type of classification structure.

#### 5.4 RQ2 – Subject matter

In total, we identified 240 unique subject matters\(^\text{11}\) for the 245 taxonomies, e.g. technical debt, architectural constraints, usability requirements, testing techniques and process models.

**Discussion**

The high number of unique subject matters means that almost each taxonomy dealt with a unique subject matter. This might be due to the following reasons:

- Existing taxonomies fit their purpose well. Therefore there is no need to define competing taxonomies.

\(^\text{11}\) See https://drive.google.com/open?id=0B2kvKPmJJREDbl0wWDBfbbNOeUk for the full list.
• The subject matters for existing taxonomies are so narrowly defined that they are not suitable for usage outside their original context. New taxonomies are therefore developed constantly.
• SE researchers do not evaluate existing taxonomies whenever there is need for organizing SE knowledge, but rather propose new ones.

One indicator for taxonomy use is the number of times each primary study is cited. This analysis is detailed in Subsection 5.7.
The list of subject matters contains mainly technical aspects of SE. Only few taxonomies deal with people-related subject matters, e.g. stakeholder-related and privacy-related.

5.5 RQ3 – Validation of taxonomies

Table 4.6 displays the approaches used to validate taxonomies. Illustration is the most frequently used approach to validate taxonomies (112 – 45.7%). Illustration includes approaches such as example, scenario and case.
Case studies have also been used to validate 32 taxonomies (13.1%). Experiments have been used to validate 10 taxonomies, while few taxonomies have also be validated through expert opinion (7 – 2.9%) and 33.9% (83) of the taxonomies have not been validated.

<table>
<thead>
<tr>
<th>Validation Method</th>
<th>F</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illustration</td>
<td>112</td>
<td>45.7</td>
</tr>
<tr>
<td>Case study</td>
<td>32</td>
<td>13.1</td>
</tr>
<tr>
<td>Experiment</td>
<td>10</td>
<td>4.1</td>
</tr>
<tr>
<td>Expert opinion</td>
<td>7</td>
<td>2.9</td>
</tr>
<tr>
<td>Survey</td>
<td>1</td>
<td>0.4</td>
</tr>
<tr>
<td>No Validation</td>
<td>83</td>
<td>33.9</td>
</tr>
</tbody>
</table>

Discussion
The results related to RQ3 show that very few taxonomies are validated through methods like case study or experiment, and a large number of taxonomies (33.9%) have not been validated by any means. We do not believe that one particular validation approach would be best for all contexts; however we believe that in most cases would not be enough just to propose a taxonomy. The researchers should make an effort to select and use
a suitable approach either to demonstrate the taxonomy with the help of an illustration or to validate the taxonomy using formal methods, such as case study.

5.6 RQ4 – Classification structure

To answer RQ4, the following data was gathered: classification structure, descriptive bases, classification procedure and classification procedure description.

Table 4.7 shows the classification structures identified for the identified taxonomies. Hierarchy was the most frequently used classification structure (125 – 51.02%), followed by faceted-based (102 – 41.63%), tree (12 – 4.9%) and paradigm (6 – 2.45%).

Table 4.7: Classification structure.

<table>
<thead>
<tr>
<th>Classification structure</th>
<th>F</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hierarchy</td>
<td>125</td>
<td>51.02</td>
</tr>
<tr>
<td>Faceted analysis</td>
<td>102</td>
<td>41.63</td>
</tr>
<tr>
<td>Paradigm</td>
<td>6</td>
<td>2.45</td>
</tr>
<tr>
<td>Tree</td>
<td>12</td>
<td>4.9</td>
</tr>
</tbody>
</table>

Table 4.8 displays the status of the taxonomies’ descriptive basis. The majority of the taxonomies have a sufficiently clear description of their elements (233 – 91.02%), followed by only 22 taxonomies (8.98%) without a sufficient description.

Table 4.8: Descriptive bases.

<table>
<thead>
<tr>
<th>Descriptive basis</th>
<th>F</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sufficiently clear described</td>
<td>223</td>
<td>91.02</td>
</tr>
<tr>
<td>Not sufficiently clear described</td>
<td>22</td>
<td>8.98</td>
</tr>
</tbody>
</table>

Table 4.9 presents the classification procedure types for the identified taxonomies. The majority of the taxonomies employed a qualitative classification procedure (238 – 97.14%), followed by quantitative (5 – 2.04%) and both (2 – 0.82%).
Table 4.9: Classification procedure types.

<table>
<thead>
<tr>
<th>Classification procedure</th>
<th>F</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qualitative</td>
<td>238</td>
<td>97.14</td>
</tr>
<tr>
<td>Quantitative</td>
<td>5</td>
<td>2.04</td>
</tr>
<tr>
<td>Both</td>
<td>2</td>
<td>0.82</td>
</tr>
</tbody>
</table>

Table 4.10 displays the status of the taxonomies’ classification procedure description. The majority of the taxonomies do not have an explicit description for the classification procedure (122 – 86.53%) and only 33 taxonomies (13.47%) have an explicit description.

Table 4.10: Classification procedure descriptions.

<table>
<thead>
<tr>
<th>Procedure description</th>
<th>F</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not explicitly described</td>
<td>212</td>
<td>86.53</td>
</tr>
<tr>
<td>Explicitly described</td>
<td>33</td>
<td>13.47</td>
</tr>
</tbody>
</table>

Discussion

The results clearly indicate that hierarchical classification structures (hierarchy, tree and paradigm) are the ones mostly used to design SE taxonomies. This is not surprising, since taxonomies were originally designed using hierarchy as classification structure (see Introduction).

Nevertheless, SE is considered as a young discipline that is in constant and fast evolution. For this reason, the existing knowledge is sometimes incomplete or unstable. This is probably the reason faceted analysis is also frequently employed as classification structure to design SE taxonomies; such classification structure is the one that is most appropriate for knowledge fields with incomplete and unstable knowledge [79]. It was interesting to observe that in almost all studies the choice to employ a specific classification structure is not explicitly motivated.

The descriptive bases are mandatory for a reasonable understanding of a taxonomy (see 2). The absence of such element can hinder the adoption of taxonomies. Not surprisingly, the 24 studies that did not provide descriptive bases are lowly cited, it might indicate the low usefulness of them.

At first glance, it appears that the main reason qualitative classification procedures are more popular is the fact that such approach is easier to
be designed; the classification using such procedure type is performed using nominal scales. However, such approach is harder to be applied, because it can leave room for ambiguities; it is not possible to establish the differences/similarities between classes because distance functions cannot be used to calculate the difference/similarity degree between classes [150]. Quantitative approaches are harder to be designed, but they are easier to be applied to classify subjects, since it is possible to use distance functions to identify the differences/similarities between classes [150].

To enable accurate subject classifications, the classification procedure of a taxonomy must be described sufficiently. Only 13.47% of the taxonomies (33) have an explicit description for their classification procedure. The remaining 86.53% of the taxonomies (212) are probably harder to be used by other people than the designers of the taxonomies themselves.

5.7 RQ5 – Taxonomy usage level

As an approximation for the usage level of SE taxonomies, we looked at the number of citations for each included primary study. Since a publication may be cited due to many different reasons, the number of citations can only be an indication of actual use of the proposed taxonomy. However, the number of citations show whether there is an interest in a subject.

Sixty one primary studies (25%) have been cited ≥ 50 times (see Table 4.11). There are only 16 studies that were never cited. Out of these 16 primary studies, 7 were published between 2012 and 2014.

Table 4.11: Number of primary studies (Frequency) by number of citations.

<table>
<thead>
<tr>
<th>Citations</th>
<th>F</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥ 300</td>
<td>8</td>
<td>3.3</td>
</tr>
<tr>
<td>≥ 200</td>
<td>16</td>
<td>6.6</td>
</tr>
<tr>
<td>≥ 100</td>
<td>30</td>
<td>12.3</td>
</tr>
<tr>
<td>≥ 50</td>
<td>61</td>
<td>25</td>
</tr>
<tr>
<td>≥ 20</td>
<td>105</td>
<td>43</td>
</tr>
<tr>
<td>≥ 10</td>
<td>143</td>
<td>58.6</td>
</tr>
<tr>
<td>≥ 1</td>
<td>228</td>
<td>93.4</td>
</tr>
<tr>
<td>0</td>
<td>16</td>
<td>6.6</td>
</tr>
</tbody>
</table>

12 The number of citations for each primary study was fetched from Google Scholar in January 2015.
Figure 4.10 shows the distribution of average number of citations per year for the primary studies. Most studies were cited 1–5 times per year on average (116 – 47.5%).

Table 4.12 lists the mean and median number of citations for the knowledge areas with at least 10 taxonomies. The data shows that for all knowledge areas, except for process, the mean value is much higher than the median, which is due to few studies with very high numbers of citations (outliers); e.g. construction has a study with 1440 citations. Maintenance has fewer taxonomies (38), as construction (52) and design (50), but it has the highest median for the number of citations.

![Figure 4.10: Distribution of average number of citations per year. The x-axis shows ranges of average citations per year. The y-axis shows the number of primary studies.](image)

**Discussion**

The results show that most studies reporting taxonomies are cited a reasonable number of times with many primary studies that are highly cited in all knowledge areas. This indicates an interest in taxonomies by the SE community. The high number of unique subject matters (see Subsection 5.4) might be due to the breadth of the SE field and a need for classifying new and evolving SE subareas.

We also observed that some recently proposed taxonomies (2012–2014) were already cited many times. These results highlight that new taxonomies are proposed continuously and cited frequently. This indicates that the classification of SE knowledge through taxonomies is a relevant topic.

---

13 Average is computed by dividing citation count of a study by number of years from publication date to 2014.
A more thorough analysis of the actual usage of the cited primary studies would be necessary though, to understand the contexts and reasons for taxonomy usage.

Table 4.12: Number of citations (mean and median) for primary papers by knowledge area.

<table>
<thead>
<tr>
<th>Knowledge area</th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construction</td>
<td>94.5</td>
<td>17</td>
</tr>
<tr>
<td>Design</td>
<td>43.4</td>
<td>14.5</td>
</tr>
<tr>
<td>Requirements</td>
<td>41.8</td>
<td>12</td>
</tr>
<tr>
<td>Maintenance</td>
<td>43</td>
<td>24</td>
</tr>
<tr>
<td>Testing</td>
<td>36.5</td>
<td>13</td>
</tr>
<tr>
<td>Quality</td>
<td>49.2</td>
<td>16</td>
</tr>
<tr>
<td>Models &amp; Methods</td>
<td>86.5</td>
<td>14</td>
</tr>
<tr>
<td>Process</td>
<td>12.1</td>
<td>7.5</td>
</tr>
</tbody>
</table>

6  A REVISED TAXONOMY DESIGN METHOD

A method to support the design of SE taxonomies was recently proposed by Bayona-Oré et al. [11], which consists of 5 phases and 24 activities (see Tables 4.13 and 4.14). Bayona-Oré et al. based their method on an aggregation of the steps suggested in 9 sources and used it to design a taxonomy of critical factors for software process deployment.

