Indonesian Journal of Electrical Engineering and Computer Science
Vol. 32, No. 3, December 2023, pp. 1485~1494
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v32.i3.pp1485-1494

Systematic literature review on global software development based software cost estimation models and cost drivers

Mehmood Ahmed[1,3], Noraini Ibrahim[1], Wasif Nisar[2], Adeel Ahmed[3]
[1]Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia Parit Raja, Batu Pahat, Malaysia
[2]Department of Computer Science, Wah Campus, COMSATS University Islamabad, Islamabad, Pakistan
[3]Department of Information Technology, The University of Haripur, Haripur, Pakistan

Article Info

Article history:

Received Nov 18, 2022
Revised Jul 17, 2023
Accepted Aug 14, 2023

Keywords:

Cost drivers
Cost estimation
Distributed software development
Global software development
Systematic literature review

ABSTRACT

Global software development (GSD) is a well-established discipline of software engineering that focuses on the advantage of a global environment. Effective cost estimation is critical for the success of GSD projects. Cost estimation in a GSD environment is a challenging task. As a result, GSD must emphasize cost estimation. Findings show that a number of researchers over the past few decades have emphasized GSD-based cost estimation in GSD; to the best of our knowledge, however, existing cost estimation have not taken into account many GSD-based cost drivers that must be considered when estimating costs. Motivated by all this, the purpose of this study is to review the existing GSD-based cost estimation models/techniques and cost drivers that influence the accuracy of cost estimation. To identify and compile relevant research papers, a systematic literature review was carried out. From twenty-seven selected studies, initially, 86 GSD-based cost drivers and 12 GSD-based cost estimation models/techniques were extracted. After filtration, 26 cost drivers were identified as significant and to be considered in GSD-based cost estimation. This study significantly identifies GSD-based cost drivers and existing cost estimation techniques.

Corresponding Author:

Mehmood Ahmed
Faculty of Computer Science and Information Technology
Universiti Tun Hussein Onn Malaysia Parit Raja
Batu Pahat, Malaysia
Email: gi180002@student.uthm.edu.my

1. INTRODUCTION

Global software development (GSD) describes an environment in which a software project is managed within a distributed setting [1], [2]. One of the main drivers behind the growth of GSD projects is cost savings. Several software companies have sought to gain market share by decreasing their product's time-to-market, cutting costs by employing people from low-wage countries, and defying duration by working on projects around the clock [1]. As an outcome, most software development projects are completed globally across various sites and countries. Companies frequently seek to outsource software development to different countries such as India and China, where labor rates are lower compared to Western countries. This allows them to reduce their development costs significantly [3]. GSD advantages can only be obtained through effective project management. The challenge of managing GSD projects is due to various issues, including time zone and cultural differences; if we do not consider these, it might take extra time and cost [3]. Avoiding budget overruns and late software delivery requires accurate cost and effort estimation [4]. Several tools and techniques for cost estimation were developed for the collocated context before the GSD concept, but these tools and techniques lack GSD cost drivers [5]. These cost estimation models are generally classified into

algorithmic, non-algorithmic, and hybrid models [6]. Estimating effort and cost for GSD projects can be difficult due to a number of issues [5]. Project managers' primary concern is cost estimation. But in GSD, due to its distributed nature, work allocation, resource allocation, and cost estimation are complex [7]. The current state of the art lacks cost drivers that can increase GSD cost estimation accuracy. Considering these cost drivers can enhance the accuracy of GSD-based cost estimation techniques and help practitioners implement them. Identification of GSD-based cost drivers is essential since the accuracy of GSD cost estimation is vital for the success of GSD-based projects. So, it is imperative to study further identification of GSD-based cost drivers and to improve cost estimation accuracy. In section 1 presents an overview of the research study; section 2 discusses the research method; section 3 discusses the findings and section 4 the conclusion.

2. RESEARCH METHOD

A systematic literature review (SLR) is a standard technique for synthesizing the most relevant research on the topic [8], [9]. SLR is a systematic, concise, and consistent process for identifying, reviewing, evaluating, and synthesizing the existing body of complete and documented research [10]–[12]. SLR was used to find and assess relevant scholarly research papers for this study [10]. SLR consists of 3 stages, i.e., review planning, review conducting, and review reporting [10].

