



<http://www.diva-portal.org>

## Postprint

This is the accepted version of a paper presented at *15th International Conference on Agile Software Development*.

Citation for the original published paper:

Britto, R., Usman, M., Mendes, E. (2014)

Effort Estimation in Agile Global Software Development Context.

In: *Agile Methods. Large-Scale Development, Refactoring, Testing, and Estimation: XP 2014 International Workshops, Rome, Italy, May 26-30, 2014, Revised Selected Papers*

[http://dx.doi.org/10.1007/978-3-319-14358-3\\_15](http://dx.doi.org/10.1007/978-3-319-14358-3_15)

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:bth-11146>

# Effort estimation in Agile Global Software Development Context

Ricardo Britto<sup>1</sup>, Muhammad Usman<sup>1</sup>, Emilia Mendes  
Department of Software Engineering, Faculty of Computing, Blekinge Institute of  
Technology, 371 79, Karlskrona, Sweden.  
{ricardo.britto, muhammad.usman, emilia.mendes}@bth.se

**Abstract.** Both Agile Software Development (ASD) and Global Software Development (GSD) are 21<sup>st</sup> century trends in the software industry. Many studies are reported in the literature wherein software companies have applied an agile method or practice GSD. Given that effort estimation plays a remarkable role in software project management, how do companies perform effort estimation when they use agile method in a GSD context? Based on two effort estimation Systematic Literature Reviews (SLR) - one in within the ASD context and the other in a GSD context, this paper reports a study in which we combined the results of these SLRs to report the state of the art of effort estimation in agile global software development (ASD) context.

**Keywords:** Agile Software Development, Global Software Development, Effort Estimation.

## 1 Introduction

The software industry is greatly impacted by globalization of world economies in 21<sup>st</sup> century. Software companies are increasingly engaging themselves in Global Software Development (GSD) in order to gain benefits such as cost savings, access to global resource pool, round the clock development [1]. Due to temporal, cultural and geographical boundaries, GSD also poses some challenges e.g. communication and coordination issues, project management, knowledge management [1]. In parallel with the GSD trend, software industry is also shifting to agile methods during last ten years or so. ASD [2] and GSD are 21<sup>st</sup> century trends in software industry. Studies have been conducted to investigate the adoption of agile methods and practices in GSD context, also called by Agile Global Software Development (AGSD) [3].

Jalali *et al.* [4] conducted a systematic literature review in order to find the state of the art of applying agile methods and practices in GSD context. In this SLR both inshore and offshore-distributed development settings were considered. The authors found that most of the existing literature consists of industrial experience reports. The authors also identified the most used agile practices in the context of GSD.

---

<sup>1</sup> First two authors have contributed equally in this work.

Hossain *et al.* [5] performed a systematic literature review that identified challenges and risk factors related to the use of Scrum practices in globally distributed projects. Strategies and practices to deal with the identified challenges and risk factors were also investigated. The authors found out that in order to be applied in a global context, Scrum practices must be adapted to deal with the additional difficulties regarding communication, coordination and collaboration processes in a globally distributed software project.

Project management is an important task in both agile and global software development contexts. Estimation is at the core of efficient project management as it guides the formulation, execution and adjustment of project plans. It is important to see what software estimation techniques or predictors or metrics have been used with agile methods when they are applied in GSD context i.e. Agile Global Software Development (AGSD). To date, no work has tried to aggregate the evidence regarding effort estimation in the context of AGSD. The aim of this paper is to report the state of the art on effort estimation in AGSD. Rest of the paper is organized as follows: Section 2 describes the research methodology; results are presented in Section 3.; Section 4 states the validity threats and conclusion is described in Section 5.

## 2 Methodology

As previously mentioned, to carry out this study we combined the outcomes of two systematic literature reviews performed by the authors of this paper ([6], [7]). So, in this section we explain the methodology used to conduct this study.

### 2.1 Research questions

The research questions of the two SLRs were combined in order to guide this work. They are as follows:

- **Question 1** - What methods/techniques have been used to estimate effort in AGSD?
  - **1a** - What metrics have been used to measure the accuracy of effort estimation methods/techniques in AGSD 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 AGSD?
- **Question 3** - What are the characteristics of the datasets used for effort estimation in AGSD?
  - **3a** - What are the domains represented in the dataset (academia/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** - Which software development phases were considered during effort estimation process?
- **Question 5** – What sourcing strategies (offshore outsourcing/offshore insourcing) are used?
  - **5a** - Which countries involved?
  - **5b** - How many sites are involved?
- **Question 6** – Which agile methods have been used?