Bayona-Oré et al.’s method includes many activities that are important for systematically designing SE taxonomies. However, their method has some issues, as follows:

- **Issue 1** – Six out of nine sources on which Bayona-Oré et al.’s method is based are either gray literature or are not available in English.
- **Issue 2** – The proposed method is a simple aggregation of all the activities suggested in the nine sources. Some of these activities are only used in one of the peer-reviewed sources.
- **Issue 3** – The scope of some activities of this method goes beyond taxonomy design. It covers steps related to project planning, recruitment, training and deployment. The method also seems to be targeted towards committee efforts.
Table 4.13: The revised taxonomy design method - part 1.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Activity according to Bayona-Oré et al. [11]</th>
<th>Revised activity</th>
<th>Rationale for change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning</td>
<td>A1: Identify the area of the study</td>
<td>B1: Define SE knowledge area</td>
<td>Similar to A1</td>
</tr>
<tr>
<td></td>
<td>A2: Define the objectives of the taxonomy</td>
<td>B2: Describe the objectives of the taxonomy</td>
<td>Combination A2 and A4</td>
</tr>
<tr>
<td></td>
<td>A3: Develop a survey of user needs</td>
<td>—</td>
<td>Deleted due to issue 3</td>
</tr>
<tr>
<td></td>
<td>A4: Define the scope of the proposed taxonomy</td>
<td>—</td>
<td>Part of B1 and B2</td>
</tr>
<tr>
<td></td>
<td>A5: Define the team in charge of developing the taxonomy</td>
<td>—</td>
<td>Deleted due to issue 3</td>
</tr>
<tr>
<td></td>
<td>A6: Identify the required resources</td>
<td>—</td>
<td>Moved to next “phase”</td>
</tr>
<tr>
<td></td>
<td>A7: Document the plan</td>
<td>—</td>
<td>Deleted due to issue 3</td>
</tr>
<tr>
<td></td>
<td>A8: Obtain commitment and support</td>
<td>—</td>
<td>Deleted due to issue 3</td>
</tr>
<tr>
<td></td>
<td>B3: Describe the subject matter to be classified</td>
<td>Based on results of this study, from Wheaton [150] and Glass and Vessey [50]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B4: Select classification structure type</td>
<td>Based on results of this study, from Kwasnik [79] and Glass and Vessey [50]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B5: Select classification procedure type</td>
<td>Based on results of this study, from Wheaton [150] and Glass and Vessey [50]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B6: Identify the sources of information</td>
<td>Same as A9; considered as a planning activity</td>
<td></td>
</tr>
<tr>
<td>Identification and extraction</td>
<td>A9: Identify the sources of information</td>
<td>—</td>
<td>Moved to planning phase</td>
</tr>
<tr>
<td></td>
<td>A10: Extract all terms and identify candidate categories</td>
<td>B7: Extract all terms</td>
<td>Partly corresponds to A10; categories’ identification is covered by B10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B8: Perform terminology control</td>
<td>Combines A11 and A13</td>
</tr>
</tbody>
</table>
Table 4.14: The revised taxonomy design method - part 2.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Activity according to Bayona-Oré et al. [11]</th>
<th>Revised activity</th>
<th>Rationale for change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design and</td>
<td>A11: Check the list of terms and define the criteria</td>
<td>—</td>
<td>Corresponds to B8</td>
</tr>
<tr>
<td>Construction</td>
<td>A12: Define the first level of the taxonomy design</td>
<td>—</td>
<td>Part of B9 and B10</td>
</tr>
<tr>
<td></td>
<td>A13: Perform terminology control</td>
<td>—</td>
<td>Moved to identification and extraction phase</td>
</tr>
<tr>
<td></td>
<td>A14: Define the subsequent levels of the taxonomy</td>
<td>—</td>
<td>Covered in B10</td>
</tr>
<tr>
<td></td>
<td>A15: Review and approve the taxonomy by stakeholders and experts</td>
<td>—</td>
<td>Deleted due to issue 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B9: Identify and describe taxonomy dimensions</td>
<td>Based on results of this study, from Kwasnik [79] and Glass and Vessey [50]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B10: Identify and describe categories of each dimension</td>
<td>Based on results of this study, from Wheaton [150], Kwasnik [79], Glass and Vessey [50] and A10, A12, A14.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B11: Identify and describe the relationships</td>
<td>Based on results of this study, from Wheaton [150], Kwasnik [79] and Glass and Vessey [50]</td>
</tr>
<tr>
<td></td>
<td>A16: Define the guidelines for using and updating the taxonomy</td>
<td>B12: Define the guidelines for using and updating the taxonomy</td>
<td>Same as A16</td>
</tr>
<tr>
<td>Testing and</td>
<td>A17: Test the taxonomy</td>
<td>B13: Validate the taxonomy</td>
<td>Combination of A17 and A18</td>
</tr>
<tr>
<td>Validation</td>
<td>A18: Incorporate improvements as a result of the tests</td>
<td>—</td>
<td>Part of B13</td>
</tr>
<tr>
<td>Deployment</td>
<td>A19: Prepare the training plan</td>
<td>—</td>
<td>Deleted due to issue 2 and 3</td>
</tr>
<tr>
<td></td>
<td>A20: Train users</td>
<td>—</td>
<td>Deleted due to issue 2 and 3</td>
</tr>
<tr>
<td></td>
<td>A21: Collect evidence of learning</td>
<td>—</td>
<td>Deleted due to issue 1 and 3</td>
</tr>
<tr>
<td></td>
<td>A22: Use the selected technology and make the taxonomy available</td>
<td>—</td>
<td>Deleted due to issue 1 and 3</td>
</tr>
<tr>
<td></td>
<td>A23: Develop the management and maintenance plan</td>
<td>—</td>
<td>Deleted due to issue 2 and 3</td>
</tr>
<tr>
<td></td>
<td>A24: Manage and maintain the taxonomy</td>
<td>—</td>
<td>Deleted due to issue 3</td>
</tr>
</tbody>
</table>
Therefore, based on the results reported in this paper and on the studies by Wheaton [150], Glass and Vessey [50] and Kwasnik [79], we revised Bayona-Oré et al.’s method, as follows:

1. We excluded phases and activities that were out of the scope of taxonomy design.
2. We analyzed the remaining activities in the light of additional literature ([79, 50, 150]) and our data analysis results.

As a result, the revised approach method has 4 phases with 13 activities. A summary of the approach, in comparison to Bayona-Oré et al.’s method, is shown in Tables 4.13 and 4.14. A more detailed description of each activity of the revised method is provided in Subsections 6.1–6.4.

6.1 Planning

This phase has 6 activities that define the taxonomy initial setting, i.e. SE knowledge area, objective, subject matter, classification structure type, classification procedure type and sources of information.

In activity B1, one must select and make clear the SE knowledge area for each a new taxonomy is to be designed. In doing so, it is easier to understand the context of the taxonomy and thus apply it.

In Activity B2, the objective, purpose, aimed users and boundaries of the new taxonomy must be clearly defined.

Activity B3 selects and describes the subject matter to be classified, which is a fundamental step to design a taxonomy (see Sections 2 and 5).

Activity B4 selects an appropriate classification structure type to ground the design of the new taxonomy, which can be hierarchy, tree, paradigm or facet-based (see Sections 2 and 5).

In activity B5, one must select an appropriate classification procedure type to ground the classification of exiting knowledge related to the new taxonomy, which can be qualitative, quantitative or both (see Sections 2 and 5).

Activity B6 determines one or more data sources and data collection methods to enable the prospection of knowledge related to the new taxonomy. Examples of data collection methods are interviews, observation, archival research, survey and simulation [152].

6.2 Identification and extraction

This phase has 2 activities that allow for extracting and controlling the terms associated to the taxonomy to be designed.
In activity B7, the terms relevant to the new taxonomy must be extracted from the collected data.

As aforementioned, taxonomies are specially useful for organizing new knowledge areas that have evolving and also volatile knowledge. It is very often the case that in such knowledge areas there is no common terminology, e.g. two different terms can mean the same or one term is used to represent different things. Activity B8 identifies and removes redundancies and inconsistencies in the terminology related to the new taxonomy.

6.3 Design and construction

This phase has 4 activities that allow for identifying and describing the dimensions, categories and relationships of the new taxonomy. In addition, guidelines for using and evolving the new taxonomy are to be provided.

In activity B9, dimensions must be identified and described. Dimensions are usually identified using knowledge acquired about the subject matter to be classified. In taxonomies that employ hierarchy or tree as classification structure, just one dimension is expected to be employed during the design and construction stage. Taxonomies that employ paradigm as classification structure must have two different dimensions. Finally, taxonomies that employ a facet-based classification structure must have at least two different dimensions.

Activity B10 identifies and describes categories for each of the identified identified dimensions; each dimension must have at least two categories.

In activity B11, one must clearly identify and describe the relationship between dimensions and categories. Note that in some cases there is no relationship between dimensions, i.e. the execution of this activity can eventually yield no results.

To facilitate the adoption and evolution of the new taxonomy, guidelines must be clear defined, which is done by activity B12.

6.4 Validation

To ensure that the selected subject matter is clearly, concisely and thoroughly classified, it is necessary to validate the new taxonomy, which is done by activity B13. A taxonomy can be validated in three ways [149]:

- **Orthogonality demonstration** - The orthogonality of the taxonomy dimensions and categories must be demonstrated.
- **Benchmarking** - The taxonomy must be compared with other similar classification schemes.
- **Utility demonstration** - The utility of the taxonomy must be demonstrated through classify existing knowledge.
The aforementioned validations can be achieved through approaches such as classification of literature, expert opinion, survey, case study and experiment.

Considering the variety of validation approaches identified during the conduction of this mapping study, it is not possible to clearly indicate each approach must be preferably employed. Independently of the employed validation method, it is fundamental to consider the result of the validation and improve the new taxonomy whenever required.

7 THREATS TO VALIDITY

Three major validity threats were identified and mitigated in this paper:

- **Coverage of the search string** - This threat means the efficiency of the applied search string to reach relevant primary studies. To mitigate this threat, the search string was designed to be as more comprehensive as possible. To achieve such goal, we not only included the SWEBOK SE related knowledge areas; the synonyms of the included knowledge areas were also included. It enabled the search string to return a high number of primary studies (1371 unique studies). The accuracy of the search string was considered as fair, since 19,76% of the 1371 were included after the selection process. Finally, an additional mitigation action was to apply the designed search string in all the most relevant SE databases.

- **Study selection** - This threat means the possibility of a study being classified in different manners by different reviewers. The first mitigation strategy was to define a clear selection criteria. The selection criteria were discussed by the authors of this paper to ensure that all of them shared the same understanding of the criteria. Whenever a paper was excluded, the reviewer was to give the reason for the exclusion. The second mitigation action was to perform cross-validations in both level-1 and level-2 screenings, as detailed in Subsection 4.3.

- **Data extraction** - This threat means the possibility of a study data being extracted in different manners by different reviewers. The first action to mitigate this threat was to design a spreadsheet to be used by all the authors during the data extraction. The spreadsheet was discussed by the authors to ensure a common understanding. The second mitigation action was to perform a cross-validation of the extracted data, as detailed in Subsection 4.4. Even considering the discussions conducted to ensure a common understanding of each item of the designed spreadsheet, the item “classification structure” was specially difficult to extract; none of the papers explicitly provided that item and it had to be inferred from the text, which led to high level of disagreement among the authors. Thus, the third mitiga-
tion action, performed jointly by the first, second and third authors, was to re-screen all the 254 studies, focusing on the "classification structure", as described in Subsection 4.4.

8 CONCLUSION

In this paper we reported the results of a systematic mapping study on taxonomies in SE discipline. The initial search returned 1371 results (without duplicates). The application of a two-phased study selection and validation processes resulted in the inclusion of a final set of 244 primary studies.

We observed a rising trend in publishing and using taxonomies in SE during recent years. Taxonomies have been published mostly in domain specific SE conferences (e.g. Requirements’ Engineering Conference) and highly reputed SE journals (e.g. IEEE Transactions in Software Engineering and Information and Software Technology). Regarding the research type, most studies were categorized as either solution proposal or philosophical paper.