2.1. Review planning

This phase comprises the most important steps required to carry out an SLR. To begin an SLR, some prerequisite steps must be performed. The steps are as: i) defining the research question; ii) specifying the research databases; iii) developing a search string generation for article extraction; iv) specifying inclusion and exclusion criteria; and v) specifying quality criteria. The following sections describe the above steps.

2.1.1. Defining the research question

Population, intervention, comparison, outcomes, and context (PICOC) approach was used to formulate research questions [13]. i) population (P): projects in global software development context, ii) intervention (I): cost drivers/estimation techniques/methods; iii) comparison (C): no comparison intervention; iv) outcomes (O): an accuracy of the tools or techniques used for cost estimation; and v) context (C): will examine GSD-based cost estimation. SLR attempts to find answers to the following questions:
Question 1. What cost drivers have been considered in GSD research?
Question 2. What techniques/methods have been applied to estimate the cost of GSD?

2.1.2. Specifying the research databases

The selection of research databases is decided as per recommendations [14]. Databases selected for this study include ACM Digital Library, Wiley Inter Science, ScienceDirect, IEEE Xplore Digital Library, Springer Link and Google Scholar. These databases have been chosen because they contain the most significant articles and conference proceedings.

2.1.3. Developing a search string to extract the articles

After developing research questions and selecting research databases, a search string, as mentioned in Table 1, is formulated to find primary studies. To avoid researcher bias, the search string is composed as: i) analyze the research questions and identify relevant keywords to the population, intervention, and outcomes; ii) check the keywords and evaluate the relevant papers; and iii) using the Boolean AND operator, connect the keywords from population, intervention, and outcome. These three steps were carried out to generate the following string:

Table 1. Search string

Search group	Search string
Software cost estimation	("cost estimation" OR "effort estimation" OR "effort prediction" OR "cost prediction" OR "cost assessment" OR " effort assessment" OR "calculating cost" OR "calculating effort" OR "measuring cost" OR "measuring effort."
	AND
Global software development	"globally software development" OR "global software engineering" OR "globally distributed development" OR "globally distributed work" OR "distributed software development" OR "distributed software engineering" 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."

2.1.4. Specifying inclusion and exclusion criteria

In order to ensure that only relevant literature is admitted to the SLR, inclusion and exclusion criteria is used as described in Table 2. Based on the recommendations of [10], inclusion and exclusion criteria for this study was determined. The following criteria is used to choose research studies:

Table 2. Inclusion and exclusion criteria

Inclusion criteria		Exclusion criteria	
IC 1	Research addressing models, techniques, and cost drivers for effort estimation for GSD AND;	EC 1	Research that doesn't address cost estimation models, techniques, cost drivers for GSD OR;
IC 2	Research studies must be written in English AND;	EC 2	Research without empirical support OR;
IC 3	Research should be published in a peer-reviewed journal, conference, technical report, or thesis.	EC 3	Research not published in English

2.1.5. Specifying quality criteria

This stage entails evaluating the quality of the selected research studies in order to enable the extraction of information for synthesis and analysis results. Quality research depends on evaluating the selected studies. Studies were assessed using a questionnaire with six questions, as mentioned in Table 3. The guidelines as per [10], [15] were followed in formulating the questions. A paper's quality rating could be 0 and 6. The above checklist questions are assessed as follows: i) addressing the research question correctly gets 1 point; ii) addressing the research question partially gets 0.5 points; and iii) not addressing the research questions gets 0 points.

Table 3. Quality assessment questions

S.#	Quality assessment questions	S.#	Quality assessment questions
Q1	Do the research objectives carefully stated?	Q4	Is there sufficient information about the data collection?
Q2	Is the research strategy well-defined and well-documented?	Q5	Does the study take into account the cost factors for GSD?
Q3	Is the research considering GSD context in cost estimation?	Q6	Do the research results answer the research objectives?

2.2. Phase 2: review conduct

Article extraction followed a standard approach. This SLR step applied a search query to select primary studies. Once retrieved, documents were compiled based on criteria, the following sections outline sub-activities.

2.2.1. Secondary and primary searches

Search strings were used on several databases as part of the main search strategy to get the primary studies published, as mentioned in Figure 1. Selected databases contain all major software engineering conferences and publications; therefore, they cover SLRs extensively.