## 2.2 Study Selection and Data Extraction

Both SLRs ([6], [7]) have used same databases/search engines for applying the search strings. These databases/search engines were:

1. Scopus.
2. IEEExplore.
3. ACM Digital Library.
4. ScienceDirect.
5. Compendex.
6. Inspec.
7. Web of Science.

Both SLRs have similar inclusion exclusion criteria with the only difference being that one was about agile and other was about global software development. The study selection process was applied in two phases. In the first phase, the inclusion and exclusion criteria were applied on titles and abstracts and in the second phase the criteria were applied on the papers' full text. The final lists of each SLR have respectively 5 papers [6] and 20 papers (25 studies) [7]. From these final lists we selected for this study those papers that:

1. Have investigated effort estimation methods or size metrics or accuracy metrics or cost drivers and
2. Have used an agile method as software development process and
3. Are carried out in a GSD context.

The application of above criteria resulted in the selection of four papers from Britto *et al.* [6], identified as G1 [8], G2 [9], G3 [10] and G4 [11] each reporting a single study; and one paper [12] from Usman *et al.* [7] reporting results from four projects identified as A1a to A1d. Therefore this study includes a total of 5 papers reporting 8 projects.

Most of the required data were available from the data extraction steps of two SLRs to carry out the study. However, since the questions 5, 5a, 5c and 5b were considered just for Britto *et al.* [6] and question 6 was considered only for Usman *et al.* [7] we had to extract the remaining data from the selected papers, in order to address all research questions.

### 3 Results and Discussion

In this section, first we provide a brief description, contexts and settings of the included primary studies. Later, results for each of the research questions are described and discussed.

#### 3.1 Study Summaries

Study G1 reports a survey that was conducted to understand the state of the practice of effort estimation in GSD projects. Survey was applied in a large multinational IT organization that has operations in countries like USA, Brazil, India etc. The software development in this organization is performed using both onshore and offshore insourcing strategies. Out of a total of 3595 employees, 551 answered the survey. Study describes that the organization uses both agile and plan-driven development processes but it does not describe the name of the agile method followed in the organization. It was concluded in this study that the teams do not have a clear criterion to select a suitable effort estimation technique in a given context.

Study G2 reports a case study that was conducted at ABB<sup>2</sup> group of companies to understand the factors that impact the management of GSD projects. Seven projects (six at ABB and one at another company) were included in this case study wherein all projects were carried out in a single company setting i.e. offshore insourcing. Study describes that three out of seven projects applied an agile method but it does not specify the name of the agile method used. It also does not describe the development method used in other four projects. Data collection was performed by means of interviews (31 participants) and an online questionnaire (40 participants) wherein participants were from different ABB sites across the globe. The study identified number of factors (cost drivers) for GSD projects and mechanisms to mitigate the risk related to each identified factor.

Study G3 also reports a case study that was conducted at three different Indian software companies, which work in development of financial service, retail, manufacturing and telecommunication software systems. These software companies apply the offshore insourcing strategy for distributed development. Participatory action research approach was applied to collect the data from these three companies. It involved 75 brainstorming sessions with study participants that lead to the identification of several cost drivers. It is interesting to note that this study considered process model (agile or otherwise) as a cost driver. Applying the identified cost drivers, case base reasoning approach was used to estimate the effort of 219 projects. The study analyzed the impact of “the knowledge about client”, “the work dispersion across sites” and “the understanding of technology” on the development effort. The authors compared their customized case based reasoning approach with standard regression based approach for estimation and found that case based reasoning

---

<sup>2</sup> ABB is a leading company in power and automation sector ([www.abb.com](http://www.abb.com))

approach performed better than regression for the studied projects. It is important to note that the study does not describe the exact development process applied.

Study G4 reports a qualitative study in which authors proposed a formal model for task allocation and effort estimation in GSD. The model includes the estimation technique and cost drivers. However, the authors only validated the cost drivers by means of semi-structured interviews with four project managers that lead to the better understanding of the identified cost drivers. Process model (agile or plan-driven) is one of cost drivers that impact the development effort. Since the proposed estimation technique was not validated, we did not include it in our analysis. It is important to note that the study does not clearly describe the sourcing strategy and exact development process applied.