About 78% of the taxonomies are proposed in 5 SE KAs (requirements, design, construction, testing and maintenance) that are known as framework activities in software process [110]. We identified 240 unique subject matters, wherein the majority of them are related to technical aspects of SE (e.g. requirements, defects, tools and techniques). More than half of the primary studies presented the taxonomies in a textual manner. However, we noticed that it was easier to understand the taxonomies presented in a graphical manner.

Although taxonomies are vastly recognized as classification mechanisms, a large number of primary studies stated taxonomy purposes that are not related to classification. However, those taxonomies are able to classify the aimed subject matters.

We identified that a large number of primary studies have either used an illustration (45.7%) to validate the proposed taxonomy or have not performed any validation (33.9%) at all. We believe that researchers should, whenever possible, select a suitable approach for validating the proposed taxonomy considering factors such as taxonomy purpose and audience.

We also identified that hierarchy (51.02%) and facet-based (41.63%) were the most frequently used classification structures. Majority of the studies used a qualitative procedure to assign subject matter instances to classes.

The overall result of this paper indicates that, except for one study, SE taxonomies have been developed without following any systematic approach. Bayona-Oré et al. have recently proposed a method to design SE taxonomies. Nevertheless, in the light of results presented in this paper, we identified a need to revise this method. Therefore, we revised Bayona-Oré et al.’s method, which is another contribution of this paper. We intend to
apply this revised method to validate and demonstrate its usefulness in designing and presenting SE taxonomies.
EXTENDING AND SPECIALIZING A GLOBAL SOFTWARE ENGINEERING TAXONOMY

1 INTRODUCTION

Throughout the years, the software industry has applied many different software development approaches, aiming at increasing process efficiency and profit. Numerous companies worldwide develop software in a globally distributed manner to achieve benefits such as reduced time-to-market and access to skillful people all over the world [114, 29, 18, 58]. Such a software development approach is called Global Software Development (GSD) [58] and is also known as Global Software Engineering (GSE) or Distributed Software Development (DSD). In general, the terms GSE and DSD are used to identify distributed software development projects, embracing both work distribution inside one country and across the boundaries of different countries [149]. On the other hand, GSD is usually used to identify software development projects performed globally, embracing work distribution over different countries.

Despite all the benefits argued to be achieved by means of GSE, there are also known challenges related to such a software development approach, such as communication and coordination between the involved entities [57]. The mitigation of such challenges has been one of the research agendas to those carrying out research in this area.

The research in GSE has evolved in the last decade and there is a lack of a common terminology and knowledge organization in such area. The absence of common terminology and knowledge organization can hinder the understanding of studies’ contexts, making harder to analyze, compare and aggregate the results from similar studies. In addition, it also complicates the identification of research gaps and the transfer of research results to industry. A classification scheme can mitigate the aforementioned problems [141].

\[1\] For the sake of simplicity, we decided to use the term GSE in the rest of this paper, which is deemed more comprehensive.
In the context of Software Engineering (SE), it has been common to use taxonomies as classification schemes to organize the different SE knowledge areas [26, 138, 149]. According to the Oxford English Dictionary [37], a taxonomy is “a scheme of classification”. This concept was initially devised to classify organisms [84], although it has been applied in many different domains, e.g. Education [16], Psychology [97] and Computer Science [120]. Originally, the taxonomy approach was designed to enable the description of terms and their relationships in the context of a knowledge area in a hierarchical way. Nevertheless, to date many different classification structures have been used to construct taxonomies, e.g. “hierarchy”, “tree” and “facet-based” [79].

Recently, Smite et al. proposed a taxonomy to classify sourcing strategies in GSE projects [149]. However, that taxonomy does not embrace all the existing factors related to the development of software in a globally distributed manner. Thus, Smite et al.’s taxonomy can be extended to also include additional global factors, thus leading to a more detailed way of classifying GSE scenarios. In addition, the taxonomy can also be specialized to classify GSE scenarios from specific perspectives, such as software testing and effort estimation.

Therefore, the goal and main contribution of this paper is three-fold:

• To extend the taxonomy proposed by Smite et al. [149].
• To specialize the taxonomy proposed by Smite et al. [149].
• To present a specialized taxonomy for effort estimation in the GSE context. This example was grounded on a previously performed systematic literature review (SLR) on effort estimation in GSE [21].

The extension and specialization presented in this paper complement the GSE-related terminology originally proposed by Smite et al. [149]. The terms associated with the designed extension and specialization are consistent and concise with Smite et al.’s taxonomy and terminology.

We anticipate the approach used in this paper to guide other GSE researchers when extending or specializing Smite et al.’s taxonomy. Thus, any new extension or specialization is expected to be consistent and concise with the original taxonomy. This is needed to avoid the fragmentation of the defined GSE terminology.

We believe that both the extended and specialized dimensions presented herein will help researchers and practitioner to find cases of interest for them; the specialized taxonomy provides a lingua franca that can be used to report new findings on effort estimation in GSE, facilitating the analysis, comparison and aggregation of results from new studies.

The remainder of this paper is organized as follows. A discussion of related work is presented in Section 2. Section 3 explains the applied research methodology. The process used for extending and specializing Smite et al.’s GSE taxonomy is presented in Section 4. Section 5 presents the designed
extension for Smite et al.’s GSE taxonomy. Section 6 presents the specialized GSE effort estimation taxonomy. The classification process that guides the usage of the specialized GSE effort estimation taxonomy is explained in Section 7. Section 8 illustrates how to classify GSE scenarios using the specialized effort estimation GSE taxonomy. Section 9 provides a summary of the research questions’ answers and a discussion on the limitations of this work. Finally, in Section 10 we draw our conclusions and present directions for future work.

2 RELATED WORK

In this section the related work is presented as follows: In Subsection 2.1 we further describe Smite et al.’s GSE taxonomy, which was used as the main basis for the specialized GSE effort estimation taxonomy proposed in this paper. In Subsection 2.2 additional relevant related work is described and compared against the proposal put forward in this paper.

2.1 The Smite et al.’s GSE taxonomy

Smite et al. [149] conducted a Delphi-inspired study with GSE researchers to develop an empirically based glossary and taxonomy, focused on the sourcing strategy aspect of GSE projects.

To construct the taxonomy, firstly, the authors investigated the state of the art in the use of GSE terminology by systematically reviewing studies from GSE-related venues. Secondly, by using a Delphi-based approach, they evaluated the literature and defined a consensual terminology. Finally, using the defined terminology the authors identified the relationship between the defined terms. To illustrate the usage of the proposed GSE taxonomy, the authors applied their proposal to classify sourcing strategies presented in 68 different studies.

The GSE taxonomy proposed by Smite et al. was developed to classify the relationship between pairs of sites, although it is equally possible to describe more complex GSE projects, with more than two sites. The taxonomy has five dimensions (Figure 5.1), as follows [149]:

- **GSE** - This dimension contains the root of the taxonomy, called sourcing. In this context, sourcing means some form of external software development.
- **Location** - A sourcing can be delegated to a site in the same country, i.e. onshore, or to a site in another country, i.e. offshore.
- **Legal entity** - Independently from the location, a sourcing can be transferred to a different branch (site) of the requester company, i.e. insourcing, or subcontracted from a different legal entity (company), i.e. outsourcing.
Figure 5.1: The GSE taxonomy (Adapted from Smite et al. [149]).
• **Geographical distance** - There are four different kinds of geographical distances, which depend on the location dimension:

  – **Close** - In onshore projects, the geographical distance is considered close when it is possible to have relatively frequent face-to-face meetings, since no flights are required to go from one site to the other.
  
  – **Distant** - In onshore projects, the geographical distance is considered distant when at least one flight is required to have face-to-face meetings, which yields time and cost increases.
  
  – **Near** - In offshore projects, the geographical distance is considered near when the required flying time is less than two hours. It means that is possible to hold a meeting of three or four hours and travel back and forth within the same day.
  
  – **Far** - In offshore projects, the geographical distance is considered far when the flying time is longer than two hours and staying overnight is usually required.

• **Temporal distance** - There are four different types of temporal distance, which depend on the location dimension:

  – **Similar** - In onshore projects, the temporal distance is considered similar when there is a time difference of one hour or less. It allows an almost complete overlap in work hours between two different sites.
  
  – **Different** - In onshore projects, the temporal distance is considered different when the time difference between two sites is longer than one hour.
  
  – **Small** - In offshore projects, the temporal distance is considered small when there is a time distance between sites of four hours or less. In that situation there is overlap of at least half of a workday between two sites.
  
  – **Large** - In offshore projects, the temporal distance is considered large when there is a time distance between two sites of more than four hours.

Smite et al.’s taxonomy was designed using faceted analysis [79] as its classification structure. The taxonomy has five facets (dimensions), which relate to each other as follows:

- The dimension “GSE” is parent of all other dimensions.
- The classification by means of the dimension “geographic distance” depends on the categories of the dimension “location”.
- The classification by means of the dimension “temporal distance” depends on the categories of the dimension “geographic distance”.
2.2 Additional related work

As a result of the performed literature survey, three taxonomies [52, 83, 149] and two ontologies [144, 89] that provide knowledge organization schemes in the GSE context were identified. It is important to remark that other taxonomies were also identified [3, 115, 61], but they were improved by Smite et al.’s taxonomy [149]. For this reason we decided not to consider them in this work. The taxonomy proposed by Narasipuram [101] was also not considered because the supporting evidence is limited.

Gumm [52] developed a taxonomy to classify GSE projects in terms of distribution dimensions. The main goal of the author was to provide a foundation to discuss the challenges regarding GSE projects. The study was based on an earlier literature study performed by the same author. The proposed taxonomy used four different dimensions to classify the ways people and/or artifacts could be distributed in a GSE project. The proposed dimensions were physical or geographical distribution, organizational distribution, temporal distribution and distribution between stakeholder groups. Each dimension could be measured on a 3-point ordinal scale (high, medium or low). The author described an onshore distributed project to validate her proposal and she argued that the taxonomy helped one to understand the scope and the distribution issues of the evaluated project.

Laurent et al. [83] proposed a taxonomy and a visual notation to address the requirements engineering aspect of GSE projects. The main goals of the authors were to design a common language for modeling the requirements of GSE projects and to allow project managers to manage distributed requirements in a better way. The proposal was derived from the findings of a broad study performed with industrial partners (seven different projects). Interviews were performed with the team leaders responsible for eliciting and gathering the requirements in each project. The proposed taxonomy was divided into three different entities: role, site and artifact. They graphically showed the taxonomy as a Unified Model Language (UML) class diagram with some attributes in each entity. These attributes were related to the entity "site", respectively called location, language and time zone. To facilitate the usage of the proposed taxonomy, the authors also designed a visual notation, which was later on applied to a real GSD project from the video gaming domain. They reported that their proposal helped to identify in advance problems regarding the management of documents and the requirements gathering process.

Vizcaino et al. [144] developed an ontology, called O-GSD, which was aimed at easing the communication and avoiding misunderstanding in GSE projects. This ontology was iteratively developed in the context of a project that involved five companies and two universities in Spain. The
authors used the REFSENO (representation formalism for software engineering) \cite{134} to create the proposed ontology.

This ontology allows for the description of GSE projects by instantiation of different factors, e.g. time zone difference and language distance, roles of the involved members, involved sites and etc. The authors designed a glossary and a UML class diagram to depict the semantic relationship between all the determined concepts. To validate their proposal, the authors used the proposed ontology to describe a real GSE project, which consisted of software related to the sale of security devices in European Union countries. The ontology was able to cover all the concepts required by the involved company to represent the GSE project. Another conclusion was that the ontology fostered a common understanding about the represented project.