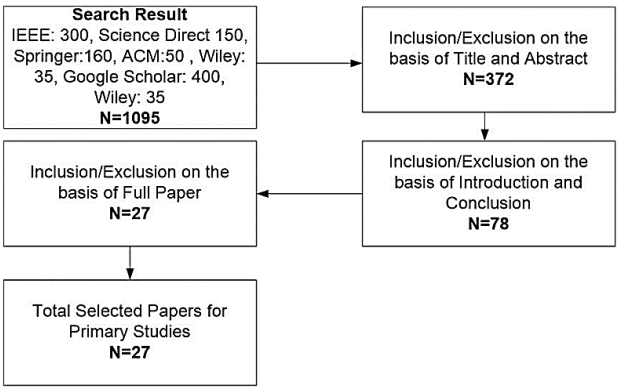


Figure 1. Study selection flowchart based on inclusion and exclusion

Furthermore, it was ensured that these databases included all primary research articles on the SLR topic. Wickramaarachchi and Lai [5], Khan *et al.* [16] have used these databases to find primary research.

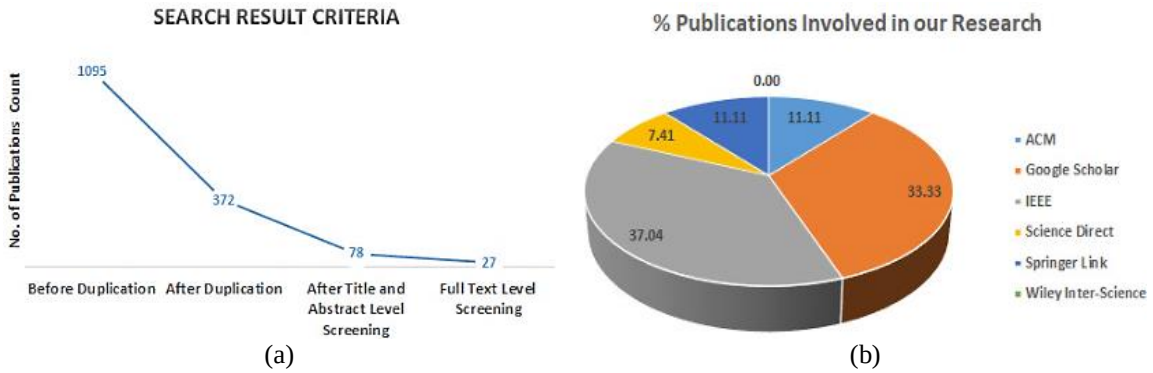
The next step was searching by consulting references. In the second step, we reviewed all the main references. In step 1, specified criteria for inclusion/exclusion were applied to the titles and abstracts of selected primary studies. This step was performed on all 372 search results, with the result that 78 studies passed the criteria. In step 2, inclusion/exclusion criteria were applied to 78 studies that passed the title and abstract level test. This step was performed independently on all 78 studies; as a result, 27 studies passed the criteria. Figure 1 illustrates the distribution of search results.

2.2.2. Data extraction and synthesis

The studies were selected using study id, publication year, author name, research technique, and associated constraints. Table 4 contains a list of the papers that were considered for inclusion. The relevant research topics were matched with the selected documents to determine relevancy. Selected studies were evaluated against research topics during synthesis and discovered 86 cost drivers and twelve GSD-specific cost estimate models from twenty seven publications. Figure 2 presents the search result. Figure 2(a) illustrate the search result based on the criteria of screeing research papers whereas Figure 2(b) present the percentage of research papers from each reseach database.