Study A1 reports a case study consisting of four projects to investigate effort estimation for testing phase in an agile software development context. Study presented a customized version of use case points estimation method for estimating testing effort only. A1 was performed in an offshore outsourcing context wherein Scrum was applied as the development method. The authors found that the new method (modified use case points method) was more accurate than the expert judgment and the original use case points method.

The details from these summarized studies are described in the following subsections.

### 3.2 Estimation Methods

Table 1 lists the estimation methods used in an AGSD context, showing that the methods used the most were expert judgment, use case points (UCP), planning poker and Delphi. Note that the use case point method is used differently – in one paper it was used to size the application and in another to estimate the testing effort. We also note that some of these ‘effort estimation’ methods are in fact size metrics, which are used in combination with some sort of productivity metrics, or cost per hour measure, to obtain the effort/cost relating to an application. Traditional algorithmic models such as COCOMO were not identified in any of the AGSD studies.

**Table 1:** Identified effort estimation methods.

Estimation method	Study ID
Case-based reasoning	G3
Planning poker	G1
Function point count	G1
Use case point count	G1, A1a to A1d
Use case point test effort estimation model	A1a to A1d
Expert judgment	G1, A1a to A1d
Delphi	G1
No estimation approach	G2, G4

### 3.3 Accuracy Metrics and Levels

Table 2 lists the accuracy metrics used in the selected papers and projects therein. Three studies (G1, G2, G3) did not report usage of any accuracy metric. Two papers (A1, G3) have used the magnitude of relative error (MRE) or its variation to assess the estimation accuracy of their techniques. Only one study (G3) has used multiple metrics (MMRE, MdMRE and Pred(25)).

**Table 2.** Identified accuracy metrics.

Accuracy metric	Study ID
MMRE	G3
MdMRE	G3
Pred(25)	G3
MRE	A1a to A1d
No accuracy metrics	G1, G2, G4

Only two papers (G3, A1) reported accuracy levels related to the estimation techniques being investigated. These values are reported in Table 3, where we can also see that case base reasoning and UC point test effort estimation model present good accuracy values [13]. MRE values for UCP method, for all four projects in study A1, are also below 25%.

**Table 3.** Identified accuracy levels

Estimation method	Study ID	Accuracy (%)
Case-based reasoning	G3	MMRE: 15.99 MdMRE: 11.67 Pred(25): 84.12
Use case point test effort estimation model	A1	Project1 – MRE:11; Project2 – MRE:2; Project3 – MRE:3; Project4 – MRE:6.
Expert judgment	A1	Project1 – MRE:32; Project2 – MRE:30; Project3 – MRE:8; Project4 – MRE:21.
Use case point	A1	Project1 – MRE:21; Project2 – MRE:20; Project3 – MRE:21; Project4 – MRE:10.

### 3.4 Cost Drivers and Size Metrics

Table 4 lists the cost drivers that were identified from the primary studies. Time, language and cultural differences are the most frequently reported cost drivers in an AGSD context. When we move from collocated development to GSD, global barriers, e.g. temporal, geographical and cultural, arise as fundamental challenges. These global challenges make communication and coordination tasks more difficult which in turn impacts all development activities (e.g. RE, estimation etc.) [1].

In addition, the process model is also reported by two studies as a cost driver. This may be due to the fact that papers in this study are applying or investigating the applicability of a different process model, e.g. agile methods, in a GSD context.

**Table 4.** Identified cost drivers.

Cost driver	Study ID
Time zone	G2, G3, G4
Language and cultural differences	G2, G3, G4
Process model	G3, G4
Communication	G4
Competence level	G2
Requirements legibility	G2
Process compliance	G2
Communication infrastructure	G2
Communication process	G2
Work dispersion	G3
Range of parallel-sequential work handover	G3
Client-specific knowledge	G3
Client involvement	G3
Design and technology newness	G3
Team size	G3
Project effort	G3
Development productivity	G3
Defect density	G3
Rework	G3
Reuse	G3
Project management effort	G3
Travel	G4
Tester efficiency factor	A1
Tester risk factor	A1

The size metrics identified in the five primary studies are listed in Table 5. Function points, LOC and UC points are used in two papers. Overall, point-based size metrics (function or UC or story points) were used in three out of five studies.