Marques et al. \cite{89} proposed an ontology for task allocation to teams in GSE projects. This ontology was developed based on the findings of a systematic mapping study performed by the authors and aimed at clarifying the concepts related to task allocation to teams in distributed projects; proposing ways to analyze and identify adequate approaches to allocate task in such projects. The authors used UML class diagrams to represent the proposed ontology. The main concepts addressed by the authors were artifact, competence and constraints. The authors performed a preliminary evaluation of the ontology by the means of interviews with five project managers. According to the authors, the preliminary evaluation suggested that the concepts and relationships embraced by the proposed ontology are suitable to be applied in real distributed projects.

By analyzing the identified taxonomies and ontologies, we noticed that there are two different kinds of knowledge organization approaches:

- **Description approach:** Three studies \cite{83, 144, 89} proposed graphical-based approaches that are more adequate to describe GSE projects rather than classify. It comes from the fact that none of these approaches has predefined scales to allow the classification of projects. It is interesting to note that these studies proposed ontologies. Laurent et al. \cite{83} called their proposal a taxonomy, but it resembles more an ontology, once the relationship between the elements are explicitly described. Like the other two ontologies, there are elements, which must be instantiated, rather then dimensions, which must be measured using some predefined scale.

- **Classification approach:** The other two studies \cite{52, 149} are more adequate to classify GSE projects. The two studies are taxonomies that are organized in dimensions. Each dimension has some sort of predefined scale, which can be used to classify GSE projects.

Therefore, not all the identified GSE knowledge organization approaches are able to classify GSE scenarios. Furthermore, none of them incorporates
all the relevant factors regarding classifying GSE scenarios focusing on effort estimation.

3 RESEARCH METHODOLOGY

In this section we present the research methodology used in this paper. The following research questions drove the work reported herein:

- **RQ1**: How can a general GSE taxonomy be extended and specialized?
- **RQ2**: What dimensions can be included as extensions to Smite et al.’s taxonomy?
- **RQ3**: What dimensions are required to specialize Smite et al.’s taxonomy focusing on the effort estimation perspective?
- **RQ4**: What is the usefulness of a specialized effort estimation GSE taxonomy?

To answer RQ1, we designed a process to extend and specialize Smite et al.’s taxonomy, which is detailed in Section 4. To answer RQ2 and RQ3, we respectively proposed some extended dimensions and specialized dimensions for Smite et al.’s taxonomy. The extensions and specializations were designed following three activities, as follows:

1. First, we selected the relevant factors to classify GSE scenarios focusing on effort estimation. The selection procedure was based on a previously performed SLR on effort estimation in GSE [21].
2. Second, a literature review was performed to gather literature on the selected factors. We searched for relevant literature using Scopus², Compendex³, Inspec⁴ and Web of Science⁵. Additionally, considering that some of the factors were not covered by the software engineering literature, we also looked at studies from other Computer Science domains, specifically from the Distributed Systems knowledge field [31].
3. Third, the knowledge of one of the authors on effort estimation was incorporated as one of the systematic process inputs, to bridge the remaining gaps related to the factors that were not covered by the state of the art (step 1).

The three above-mentioned activities are further elaborated in Section 5.

To answer RQ4, we illustrate the usage of the specialized effort estimation GSE taxonomy via the classification of eight different finished GSE projects. Further description of the illustration is provided in Section 8.

---

2 www.scopus.com
3 www.engineeringvillage.com
4 www.engineeringvillage.com
5 apps.webofknowledge.com
The designed process used for extending and specializing Smite et al.’s taxonomy has four activities (Figure 5.2), which are described as follows.

The first activity consists of determining the taxonomy classification perspective. A perspective delimits the factors that are mandatory to classify GSE scenarios. Perspectives can be general or specific. General classification perspectives require solely global factors (e.g. temporal and geographic distances) to classify GSE scenarios. Specific classification perspectives require factors related to specific software engineering activities. For example, we can cite “communication” as a general classification perspective and “effort estimation” as a specific classification perspective.

Once the classification perspective is defined, the new factors to be incorporated must be selected (second activity). New factors are introduced into the taxonomy by means of dimensions. The way a new dimension is incorporated depends on the selected classification perspective. General classification perspectives allow for only extensions. Specific classification perspectives allow for both extensions and specializations.

An extended dimension (third activity in Figure 5.2) is based on a global factor and can be reused to perform different types of GSE scenario classifications. For example, “language distance” is relevant from a general perspective, so that it could be a new “extended dimension” of the original GSE taxonomy. At the same time, such an “extended dimension” would be readily applicable for classifying focusing on many different specific classification perspectives (e.g. effort estimation). The third activity of the proposed systematic process can be also used to adapt already existing dimensions of the original GSE taxonomy, so that an existing dimension can have its representativeness augmented whenever necessary.

A specialized dimension (fourth activity in Figure 5.2) is based on a factor that is relevant just from a given specific classification perspective and cannot be readily reused to perform other different types of specific classifications. For example, an “effort estimation” role is relevant only within the
effort estimation perspective, so that it could be a new specialized dimension of the original GSE taxonomy.

5 EXTENSION FOR SMITE ET AL.’S GSE TAXONOMY

In this section we describe the extended dimensions incorporated into Smite et al.’s GSE taxonomy and how these extended dimensions interact with the original dimensions described in Section 2.

The classification perspective that guided both the extension (described in this section) and the specialization (described in Section 6) of Smite et al.’s GSE taxonomy is “effort estimation” (specific perspective). Note that the chosen classification perspective does not restrain the reuse capacity of the extended dimensions.

5.1 Selection of factors

We based the selection of the extended factors on Britto et al.’s SLR on effort estimation in GSE [21]. First, we looked at the cost drivers that were identified in more than one primary study, since the larger the number of studies supporting a cost driver, the more relevant it was judged to be. Thus, we identified four cost drivers, respectively called “time zone” (supported by four studies), “language and cultural differences” (supported by three studies), “communication” (supported by two studies) and “software process model” (supported by two studies).

Considering the outcomes of the SLR and the general literature regarding both effort estimation and GSE, we have noticed that the following aspects were not covered in Smite et al.’s GSE taxonomy:

• It is very well-known by the GSE community that language and cultural distance perform an important role in GSE projects [58]. We also identified both aspects as one of the most frequent cost drivers in our previous work [21]. Despite that both distances were discussed in Smite et al. [149], their taxonomy does not have dimensions to represent language and cultural distances.
• Since the distributed aspect of GSE projects appears to affect the effort estimation process itself, to classify GSE scenarios focusing on effort estimation it is necessary to have a dimension regarding the organization of the sites and their roles within the effort estimation process. The GSE taxonomy proposed by Smite et al. does not cover these factors (specialization, see Section 6).
• The Smite et al.’s taxonomy was developed to classify relationships between pairs of sites. Nevertheless, some aspects that can have a bearing on the effort related to GSE scenarios or on the effort esti-
mation process would be better classified in the granularity level of sites rather than the granularity level of relationships between sites.

Considering the above discussion and the needs in effort estimation, only “time zone” is originally part of the Smite et al.’s GSE taxonomy. This factor is represented by the dimension called “time distance”. To complement the original taxonomy, we selected the following factors: “language distance”, “cultural distance” and “software process model”.

Although it has been supported by two studies, we discarded the factor “communication” because rather than being a cost driver itself, it is a factor that depends on other cost drivers (e.g. temporal, language and cultural distances). To include such a factor would be redundant, since its impact is already represented by the dimensions “temporal distance”, “language distance” and “cultural distance”.

Furthermore, based on the existing literature, it was necessary to split the “cultural distance” dimension into two different dimensions, called “power distance” and “uncertainty avoidance”.

Finally, the granularity level originally considered in Smite et al.’s GSE taxonomy (relationship between sites) is not representative enough for the selected specific perspective. Thus, we included a new extended dimension called “setting” for classifying GSE scenarios from both “site” and “relationship between sites” settings. The root of the original GSE taxonomy, placed on the “GSE” dimension, was changed from “sourcing” to “project” to keep the consistency.

Figure 5.3 displays the extended GSE taxonomy and the extended dimensions are further discussed in Subsections 5.2, 5.3, 5.4 and 5.5, respectively.

The relationships between the extended dimensions and the original dimensions of Smite et al.’s GSE taxonomy are as follows:

- The dimension “GSE” is parent of all other dimensions.
- The classifications by means of the dimensions “software process”, “power distance”, “uncertainty avoidance” and “language distance” are related to the category “site” of the dimension “setting”.
- The classifications by means of the dimensions “legal entity” and “location” are related to the category “relationship” of the dimension “setting”.
- The classification by means of the dimension “geographic distance” depends on the categories of the dimension “location”.
- The classification by means of the dimension “temporal distance” depends on the categories of the dimension “geographic distance”.

It is important to emphasize that the extended dimensions are generic, i.e. independent of the software engineering perspective, and readily reusable.
Figure 5.3: Extended GSE taxonomy.
5.2 Setting

A taxonomy that is only able to classify merely relationships between sites is not sufficient to comply with the effort estimation specific classification perspective. It comes from the fact that some of the factors that we selected are site-specific. For this reason, we designed the dimension called “setting”, to allow for the classification of projects in terms of sites (Site) and their relationships (Relationship).

5.3 Software process model

“Software process model” was one of the cost drivers identified in two of the primary studies in our SLR [21]. In both studies, this factor was related to the type of software development process applied in a site, which can be agile-based, plan-based (traditional) or a combination of both approaches.

Plan-driven software development, also called plan-based or traditional development, is a software engineering approach in which the entire development process must be detailed in a plan [128]. Details like the development schedule, tasks to be accomplished and the person responsible for each task must be recorded in such a plan.

That type of software development is deemed as heavy and bureaucratic to deal with certain types of projects, specially the ones where the requirements are unclear and uncertain [43]. Therefore, the main criticism regarding plan-driven development is that many decisions that are taken early on must be reappraised later on, since software development deals with a lot of uncertainty in the early stages of a project [108]. On the other hand, this approach allows for planning organizational aspects earlier, besides fostering the discovery of potential problems before the start of a given project.

The agile software development was proposed as an answer to the needs of the software industry, which required more flexibility and efficiency to perform software projects [117]. In order to deliver software without budget and schedule overruns, agile methods, such as Scrum [123], Extreme programming (XP) [12], keep close collaboration between development teams and customers and respond quickly to needs for change in an incremental-adaptive fashion [128]. Differently from plan-driven approaches, in agile approaches the functionality is not planned in advance. Rather, it is planned incrementally.

Although it is argued that agile methods are more suitable to deal with projects that present unclear and uncertain requirements, they also have drawbacks. For example, to be successful, an “agile software project” must have close collaboration of the customer and the development team [12]. Besides, organizations and customers can be more familiar with plan-
driven approaches and can find it hard to trust and follow an agile-based approach [46]. Pure agile-based software processes are difficult to scale. They are more adequate to small and medium size projects [46].

Evidence shows that the industry applies software processes that include practices from both plan-based and agile-based software development methods. The state of the usage of each paradigm in a given company depends on factors e.g. the experience of the team members of the paradigm and the size of the project [128].

In GSE projects, the type of the software development process used is an aspect that impacts upon the effort required to perform a project [21]. Two studies evaluated the state of the use of agile methods in GSE projects [63, 69]. Both studies identified that agile practices are not readily applied to GSE projects, requiring some adaptations to be adequate to such situations.

In general, the adaptations were related to the planning of increments and the way communication should be performed between sites. These adaptations bring "agile teams" closer to plan-driven practices.

Another point that must be discussed is the difference between the software process used by different sites. It is possible that in a given GSE project the involved sites apply different software processes, which could be mainly either agile or plan-driven. We believe that such kind of difference could lead to problems in the communication and loss of trust between sites [114].