Table 4. List of included articles

Study ID	Citation	Database	Conference/Journal/Thesis
S1	[17]	IEEE	Conference
S2	[18]	Google Scholar	Journal
S3	[19]	IEEE	Conference
S4	[20]	IEEE	Conference
S5	[21]	Google Scholar	Journal
S6	[7]	Google Scholar	Thesis
S7	[22]	Google Scholar	Journal
S8	[16]	IEEE	Journal
S9	[5]	Springer	Journal
S10	[2]	Google Scholar	Journal
S11	[23]	IEEE	Journal
S12	[24]	Science Direct	Journal
S13	[25]	IEEE	Conference
S14	[26]	ACM	Journal
S15	[27]	IEEE	Conference
S16	[28]	Springer Link	Conference
S17	[29]	ACM	Conference
S18	[30]	Google Scholar	Journal
S19	[31]	IEEE	Conference
S20	[32]	Google Scholar	Journal
S21	[33]	ACM	Workshop
S22	[34]	Springer Link	Conference
S23	[35]	Science Direct	Conference
S24	[36]	Google Scholar	Journal
S25	[37]	IEEE	Journal
S26	[38]	IEEE	Conference
S27	[39]	Google Scholar	Conference



(a) (b)

Figure 2. Distribution of study results (a) search results based on criteria and (b) percentage publication from research databases

2.3. Phase 3: review reporting

In reporting phase, the quality assessment questions are used to evaluate the selected primary studies. The quality assessment criteria described in phase 1 were used to assess the quality of the selected studies. The study quality rating might be between 0 to 6. Quality results are cited in Table 5. Table 6 demonstrates that more articles have been published on cost estimation in the GSD context over recent years. One could say that the increase in publications shows that practitioners and researchers are becoming more interested in software cost estimation for GSD projects.

Table 5. Quality assessment score for selected studies

Study ID	Q1	Q2	Q3	Q4	Q5	Q6	Total score
S1	1	1	0.5	0.5	0.5	0.5	4
S2	1	1	0.5	0.5	0.5	0.5	4
S3	1	1	1	0.5	1	0.5	5
S4	1	0.5	1	0	1	0.5	4
S5	1	1	0.5	1	0.5	1	5
S6	1	1	0.5	0.5	0.5	0.5	4
S7	1	1	0.5	1	0.5	1	5
S8	1	1	1	1	1	0.5	5.5
S9	1	1	1	1	1	0.5	5.5
S10	0.5	0.5	1	0.5	0	0.5	3
S11	1	1	0.5	0.5	0	1	4
S12	1	0.5	0.5	0.5	0.5	0.5	3.5
S13	1	0.5	1	1	0.5	1	5
S14	1	1	0.5	0.5	0.5	0.5	4
S15	1	0.5	0.5	0.5	0.5	0.5	3.5
S16	1	1	1	0.5	0.5	0.5	4.5
S17	0.5	1	1	0.5	0	0.5	3.5
S18	1	1	1	1	0	0.5	4.5
S19	1	1	0.5	1	0	0.5	4
S20	1	0.5	0.5	0.5	0	0.5	3
S21	1	1	1	0	0	1	4
S22	1	1	1	0.	0	1	4
S23	1	1	1	0	0	0.5	3.5
S24	1	1	0.5	0.5	0.5	0.5	4
S25	1	1	1	1	1	0.5	5.5
S26	1	1	1	0.5	0.5	1	5
S27	1	1	1	0.	0	1	4

Table 6. Number of published articles by year interval

Year	2001-2005	2006-2010	2011-2015	2016-2022
No. of articles	0	6	10	11
Percentage	0	22.22	37.03	40.74

3. RESULTS AND DISCUSSION

This section discusses how this study can be used to support future research. This study also highlights those aspects where a researcher can further work. This study was done to find answers to two main questions:

3.1. Question 1. What cost drivers have been considered in GSD research?

Table 7 in Appendix lists the cost drivers collected from the literature. The lables include cost driver's name, frequency, percentage, and references. Cost drivers are those factors that have an intricate effect on global software development; the results could be either positive or negative. Due to the distributed nature of this type of development, these cost drivers are hidden in GSD. Due to the distributed nature of GSD, these cost drivers are often overlooked, which causes cost overruns later in the project. These cost drivers must be considered throughout the estimating process for a reasonable estimate of the effort and resources. Out of 27, 24 publications investigated the cost drivers [17]-[38], and some focused particularly on cost-estimating techniques for GSD. Although cost drivers were considered in previous research but did not provide data to support their validity. We first compiled a list of cost drivers from the literature and then conducted the analysis.

The study adopted criteria to rank cost drivers based on their frequency. Similar criteria are adopted in studies [16], [32]. Cost drivers with a frequency ≥25 are considered significant cost drivers (SFCD), and cost drivers with <25 frequency are designated as low significant cost drivers (LSFCD) and will not be considered in further investigation. The cost drivers mentioned in Table 8 are significant cost drivers.