**Table 5.** Identified size metrics.

Size metric	Study ID
Function points	G1, G4
Lines of code	G3, G4
Use case points	G1, A1
Story points	G1
No size metric used	G2

### 3.5 Dataset Domain and Type

All primary studies used industrial datasets to evaluate the estimation methods. This is viewed as a positive sign given that the use of industrial datasets may increase the external validity of the results.

Another issue attached with the use of a dataset is the dataset type, which could be single company or cross company dataset. Three papers (G2, G3 and A1) used single company datasets, while two did not state the type of their datasets.

**Table 6.** Identified dataset domains.

Domain	Study ID
Industry	G1, G2, G3, G4, A1
Academia	none

**Table 7.** Identified dataset types.

Type	Study ID
Single-company	G2, G3, A1
Not stated	G1, G4

### 3.6 Application Type

Application type is only documented in one primary study (A1). In A1, two projects were Web-based systems while the other two were mobile applications.

**Table 8.** Identified application types.

Type	Study ID
Web-based	A1a, A1c
Mobile	A1b, A1d
Not stated	G1, G2, G3, G4

### 3.7 Sourcing Strategies and Countries

Three primary studies reported studies (G1, G2, G3) that are conducted in offshore insourcing environments, i.e. same company had multiple development sites in different parts of the world. Only one paper reports projects that are conducted in offshore outsourcing arrangements, i.e. the multiple sites involved in GSD project belong to different companies. Table 9 lists the sourcing strategies identified in the five selected papers.

**Table 9.** Identified sourcing strategies.

Sourcing strategy	Study ID
Offshore insourcing	G1, G2, G3
Offshore outsourcing	A1
Not stated	G4

Three studies did not report the number of countries (or sites) involved in GSD projects. GSD projects in Study G1 and G2 were considerably complex as they included seven and ten countries respectively.

**Table 10.** Identified number of involved countries.

Number	Study ID
7	G2
10	G1
Not stated	G3, G4, A1

Three studies did not state the name of the countries where development sites were located, while USA, China and India were reported by two studies. Additionally, primary study G1 reported UK, Malaysia, Japan, Taiwan, Ireland, Brazil and Slovak Republic. Finally, primary study G2 reported Finland, Germany, Norway and Sweden.

**Table 11.** Identified countries.

Name	Study ID
USA	G1, G2
China	G1, G2
India	G1, G2
Not stated	G3, G4, A1

### 3.8 Development Phase

Which software development phase or activity is being estimated is also an important concern. Four out of five selected papers did not state the development phase or activity being estimated. One possible explanation for not stating the development phase could be that all development activities are being estimated. Only one paper (A1) clearly states that testing effort is being estimated. Table 12 provides the breakdown for this facet.

**Table 12.** Considered phases in the effort estimation process

Phase	Study ID
Requirements	none
Design	none
Coding	none
Testing	A1
Transition	none
Not stated	G1, G2, G3, G4

### 3.9 Agile Method

Another important facet is to see what agile methods are being investigated in estimation studies in AGSD context. Table 13 gives the breakup of studies with respect to the agile method used. Only one paper (A1) states the agile method used (Scrum in this case) in its projects. Other studies only mention that they are using agile software development but did not specify the exact method. We are not sure

why a study would only state that they are following an agile software development without disclosing the exact method used.

It is also interesting to note that two primary studies (G3, G4) considered the usage of agile method as a cost driver. However, those studies did not explain the impact on the effort estimation process of the usage of agile methods.

**Table 13.** Identified agile methodologies.

Agile methodology	Study ID
Scrum	A1
Not stated	G1, G2, G3, G4

## 4 Threats to validity

We believe that the main threat to the validity of this work is related to the coverage of the available literature on effort estimation in agile global software development. We applied very comprehensive search strategies in both SLRs, which were used as basis for this work. However, Britto *et al.* [6] just considered effort estimation in GSD context and Usman *et al.* [7] just looked at effort estimation in the ASD context.