Considering the discussion above, we define the dimension “software process model” as having two categories:

- **Agile** - The software process used in a given site is mainly based on agile practices.
- **Plan-driven** - The software process used in a given site is mainly based on plan-driven practices.

As aforementioned, because of the nature of GSE projects, it is often the case that many companies employ both agile and prescriptive practices when carrying out GSE projects. Therefore, when classifying the sites of a project using the dimension described in this subsection, one must analyze each site’s software process and decide which category best suits in each situation; if the majority of the practices employed in software process are agile, then the site should be classified as agile; otherwise it should be classified as plan-driven.

5.4 Cultural distance

According to Hofstede et al. [62], “culture is the collective programming of the mind, which distinguishes the members of one group or category of people from another”. It is based on beliefs, values, assumptions, rituals and behaviors inherent to groups of people, which guide individuals.
There are many levels of culture that define the mental “programming” of people. The national level is related to the culture of the country where someone was born or lives, whereas the regional level is related to the different cultures that can exist inside a country. The organizational level is connected to the culture of a company that employs someone.

We identified in Britto et al’s SLR “cultural distance” as one of the most frequent cost drivers [21]. The culture level that affects GSE projects the most is the national level, since the presence of different countries brings different national cultures. The organizational level is also important, but it is influenced by the national level, and the national culture tends to have greater impact in global scenarios than the organizational culture [62]. For this reason, we focused upon the national level of culture and its impact on the organizational level.

The core of a culture is composed by values, which are subjective measures about people’s preferences comprising plus and minus sides (e.g. evil x good, dirty x clean and irrational x rational). This characteristic makes the comparison between cultures difficult to be performed, specially at the national level.

Many approaches have been proposed to compare national cultures [127, 136, 64, 62]. However, Hofstede’s national culture framework [62] is the most evaluated and applied approach, even in the GSE context [68, 2, 60].

Hofstede’s national culture framework was originally developed aiming at understanding how different national cultures managed four different anthropological issues: inequality, uncertainty, relationship between individual and primary groups, gender impact. That original framework was based on survey data that had values related to people from fifty countries and three multi-country regions. The survey population encompassed IBM employees. Afterwards, the framework was widened by the inclusion of new countries and two more dimensions. Ninety countries and three multi-country regions have scores regarding the six dimensions. The current dimensions of Hofstede’s framework are as follows:

- **Power distance (PDI)** - Measures how people manage inequality in hierarchical relationships, i.e. boss-subordinates. In nations with high PDI, the employees depend more on the bosses to take decisions. On the other hand, in nations with low PDI, the competences of the employees are more considered than their hierarchical position.

- **Uncertainty avoidance (UAI)** - Measures how people manage uncertainty, how they feel threatened by uncertain situations and try to avoid or mitigate such situations. In nations with strong uncertainty avoidance, they have strict laws and rules. On the other hand, nations with weak uncertainty avoidance have as few rules as possible, which make their people more tolerant to uncertain situations.

---

6 www.ibm.com
• **Individualism (IDV)** - Measures the trend people have to value more the individual as opposed to the group. In individualist nations, everyone cares more about himself/herself and him/her immediate family. Relationships between employer and employee are based on mutual advantage. On the other hand, in collectivist nations, people are integrated into strong and cohesive in-groups relationships, with an unquestioning sense of loyalty. Relationships between employer and employee are based in moral terms, resembling family links.

• **Masculinity (MAS)** - Measures the dominant values of a society in terms of the impact of each gender, which can be “masculine” (e.g. to be more assertive and competitive) or “feminine” (e.g. to be more modest, tender, and concerned about life quality) values. “Feminine” nations tend to have people that “work to live”. In those nations, managers struggle to reach consensus. On the other hand, in “masculine” nations people use to “live to work” and managers are expected to behave in a decisive and assertive way.

• **Long-term orientation (LTO)** - Measures the trend people have to value more virtues related to future rewards (e.g. saving, persistence, and adapting to changes) or to the past/present virtues (e.g. national pride, respect for tradition and fulfilling social obligations). Nations that look to the future are called “long-term oriented”. On the other hand, nations that look to the past and present are called “short-term oriented”.

• **Indulgence versus restraint (IVR)** - Measures people’s degree of control regarding their desires and impulses. “Indulgent” nations tend to allow free gratification of human behaviors related to enjoying life and having fun. “Restrained” nations tend to regulate such gratifications by strict social norms.

Each of the 93 countries/multi-countries region evaluated has its own score for each one of the above-described dimensions. The dimensions were designed to be scored from 1 to 100. However, according to Hofstede et al. [62], it was required to extrapolate the previously defined scale, so that there are some countries that scored above 100.

The scores are relative and just make sense when used to compare different national cultures. Another important point is that the relative scores are quite stable over the years, due the fact that when a national culture shifts, the force that causes such change has global or continental impact, keeping stable the relative position between national cultures.

Organizations are structured to match the cultural needs of their members. Considering multinational companies, e.g. IBM, is expected that their organizational cultures are influenced by the national cultures of their headquarters’ countries. However, since such kind of company has many subsidiaries around the world, it is important to understand their branches’
national cultures to improve their employees’ environment and productivity [62].

The way employees think about their employers is directly affected by their national culture. Hofstede et al. [62] stated that “organizing” is related to answer two questions:

1. “Who has the power to decide what?”
2. “What rules or procedures will be followed to attain the desired ends?”

The answer for first question is directly connected to the PDI of the employees’ national culture. The second one is influenced by the UAI. There is empirical evidence supporting only PDI and UAI as the dimensions that affect the organizational culture the most [62]. Figure 5.4 shows the relationship between those two dimensions.

In GSE projects, the way decisions and the required work are performed can vary between sites, impacting negatively on the communication and trust [33]. Since projects are performed at the organizational level, we decided to include just the dimensions from Hofstede’s framework that have empirical evidence supporting their influence on the culture’s organizational level, i.e. PDI and UAI. In addition, only PDI and UAI were empirically categorized.

Therefore, the original dimension called cultural distance was divided in two different dimensions, called power distance (PD) and uncertainty avoidance (UA).

Our taxonomy’s PD dimension has the following categories:

• Small - Sites placed in countries with PDI \(\leq 50\) have small power distance.
• Large - Sites placed in countries with PDI \(> 50\) have large power distance.

Our taxonomy’s UA dimension has two categories as follows:

• Weak - Sites placed in countries with UAI \(\leq 63\) have small power distance.
• Strong - Sites placed in countries with UAI \(> 63\) have large power distance.

The threshold values used to differentiate sites with Small or Large PD and Weak or Strong UA were defined based on Hofstede et al.’s empirical study [62]. To choose the proper UA and PD categories for a site, one must consult the UAI and PDI scores for the countries involved in a project under classification, which are available in Hofstede et al. book [62]. For the sake of space, Table 5.1 only displays the PDI and UAI scores and PD
Figure 5.4: Relationship between PDI and UAI dimensions (Adapted from Hofstede et al. [62]).
Table 5.1: PDI and UAI scores and PD and UA classifications.

<table>
<thead>
<tr>
<th>Country</th>
<th>PDI</th>
<th>PD</th>
<th>UAI</th>
<th>UA</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>40</td>
<td>Small</td>
<td>46</td>
<td>Weak</td>
</tr>
<tr>
<td>Great Britain</td>
<td>35</td>
<td>Small</td>
<td>35</td>
<td>Weak</td>
</tr>
<tr>
<td>India</td>
<td>77</td>
<td>Large</td>
<td>40</td>
<td>Weak</td>
</tr>
<tr>
<td>China</td>
<td>80</td>
<td>Large</td>
<td>40</td>
<td>Weak</td>
</tr>
<tr>
<td>Malaysia</td>
<td>104</td>
<td>Large</td>
<td>36</td>
<td>Weak</td>
</tr>
<tr>
<td>Japan</td>
<td>54</td>
<td>Large</td>
<td>92</td>
<td>Strong</td>
</tr>
<tr>
<td>Taiwan</td>
<td>58</td>
<td>Large</td>
<td>69</td>
<td>Strong</td>
</tr>
<tr>
<td>Ireland</td>
<td>28</td>
<td>Small</td>
<td>35</td>
<td>Weak</td>
</tr>
<tr>
<td>Brazil</td>
<td>69</td>
<td>Large</td>
<td>76</td>
<td>Strong</td>
</tr>
<tr>
<td>Slovakia</td>
<td>104</td>
<td>Large</td>
<td>51</td>
<td>Weak</td>
</tr>
<tr>
<td>Finland</td>
<td>33</td>
<td>Small</td>
<td>59</td>
<td>Weak</td>
</tr>
<tr>
<td>Germany</td>
<td>35</td>
<td>Small</td>
<td>65</td>
<td>Strong</td>
</tr>
<tr>
<td>Norway</td>
<td>31</td>
<td>Small</td>
<td>50</td>
<td>Weak</td>
</tr>
<tr>
<td>Sweden</td>
<td>31</td>
<td>Small</td>
<td>29</td>
<td>Weak</td>
</tr>
<tr>
<td>Pakistan</td>
<td>55</td>
<td>Large</td>
<td>70</td>
<td>Strong</td>
</tr>
</tbody>
</table>
and UA classifications for the countries identified in the SLR conducted by Britto et al. [21].

Note that the PD and UA classification outcomes for sites must be compared. For example, let us consider a project with two sites, respectively named X and Y. Site X is placed in Germany and site Y is placed in Brazil, i.e. X’s PD is Small and UA is Strong, while Y’s PD is Large and UA is Strong. In this example, there is no much concern about the impact of UA on the GSE project, since both sites are classified in the same category. However, PD could negatively impact the GSE project, since the sites are classified in different categories.

In some situations, companies source human resources from different countries to compose teams. It means that the main national culture of a given site is not necessarily the same national culture of the country wherein such site is placed. For example, Ramasubbu and Balan [112] report a project with two sites, one placed in the USA and the other one located in India. Although both sites are placed in countries with different PD and UA, the human resources of both sites were sourced in India. In such situation, the actual cultural distance between the two sites is expected to be zero.

Therefore, it is important to account for the predominant nationality of the human resources of a given site to define the appropriate PD and UA.

5.5 Language distance

In a GSE project, it is very likely that involved sites do not have the same native language, which may lead to misunderstandings between sites and generate delays in the entire project [4].

Nowadays, English is the most used language when there is need for a “lingua franca” [86]. A “lingua franca” is a language used between people who do not have the same native language. For example, let us consider a GSE project with two sites, one placed in Sweden and the other one placed in China. It is more realistic to think that both sites will communicate in English instead of speaking either Swedish or Chinese.

Thus, instead of calculating the distance between sites’ languages, we measure the distance between each site’s language and English. In our extended dimension, language distance means the difficulty to learn/speak English given a site’s native language.

We used herein the proposal from Chiswick et al. [27], which measures quantitatively the distance between a language and English. This study was based on Hart-Gonzalez et al. [55].

Hart-Gonzalez et al. [55] present an evaluation performed with people who were born in USA. The main goal of this study was to evaluate the difficulty for native English-speaking Americans to learn a different lan-
guage. Different groups of Americans were trained in different languages for a period of 24 weeks. After the training period, the proficiency of the subjects in the trained language was evaluated.

Based on the evaluation of the subjects, a number called language score ($L_s$) was attributed to each language. The language score could be 1, 1.25, 1.5, 1.75, 2, 2.25, 2.5, 2.75 or 3. The bigger the language score, the easier to the English speaker to learn the evaluated foreign language.

Chiswick et al. [27] used the scores calculated in Hart-Gonzalez et al. [55] to evaluate the difficulty that immigrants from non-English-speaking countries faced to learn English in USA and Canada. Hence, rather than using $L_s$ to identify the language distance between a given language and English, they decided to use the inverse of $L_s$ (Equation 5.1), which they called $L_d$ (language distance). Therefore, the possible values for $L_d$ were 0.33, 0.36, 0.4, 0.44, 0.5, 0.57, 0.67, 0.8 and 1.

$$L_d = \frac{1}{L_s} \quad (5.1)$$

The bigger the $L_d$ value, the farther a given language $L$ is from English, which is also a measure of how difficult it is for people who speak $L$ to learn to speak English. Thus, the larger the $L_d$, the higher the likelihood that the proficiency in English will not be very good [27]. The lower the level of proficiency in English, the higher the probability of problems regarding the communication between sites [58].

Chiswick et al. presented a table containing scores for a large number of languages [27]. Table 5.2 is a fragment of the original table, including only languages spoken in countries we have identified in our previous SLR [21].

Based on the findings from Chiswick et al. [27], we defined the dimension “language distance”, which comprises four categories:

- **No distance** - When the mother language of a site is English or no “lingua franca” is required, there is no language distance in such site, since people from both sites could communicate in their native tongue.
- **Small** - When $0 < L_d \leq 0.4$, the language distance of a site is considered small. It means that it is more likely that people from such site have an acceptable level of proficiency in English, since it is easier for them to learn it.
- **Medium** - When $0.4 < L_d \leq 0.57$, the language distance of a site is considered medium. It means that it is more likely that people from such site struggle somewhat to learn English, which affects the proficiency. However, they can learn and speak English by applying more effort than people from the previous group.
- **Large** - When $0.57 < L_d \leq 1$, the language distance of a site is considered large. It means that it is more likely that people from such site
Table 5.2: Language scores and distances

<table>
<thead>
<tr>
<th>Country</th>
<th>Main language</th>
<th>Language</th>
<th>Score</th>
<th>Language distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>English</td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Great Britain</td>
<td>English</td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>India</td>
<td>Hindi</td>
<td></td>
<td>1.75</td>
<td>0.57</td>
</tr>
<tr>
<td>China</td>
<td>Mandarin</td>
<td></td>
<td>1.5</td>
<td>0.67</td>
</tr>
<tr>
<td>Malaysia</td>
<td>Malay</td>
<td></td>
<td>2.75</td>
<td>0.36</td>
</tr>
<tr>
<td>Japan</td>
<td>Japanese</td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Taiwan</td>
<td>Mandarin</td>
<td></td>
<td>1.5</td>
<td>0.67</td>
</tr>
<tr>
<td>Ireland</td>
<td>English</td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Brazil</td>
<td>Portuguese</td>
<td></td>
<td>2.5</td>
<td>0.4</td>
</tr>
<tr>
<td>Slovakia</td>
<td>Slovak</td>
<td></td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>Finland</td>
<td>Finnish</td>
<td></td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>Germany</td>
<td>German</td>
<td></td>
<td>2.25</td>
<td>0.44</td>
</tr>
<tr>
<td>Norway</td>
<td>Norwegian</td>
<td></td>
<td>3</td>
<td>0.33</td>
</tr>
<tr>
<td>Sweden</td>
<td>Swedish</td>
<td></td>
<td>3</td>
<td>0.33</td>
</tr>
<tr>
<td>Pakistan</td>
<td>Urdu</td>
<td></td>
<td>1.75</td>
<td>0.57</td>
</tr>
</tbody>
</table>
struggle even more to learn English. In general, those languages have almost no commonalities with English, which requires more effort to learn English.

The first category of this dimension ("No distance") was designed to comply with the first aforementioned limitation. The other three categories were defined by dividing the language distance scale in three different equal parts ("Small", "Medium" and "Large"), so that there is enough representativeness to classify the existing spectrum of language distance values.

Note that this dimension can be used only in projects that require no lingua franca to enable communication between sites, i.e. there is no language distance, or when the chosen lingua franca is English.

There are some particularities regarding some countries that are former British colonies, e.g. India and Pakistan. In these countries there is a considerable number of people who have a good proficiency in English. The language distance dimension of our taxonomy focuses on the language that is spoken the most in a given country. We did so because it would be very difficult to embrace the particularities of countries that have more than one official language. Thus, when using this dimension to classify the language distance of each site, one must identify the language spoken the most by the team of each site and used it as basis for selecting the language distance category that fits better.

6 SPECIALIZED GSE EFFORT ESTIMATION TAXONOMY

In this section we describe the specialized dimensions incorporated into Smite et al.’s GSE taxonomy and how these specialized dimensions interact with the extended dimensions described in Section 5 and with the original dimensions described in Section 2.

6.1 Selection of factors

Effort estimation, i.e. the process used to predict the effort needed to fulfill a given task in software projects [91], is affected by global challenges (for example, communication and coordination) [113]. Therefore, it is important to consider the impact of global challenges on the effort estimation process and on the obtained effort estimates. To do so, it is necessary to understand the effort estimation related factors affected by global challenges.

The factors related only to the effort estimation process are not eligible to be readily applicable in other specific classification perspectives. Therefore, such factors were incorporated into the original GSE taxonomy as specialized dimensions, as follows: “estimation stage”, “estimation process role”
and “estimation process architectural model”. The specialized GSE effort estimation taxonomy is presented in Figure 5.5.

Figure 5.5: Specialized GSE effort estimation taxonomy.

Two of the dimensions (“estimation process role” and “estimation stage”) were defined by using the extensive knowledge of one of the authors in effort estimation [95, 76, 94]. In addition, we used architectural models
from Distributed Systems [31] as inspiration to describe the “estimation process architectural model” dimension, since this area shares some basic concepts and challenges with GSE. The specialized dimensions are further discussed in Subsections 6.2, 6.3 and 6.4 respectively.

The relationships between the specialized dimensions, the extended dimensions and the original dimensions of Smite et al.’s GSE taxonomy are as follows:

- The dimension “GSE” is parent of all other dimensions.
- The classifications by means of the dimensions “estimation stage” and “estimation process role” are related to the category “site” of dimension “setting”.
- The classification by means of the dimension “estimation process architectural model” is related to category “relationship” of dimension “setting”.

Note that the specialized dimensions are dependent of the effort estimation perspective and therefore are not expected to be readily reusable.

6.2 Estimation process role

Considering the activities that are part of the effort estimation process, a site can play different roles. We defined the following possible categories:

- **Estimator** - A site plays this role when it is responsible for applying an effort estimation technique to obtain effort estimates.
- **Supplier** - A supplier site is the one that is not able or allowed to obtain effort estimates by itself. Thus, it should supply historical data or expert’s knowledge to an estimator site. Eventually, a site that plays this kind of role could give feedback on the obtained effort estimates.
- **Estimator/Supplier** - A site of this type is capable of calculating its own effort estimates, considering its own historical data and/or expert’s knowledge. However, to understand the overall effort related to a given project, the effort estimates obtained by sites of this kind have to be combined.

6.3 Estimation stage

During the lifetime of a given software project, the effort estimates can be calculated or refined many times. Considering the defined roles that a site can play into a given effort estimation process, it is also important to describe when historical data from past projects or knowledge from experts are collected and when the effort estimates are obtained. The difficulty to
perform these activities can vary depending on the moment in time they are carried out.

Thus, we have defined three possible stages upon when a GSE project when data or knowledge collection and effort estimates’ attainment can be carried out by a given site, which led to the following categories:

- **Early** - When the effort estimation process is performed just after requirements’ elicitation, we call it early effort estimation process.
- **Late** - When the effort estimation process is performed after requirements’ analysis and design, we call it late effort estimation process. In general, this type of estimation is expected to be performed to refine early effort estimates.
- **Early/Late** - A combination of both above-mentioned types.

### 6.4 Estimation process architectural model

The purpose of estimating effort as part of managing a project is to predict the amount of persons-time required to accomplish the set of tasks needed as part of a project’s life cycle, based on a set of inputs such as the knowledge/data of previous similar projects and other application and project characteristics that are believed to be related to effort [91].

A generic effort estimation process encompasses four different activities, as follows [91]:

- **Activity 1** - Collect data and/or knowledge from similar finished projects.
- **Activity 2** - Build an estimation model based on the collect data/knowledge.
- **Activity 3** - Estimate the size and determine the cost drivers’ values of the new software project.
- **Activity 4** - Estimate the effort using the estimated size and determined cost drivers’ values as inputs.

Effort estimation techniques can be classified as expert-based, algorithmic or artificial intelligence [126]. The sequence in which the aforementioned activities are executed into the effort estimation process depends on the type of the effort estimation technique. In addition, the activity 2 is not explicitly executed by expert-based effort estimation techniques, since the estimates are subjectively obtained by this kind of technique [91].

On one hand, there is no major concern about the way in which the activities of the effort estimation process are coordinated and performed in collocated projects, since all the steps are performed in the same place. On the other hand, the distributed nature of GSE projects allows for different interaction manners between sites, which can affect entirely the effort estimation process.
To the best of our knowledge, there is no existing classification of possible collaboration patterns when estimating effort in GSE projects. Therefore, we combined effort estimation and the distributed systems literature to design this dimension of our taxonomy.

According to Coulouris et al. [31], a distributed system is defined as “a system in which hardware or software components located at networked computers communicate and coordinate their actions only by passing messages”. The computers of a distributed system can be positioned in the same room, in the same building, in different countries or even in different continents. If we replace the word “computer” by “teams”, the definition of distributed systems becomes very similar to the definition of GSE.

In distributed systems, an architectural model describes the structure of a system in terms of its components and their relationships. There are two main architectural models: Client-server and peer-to-peer.

In the client-server architectural model (Figure 5.6), client computers consume the resources managed by server computers. Each computer in a given distributed system is called an element. The roles of both client and server elements are well-determined, despite that in some cases a server can be a client from the point of view of another server. In general, when a distributed system is built based on the client-server model, the server elements have better hardware resources than the client ones.

Client-server systems are easier to construct and manage. However, it is difficult to scale systems based on this model, since the services (or “expensive” tasks) are supplied by just a few system elements.

Figure 5.6: Client-server model (Adapted from Coulouris et al. [31]).

In the peer-to-peer architectural model (Figure 5.7), all elements (computers) of a distributed system play similar roles, working in a cooperative way to fulfill tasks. There is no such distinction between client and server
elements, like in the client-server model. Although peer-to-peer systems tend to be harder to construct and manage, they are highly scalable, since each element can carry out “expensive” tasks.

Figure 5.7: Peer-to-peer model (Adapted from Coulouris et al. [31]).

The effort estimation process architectural model dimension of our taxonomy was inspired by the above-described concepts from the distributed systems knowledge area. Thus, we defined three different architectural models, which led to three different dimensions in this dimension, respectively called centralized, distributed and semi-distributed.

The **centralized** model (Figure 5.8) is based on the client-server distributed systems’ model. Considering a given relationship between pairs of sites, one of the sites plays a role as a “server” (site x), centralizing all the effort estimation process, in a top-down fashion. The other site of the pair can be seen as a “client” (site y), being able of just using the effort estimates calculated by the “server” site. Eventually, the client site can supply complementary data to support the effort estimation.

The **distributed** model (Figure 5.9) is based on the peer-to-peer distributed systems’ model. Considering a given relationship between two sites, both sites (sites x and y) are responsible for calculating their own effort estimates based on the local data or knowledge, in a bottom-up fashion. The effort estimates calculated in each site are later on combined and shared by the “integrator” site (site x), to allow both sites to understand the overall effort inherent to the project.