Another possible threat relates to the external validity of our findings. As we only have five papers (8 projects) in this study, it is not reasonable to generalize our findings outside the context of the projects that were presented herein. Nevertheless, given the number of companies embarking on GSD and agile practices, we refrain from taking the stance that companies may not combine both approaches; rather, we believe that such results may suggest the need for researchers to amplify the number of studies within the context of AGSD, so we can understand much better not only effort estimation but also other aspects relating to software development and management under such context.

## 5 Conclusions

This paper presents a study on effort estimation in AGSD by combining the results from two SLRs respectively on effort estimation in agile contexts and effort estimation in global software development contexts. Five papers, from the list of primary studies of both SLRs, fulfilled AGSD criteria set up for this study.

We found that most of the studies did not document some aspects such as the agile method applied, the GSD strategy used, the number of development sites, the countries involved, and the development phase being estimated. Methods such as expert judgment are used in multiple studies. Global barriers of time and culture are the most frequently reported cost drivers in an AGSD context.

It is interesting to note a positive pattern, where all studies used industrial data sets to validate the estimation techniques. Offshore insourcing is the most frequently used GSD strategy in effort estimation studies in AGSD context.

**Acknowledgments.** We would like to thank CNPq and INES, for partially supporting this work.

## References

1. Herbsleb, J., Moitra, D.: Global software development. *IEEE Softw.* 18, 16–20 (2001).
2. Schwaber, K., Beedle, M.: *Agile Software Development with Scrum*. Prentice Hall (2001).
3. Kamaruddin, N. K., Arshad, N.H., Mohamed, A.: Chaos issues on communication in agile global software development. *Proceedings of IEEE Business, Engineering and Industrial Applications Colloquium - BEIAC'12*. pp. 394–398 (2012).
4. Jalali, S., Wohlin, C.: Global software engineering and agile practices: A systematic review. *J. Softw. Evol. Process.* 24, 643–659 (2012).
5. Hossain, E., Ali Babar, M., Paik, H.-Y.: Using scrum in global software development: A systematic literature review. *Proceedings of 4th IEEE International Conference on Global Software Engineering - ICGSE '09*. pp. 175–184. , Limerick, Ireland (2009).
6. Britto, R., Freitas, V., Mendes, E., Usman, M.: Effort Estimation in Global Software Development: A Systematic Literature Review. *Proceedings of 9th IEEE International Conference on Global Software Engineering - ICGSE'14*. , Shanghai, China (2014).
7. Usman, M., Mendes, E., Weidt, F., Britto, R.: Effort Estimation in Agile Software Development: A Systematic Literature Review. *Proceedings of 10th International Conference on Predictive Models in Software Engineering - PROMISE'14*. pp. 82–91. , Turin, Italy (2014).
8. Peixoto, C.E.L., Audy, J.L.N., Prikladnicki, R.: Effort Estimation in Global Software Development Projects: Preliminary Results from a Survey. *Proceedings of 5th IEEE International Conference on Global Software Engineering - ICGSE'10*. pp. 123–127. Ieee, Princeton, USA (2010).
9. Bjorndal, P., Smiley, K., Mohapatra, P.: Global Software Project Management: A Case Study. In: Nordio, M and Joseph, M and Meyer, B and Terekhov, A. (ed.) *Lecture Notes in Business Information Processing - International Conference on Software Engineering Approaches for Offshore and Outsourced Development (SEAFOOD)*. pp. 64–70. , Saint Petersburg, Russia (2010).
10. Ramasubbu, N., Balan, R.K.: Overcoming the challenges in cost estimation for distributed software projects. *Proceedings of 34th International Conference on Software Engineering - ICSE'12*. pp. 91–101. IEEE, Zurich, Switzerland (2012).
11. Narendra, N.C., Ponnalagu, K., Zhou, N., Gifford, W.M.: Towards a Formal Model for Optimal Task-Site Allocation and Effort Estimation in Global Software Development. *Proceedings of 2012 Service Research and Innovation Institute Global Conference*. pp. 470–477. IEEE, California, USA (2012).
12. Parvez, A.W.M.M.: Efficiency factor and risk factor based user case point test effort estimation model compatible with agile software development. *Proceedings of the International Conference on Information Technology and Electrical Engineering - ICITEE'13*. pp. 113–118. , Yogyakarta, Indonesia (2013).
13. Conte, S.D., Dunsmore, H.E., Shen, V.Y.: *Software Engineering Metrics and Models*. Benjamin-Cummings Publishing (1986).