The **semi-distributed** model (Figure 5.10) is based on both distributed systems’ models. In this model, a “server” site (site x) is responsible for calculating all the effort estimates regarding both sites of the pair. However, differently of the centralized model, the client site (site Y) can provide
Figure 5.8: Centralized model.

Figure 5.9: Distributed model.
feedback about the calculated effort estimates, so that the “server” site could perform further refinements. Likewise the centralized model, the client site can eventually supply complementary data to support the effort estimation.

Figure 5.10: Semi-distributed model.

7 CLASSIFYING GSE SCENARIOS FOCUSING ON EFFORT ESTIMATION

In this section we present the classification process and notation that guide the usage of the proposed specialized GSE effort estimation taxonomy.

7.1 Classification process

The subject matter to be classified using the GSE specialized taxonomy is a GSE project. Herein, we consider that a GSE project has sites (at least two) and relationships between sites (at least one). The first activity of the process displayed in Figure 5.11 produces as outcome a collaboration diagram. Collaboration diagrams present graphically the interplay (edges) between the sites (nodes) of GSE projects. The layout of a given collaboration diagram is strictly dependent on the effort estimation process architectural model employed in the GSE project under classification.
A collaboration diagram is further described by means of activities two and three. As a result of such activities, the sites and sites’ relationships of a given GSE project are classified, according to the GSE specialized taxonomy dimensions.

![Collaboration Diagram](image)

**Figure 5.11: Classification process.**

### 7.2 Collaboration diagram notation

We defined a notation to systematically draw collaboration diagrams, which is displayed in Figure 5.12. The notation is composed by two different types of nodes (“central node” and “peripheral node”) and four different types of edges (“centralized edge”, “distributed edge”, “semi-distributed edge” and “no-interaction edge”).

#### 7.2.1 Nodes

The role of “central nodes” in the collaboration diagram changes according to the employed effort estimations process architectural model (Section 6). In both centralized and semi-distributed architectural models, a central node represents the server site. In the distributed architectural model, such kind of node represents the integrator site. A collaboration diagram should have only one central node.
The role of “peripheral nodes” in the collaboration diagram also changes according to the employed effort estimation process architectural model (Section 6). In both centralized and semi-distributed models, peripheral nodes represent client sites. In the distributed model, such kind of node represents an ordinary server site. A collaboration diagram should have at least one peripheral node.

Regardless of its type, a node must be identified using rather capitalized letters from the Latin standard alphabet. Whenever required, capitalized letters from the Greek alphabet should be used to complement the Latin alphabet.

### 7.2.2 Edges

In a collaboration diagram, an edge must exist only between sites that have some kind of relationship in the context of the project. A “no-interaction edge” must be used when there is relationship between two sites, but not in the context of effort estimation. The other three edge types are directed related to the employed effort estimation process architectural model. “Centralized edges” must be used to connect client sites to the server site in a centralized model. “Distributed edges” must be used to connect ordinary server sites to the integrator site in a distributed model. Finally, “semi-distributed edges” must be used to connect client sites to the server site in a semi-distributed model.

Regardless of the type of the edge, its identifier is composed by the combination of its nodes’ identifiers.

![Diagram](image_url)
To illustrate the usage of the specialized GSE effort estimation taxonomy (and extended dimensions), we selected the five primary studies included in Britto et al.’s SLR ([105, 15, 100, 102, 113]). The objective was to classify the empirical cases (GSE projects) reported in each one of those studies. Nevertheless, the data available in such papers did not fully enable the classification of the reported GSE projects using the specialized GSE taxonomy proposed herein. Thus, we contacted the authors of those five primary studies to ask for the additional data required to classify the GSE projects reported in such studies.

We only obtained an answer from the authors of Ramasubbu et al. ([113]), who provided additional data on eight different GSE projects. The data was originally gathered from four large software companies. The performed projects were on new software development or high scale maintenance in the domain of financial services, retail, manufacturing and telecommunications industry. The projects were executed by the companies between 1999 and 2009.

Due to a non-disclosure agreement with the companies, the name of the companies and the domain of the software applications developed/main- tained were not provided. The cities wherein the sites were placed were also not explicitly stated because of the non-disclosure agreement. Rather, they presented the location of sites in terms of countries. In the case of sites placed in the USA, the region of each site was also provided. In addition, the authors also clarified that the team members of all projects were fluent in English. Thus, we considered that there was no language distance between the sites of the projects.

We classified all the eight GSE projects provided by Ramasubbu et al. ([113]) to illustrate the usage of the specialized GSE effort estimation taxonomy, along with the extended dimensions. The classification for two of the projects are presented in Subsections 8.1 and 8.2 respectively. Those two projects were selected because they have diverse settings and enable for presenting different classification possibilities. The classification for each one of the other six projects can be found in Appendix 1.

8.1 Project 1

Project 1 was performed by seven different sites. The project used a semi-distributed effort estimation process architectural model. Both early and late effort estimation were performed. Site A was located in the USA’s east coast and applied an agile-based software process (server site). Site B was located in the USA’s west coast and applied an agile-based software
Sites C and D were both located in India and applied a plan-driven software process. Site E was located in Singapore and applied a plan-driven software process. Site F was located in Germany and applied an agile-based software process. Site G was located in Australia and applied an agile-based software process. All the sites only interacted with Site A. The collaboration diagram of project 1 is displayed in Figure 5.13.

![Collaboration diagram of project 1](image)

Figure 5.13: Collaboration diagram of project 1.

The result of the sites’ classification for project 1 is as follows:

- **Site A**
  - **Software process** - Agile.
  - **Power distance** - Small.
  - **Uncertainty avoidance** - Weak.
  - **Language distance** - No distance.
  - **Estimation stage** - Early/Late.
  - **Estimation process role** - Estimator/Supplier.

- **Sites B and G**
  - **Software process** - Agile.
  - **Power distance** - Small.
  - **Uncertainty avoidance** - Weak.
  - **Language distance** - No distance.
  - **Estimation stage** - Early/Late.
  - **Estimation process role** - Supplier.

- **Sites C, D and E**
  - **Software process** - Plan-driven.
– Power distance - Large.
– Uncertainty avoidance - Weak.
– Language distance - No distance.
– Estimation stage - Early/Late.
– Estimation process role - Supplier.

• Site F
  – Software process - Agile.
  – Power distance - Small.
  – Uncertainty avoidance - Strong.
  – Language distance - No distance.
  – Estimation stage - Early/Late.
  – Estimation process role - Supplier.

The result of the relationships’ classification for project 1 is as follows:

• Relationship AB
  – Location - Onshore.
  – Legal entity - Insourcing.
  – Geographic distance - Distant.
  – Temporal distance - Different.

• Relationships AC, AD, AE, AF and AG
  – Location - Offshore.
  – Legal entity - Insourcing.
  – Geographic distance - Far.
  – Temporal distance - Large.

8.2 Project 2

All the four sites of Project 2 were placed in the USA, respectively in the east coast, north east, mid-west and west coast. The project used a distributed effort estimation process architectural model. Only early effort estimation was performed. All the sites employed an agile-based software process. Site A played the role as the integrator site. All sites related to each other in such project. The collaboration diagram of project 2 is displayed in Figure 5.14.

The result of the sites’ classification for project 2 is as follows:

• Sites A, B, C and D
  – Software process - Agile.
  – Power distance - Small.
  – Uncertainty avoidance - Weak.
Figure 5.14: Collaboration diagram of project 2.

- **Language distance** - No distance.
- **Estimation stage** - Early.
- **Estimation process role** - Estimator/Supplier.

The result of the relationship classification for project 2 is as follows:

- **Relationship AB**
  - **Location** - Onshore.
  - **Legal entity** - Insourcing.
  - **Geographic distance** - Distant.
  - **Temporal distance** - Similar.
  - **Estimation process architectural model** - Distributed.

- **Relationships AC and AD**
  - **Location** - Onshore.
  - **Legal entity** - Insourcing.
  - **Geographic distance** - Distant.
  - **Temporal distance** - Different.
  - **Estimation process architectural model** - Distributed.

- **Relationships BC and CD**
  - **Location** - Onshore.
  - **Legal entity** - Insourcing.
– Geographic distance - Distant.
– Temporal distance - Similar.
– Estimation process architectural model - No interaction.

• Relationship BD
– Location - Onshore.
– Legal entity - Insourcing.
– Geographic distance - Distant.
– Temporal distance - Different.
– Estimation process architectural model - No interaction.

9 SUMMARY AND LIMITATIONS

The research questions outlined in Section 3 are addressed as follows.

9.1 RQ1 - How can a general GSE taxonomy be extended and specialized?

RQ1 was addressed by means of the process depicted in Section 4. By applying this process, we were able to develop the required extension and specialization, which are consistent and concise with the original taxonomy. Therefore, the terminology defined by Smite et al. [149] was preserved.

We believe that there is no major limitation related to the designed process; it is straightforward to extend and specialize the original taxonomy. Thus, this process can be used as basis for extending the general aspects of the taxonomy Smite et al.’s and specializing the taxonomy for other specific perspectives. Nevertheless, other studies must be conducted to verify the effectiveness of the proposed process in different SE perspectives.

9.2 RQ2 - What dimensions can be incorporated as extensions in Smite et al.’s taxonomy?

RQ2 was addressed by incorporating five new general dimensions into Smite et al.’s taxonomy, as depicted in Section 5 (“setting”, “software process model”, “power distance”, “uncertainty avoidance” and “language distance”). These dimensions were deemed as necessary to classify GSE scenarios focusing on effort estimation (although the dimensions are general), but they do not represent an exhaustive list. Therefore, we believe that different SE perspectives can demand additional dimensions.
9.3 **RQ3 - What dimensions can be incorporated as specializations in Smite et al.’s taxonomy focusing on the effort estimation perspective?**

RQ3 was addressed by incorporating three new specialized dimensions into Smite et al.’s taxonomy, as explained in Section 6 (“estimation stage”, “estimation process role” and “estimation process architectural model”). In addition, a classification process and a notation were provided in Section 7 to guide the usage of the designed specialized GSE effort estimation taxonomy. Although the provided notation was originally designed to illustrate the collaboration between sites in the context of effort estimation, it is equally possible to have a general overview of the overall collaboration between the sites of a GSE project by means of the collaboration diagram proposed in this paper.

9.4 **RQ4 - What is the usefulness of a specialized effort estimation GSE taxonomy?**

RQ4 was addressed by classifying eight finished GSE projects, each one of them with different GSE and effort estimation process configurations (Section 8 and 1). The illustration shows that the specialized GSE taxonomy has enough representativeness to classify different context focusing on effort estimation in GSE. The proposed notation for designing the collaboration diagrams is able to represent different effort estimation process configurations.

The usefulness of a consistent and concise classification scheme and terminology became even more clear when we conducted the illustration of the GSE specialized effort estimation taxonomy; the presence of a common terminology and classification scheme provides a *lingua franca* that can be used by researchers to report their findings in a way that becomes easier to analyze, compare and aggregate their findings. Since this paper is the first endeavor towards a GSE specialized effort estimation taxonomy, the existing studies on effort estimation in GSE does not report many relevant aspects, e.g. effort process architectural model. For this reason, to enable the classification of the empirical cases presented in Section 8 and 1, we had to contact the authors of the selected papers.

9.5 **Limitations**

As most studies, the research reported herein comes with some limitations:

- Despite the extensive literature review that grounded the extension and specialization process, there are just few studies on effort estimation in the GSE context [21]. It means that the existing evidence, which was the basis for the new dimensions, is not very strong.
Although some of the authors of this paper have extensive background in effort estimation or GSE, we believe that more feedback from other experts can contribute further to the correctness and usefulness of the specialized GSE taxonomy.

It was not possible to illustrate the centralized effort estimation process architectural model, because the data on finished GSE projects used to illustrate the specialized GSE taxonomy did not cover this case.

In spite of the above-mentioned limitations, we believe that the usage of both the extended and specialized dimensions will help researchers and practitioners to report new findings on effort estimation in GSE, which will facilitate the analysis, comparison and aggregation of new studies.

10 CONCLUSIONS

This paper contributed by extending the GSE taxonomy proposed by Smite et al. [149]. Furthermore, it presented a specialized GSE effort estimation taxonomy, which was designed by specializing the extended general GSE taxonomy by Smite et al. [149]. Both the extended and specialized dimensions proposed in this paper are consistent and concise with the original taxonomy, i.e. the original terminology proposed by Smite et al. was preserved; new terms were incorporated to the original terminology, without changing the meaning of the original terms.

The designed extension and specialization were based on a literature review and expert opinion. As a result, five extended and three specialized dimensions were incorporated into the original GSE taxonomy. In addition, a classification process and a notation were defined to guide the usage of the specialized GSE effort estimation taxonomy proposed in this paper.

To illustrate the usage of the specialized GSE taxonomy, eight finished GSE projects were classified. The specialized taxonomy was representative enough to classify the projects in a clear and simple manner. Nevertheless, considering the amount of possible scenarios, it was not possible to illustrate all possible GSE configurations based on the classified projects.

The GSE specialized effort estimation taxonomy has some limitations, such as the weak evidence in the state of the art of effort estimation in GSE (used as basis to select incorporated factors) and the reduced number of experts who participated during its design.

We believe that further research can be conducted to develop and evaluate an instrument to enable the use of the specialized GSE effort estimation taxonomy in the software industry. By doing so, practitioners could better understand the different GSE scenarios, so that it may help them to optimize their effort estimation processes and techniques.
In addition, more studies can be performed to extend or specialize Smite et al.’s taxonomy focusing on different SE perspectives, e.g. as software testing.

1 APPENDIX - PROJECTS’ CLASSIFICATION

In this appendix we present the classification for each of the six remaining projects, as discussed in Section 8.

1.1 Project 3

Project 3 was performed by four different sites. The project used a semi-distributed effort estimation process architectural model. Both early and late effort estimation were performed. Site A was located in India and applied a plan-driven software process (server site). Sites B, C and D were located in the USA, respectively in the east coast, mid-west and west coast. An agile-based software process was employed by all the sites. All the sites only interacted with Site A. The collaboration diagram of project 3 is displayed in Figure 15.

![Collaboration diagram of project 3.](image)

The result of the sites’ classification for project 3 is as follows:

- **Site A**
  - **Software process** - Plan-driven.
  - **Power distance** - Large.
  - **Uncertainty avoidance** - Weak.
  - **Language distance** - No distance.
  - **Estimation stage** - Early/Late.
  - **Estimation process role** - Estimator/Supplier.
• Sites B, C and D
  – Software process - Agile.
  – Power distance - Small.
  – Uncertainty avoidance - Weak.
  – Language distance - No distance.
  – Estimation stage - Early/Late.
  – Estimation process role - Supplier.

The result of the relationships’ classification for project 3 is as follows:

• Relationships AB, AC and AD
  – Location - Offshore.
  – Legal entity - Insourcing.
  – Geographic distance - Far.
  – Temporal distance - Large.

1.2 Project 4

Project 4 was performed by four different sites. The project used a semi-distributed effort estimation process architectural model. Both early and late effort estimation were performed. Site A was located in India and applied a plan-driven software process (server site). Site B was located in Germany and applied an agile-based software process. Site C was located in Spain and applied an agile-based software process. Site D was located in England and applied an agile-based software process. All the sites only interacted with Site A. The collaboration diagram of project 4 is displayed in Figure 16.

![Collaboration diagram of project 4.](image)

The result of the sites’ classification for project 4 is as follows:
• Site A
  – Software process - Plan-driven.
  – Power distance - Large.
  – Uncertainty avoidance - Weak.
  – Language distance - No distance.
  – Estimation stage - Early/Late.
  – Estimation process role - Estimator/Supplier.

• Sites B and C
  – Software process - Agile.
  – Power distance - Small.
  – Uncertainty avoidance - Strong.
  – Language distance - No distance.
  – Estimation stage - Early/Late.
  – Estimation process role - Supplier.

• Site D
  – Software process - Agile.
  – Power distance - Small.
  – Uncertainty avoidance - Weak.
  – Language distance - No distance.
  – Estimation stage - Early/Late.
  – Estimation process role - Supplier.

The result of the relationships’ classification for project 4 is as follows:

• Relationships AB, AC and AD
  – Location - Offshore.
  – Legal entity - Insourcing.
  – Geographic distance - Far.
  – Temporal distance - Large.

1.3 Project 5

Figure 17: Collaboration diagram of project 5.
Project 5 was performed by two different sites. The project used a distributed effort estimation process architectural model. Both early and late effort estimation were performed. Site A was located in India and applied a plan-driven software process (integrator site). Sites B was located in Japan and applied a plan-driven software process. The collaboration diagram of project 5 is displayed in Figure 17.

The result of the sites’ classification for project 5 is as follows:

- **Site A**
  - **Software process** - Plan-driven.
  - **Power distance** - Large.
  - **Uncertainty avoidance** - Weak.
  - **Language distance** - No distance.
  - **Estimation stage** - Early/Late.
  - **Estimation process role** - Estimator/Supplier.

- **Site B**
  - **Software process** - Plan-driven.
  - **Power distance** - Large.
  - **Uncertainty avoidance** - Strong.
  - **Language distance** - No distance.
  - **Estimation stage** - Early/Late.
  - **Estimation process role** - Estimator/Supplier.

The result of the relationship classification for project 5 is as follows:

- **Relationship AB**
  - **Location** - Offshore.
  - **Legal entity** - Insourcing.
  - **Geographic distance** - Far.
  - **Temporal distance** - Large.
  - **Estimation process architectural model** - Distributed.

### 1.4 Project 6

Project 6 was performed by two different sites. The project used a distributed effort estimation process architectural model. Only early effort estimation was performed. Site A was located in Germany and applied an agile-based software process (integrator site). Site B was located in India and applied an agile-based software process. The collaboration diagram of project 6 is displayed in Figure 18.

The result of the sites’ classification for project 6 is as follows:

- **Site A**
  - **Software process** - Agile.
– Power distance - Small.
– Uncertainty avoidance - Strong.
– Language distance - No distance.
– Estimation stage - Early.
– Estimation process role - Estimator/Supplier.

• Site B
– Software process - Agile.
– Power distance - Large.
– Uncertainty avoidance - Weak.
– Language distance - No distance.
– Estimation stage - Early.
– Estimation process role - Estimator/Supplier.

The result of the relationship classification for project 6 is as follows:

• Relationship AB
  – Location - Offshore.
  – Legal entity - Insourcing.
  – Geographic distance - Far.
  – Temporal distance - Large.
  – Estimation process architectural model - Distributed.

1.5 Project 7

All the four sites of Project 7 were placed in India. The project used a distributed effort estimation process architectural model. Both early and late effort estimation were performed. Site A was the integrator site. All sites related to each other in such project. All sites employed a plan-driven software process. The collaboration diagram of project 7 is displayed in Figure 19.

The result of the sites’ classification for project 7 is as follows:

• Sites A, B, C and D
  – Software process - Plan-driven.
The result of the relationship classification for project 7 is as follows:

- **Power distance** - Large.
- **Uncertainty avoidance** - Weak.
- **Language distance** - No distance.
- **Estimation stage** - Early/Late.
- **Estimation process role** - Estimator/Supplier.

The result of the relationship classification for project 7 is as follows:

- **Relationships AB, AC and AD**
  - **Location** - Onshore.
  - **Legal entity** - Insourcing.
  - **Geographic distance** - Distant.
  - **Temporal distance** - Similar.
  - **Estimation process architectural model** - Distributed.

- **Relationships BC, BD and CD**
  - **Location** - Onshore.
  - **Legal entity** - Insourcing.
  - **Geographic distance** - Distant.
  - **Temporal distance** - Similar.
  - **Estimation process architectural model** - No interaction.
1.6 Project 8

Project 8 was performed by four different sites. The project used a distributed effort estimation process architectural model. Both early and late effort estimation were performed only by Site A. The other sites performed only early effort estimation. Site A was located in the USA’s east coast and applied an agile-based software process (integrator site). Sites B, C and D were located in India and applied a plan-driven software process. All the sites only interacted with Site A. The collaboration diagram of project 8 is displayed in Figure .20.

![Collaboration diagram of project 8](image)

Figure .20: Collaboration diagram of project 8.

The result of the sites’ classification for project 8 is as follows:

- **Site A**
  - **Software process** - Agile.
  - **Power distance** - Small.
  - **Uncertainty avoidance** - Weak.
  - **Language distance** - No distance.
  - **Estimation stage** - Early/Late.
  - **Estimation process role** - Estimator/Supplier.

- **Sites B, C and D**
  - **Software process** - Plan-driven.
  - **Power distance** - Large.
  - **Uncertainty avoidance** - Weak.
  - **Language distance** - No distance.
  - **Estimation stage** - Early.
  - **Estimation process role** - Estimator/Supplier.

The result of the relationships’ classification for project 8 is as follows:
• Relationship AB, AC and AD
  – **Location** - Offshore.
  – **Legal entity** - Insourcing.
  – **Geographic distance** - Far.
  – **Temporal distance** - Large.
  – **Estimation process architectural model** - Distributed.
BIBLIOGRAPHY


[84] C. Linnaeus. System of nature through the three kingdoms of nature, according to classes, orders, genera and species, with characters, differences, synonyms, places (in Latin). Laurentius Salvius, 10th edition, 1758. (Cited on pages 59, 80, and 96.)


[90] E. A. Matusse, E. H. M. Huzita, and T. F. C. Tait. Metrics and indicators to assist in the distribution of process steps for distributed


ABSTRACT

Background: Global Software Engineering (GSE) has become a widely applied operational model for the development of software systems; it can increase profits and decrease time-to-market. However, there are many challenges associated with the development of software in a globally distributed fashion. There is evidence that these challenges affect many processes related to software development, such as effort estimation. To the best of our knowledge, there are no empirical studies to gather evidence on effort estimation in the GSE context. In addition, there is no common terminology for classifying GSE scenarios focusing on effort estimation.

Objective: The main objective of this thesis is to support effort estimation in the GSE context by classifying the existing knowledge in this field.

Method: In this thesis, we employed systematic literature review, survey, systematic mapping and literature survey as research methods.

Results: Our research on the state of the art and the state of the practice on effort estimation in the GSE context shows that the effort estimation techniques employed in GSE projects do not differ from the techniques employed in collocated projects. It was also identified that global aspects, e.g. time, geographical and socio-cultural distances, are accounted for as cost drivers, although it is not clear how they are measured. Our systematic mapping study on software engineering taxonomies showed that most taxonomies are designed in an ad-hoc way; hence, an existing method was revised to support the development of SE taxonomies in a more systematic way. The aforementioned results were combined to specialize an existing GSE taxonomy, with focus on effort estimation. The usage of the specialized taxonomy was illustrated by classifying 8 completed GSE projects. The results show that the specialized taxonomy is comprehensive enough to classify effort estimation aspects of GSE projects.

Conclusions: We believe that the taxonomy presented in this thesis can help researchers and practitioners to report new research on effort estimation in the GSE context; researchers and practitioners will be able to gather evidence, compare new studies and find new gaps in an easier way. Finally, further research should be conducted to complement the findings of this thesis, e.g. both the way cost drivers impact the effort of GSE projects and the way they are measured are not fully clear.