Table 8. List of significant cost drivers

S.No	Cost driver
1	Language and cultural differences
2	Time zone difference
3	Communication
4	Geographic distance
5	Project management effort
6	Team trust
7	Process model
8	Knowledge management
9	Competence level
10	Team size
11	Client involvement
12	Requirement legibility
13	Development productivity
14	Design and technology newness
15	Reuse
16	Process maturity
17	Process compliance
18	Project management effort
19	Travel cost
20	Rework
21	Shared resources
22	Work pressure
23	Delay response
24	Task allocation
25	Project effort
26	Contract design

3.2. Question 2. What techniques have been applied to estimate the cost of GSD?

Due to the additional challenges associated with global software development [16], neither algorithmic nor non-algorithmic cost estimation approaches are suitable for GSD. Different studies mentioned in Table 9 are trying to overcome these challenges. Lamersdorf *et al.* [25] introduced a hybrid model, i.e., the COBRA-based model, which amplified with cost overhead. The researches [27], [30], [33], [34], [36], [37], [39] also proposed algorithmic approaches, and most of the techniques are COCOMO II-based amplification for GSD. These methods are currently in the experimental stage and are not yet validated. There are not sufficient cost drivers in these approaches.

Table 9. Identified GSD-based cost estimation models

Study ID	Citation	Category of model	Technique/approach	Number of GSD cost drivers used
S7	[7]	Non-algorithmic models	Casual model	11
S13	[25]	Hybrid models	COBRA-based cost estimation	13
S15	[27]	Algorithmic models	COCOMO II and SLIM	11
S17	[29]	Algorithmic models	Scheduling based model	03
S18	[30]	Algorithmic models	COCOMO II-based model	11
S19	[31]	Non-algorithmic models	Analogy-based cost estimation	12
S21	[33]	Algorithmic models	COCOMO II-based model	9
S22	[34]	Algorithmic models	Cost xpert model	8
S23	[35]	Machine learning based	Machine learning based	7
S24	[36]	Algorithmic models	COCOMO II and SLIM based	11
S25	[37]	Algorithmic models	COCOMO II	8
S27	[39]	Algorithmic models	Use case point (UCP) model	3

4. CONCLUSION

This study shows the results of a systematic literature review of software cost estimation in the GSD context. This result highlights that the software industry is rapidly adopting a GSD culture. This pace of globalization encouraged us to identify the cost factors that can affect software project management in general and software cost estimation specifically for GSD projects. The objective of this paper was to present the state-of-the-art software cost estimation model and techniques available with GSD context, along with cost drivers affecting the accuracy of cost estimation. We used SLR to accomplish the stated research goal to achieve our desired goal. In the primary search after title scanning, 372 articles were selected for introduction and conclusion, whereas only 78 articles were selected for full-text scanning. Finally, 27 research articles were selected for the primary study. It is important to highlight that out of 27 selected articles, only 24 discussed cost drivers, and 12 focused on cost estimation techniques. SLR identified 86 cost drivers in the GSD context;

however, based on frequency, 26 cost drivers are considered significant cost drivers and recommended to be considered in GSD-based cost estimation. Studies highlight that researchers are still working on GSD-based cost estimation model/techniques. Current reported techniques are still in experimental stages and most models are yet to validate.

APPENDIX

Table 7. Identified GSD-based cost drivers

S. No	Cost driver	Studies	Frequency	Percentage
1	Language and cultural differences	S1, S2, S3, S4, S6, S7, S8, S9, S10, S11, S12, S14, S15, S16, S17, S18, S21, S24, S25, S26	20	83.33
2	Time zone difference	S1, S3, S4, S5, S6, S7, S8, S9, S10, S11, S12, S15, S16, S17, S18, S24, S25, S26	18	75.00
3	Communication	S1, S2, S3, S5, S6, S7, S8, S9, S10, S11,S12, S13, S14, S16, S25, S26	16	66.67
4	Geographic distance	S1, S2, S4, S6, S7, S8, S9, S10, S12, S14, S15, S16, S21, S25, S26	15	62.50
5	Project management effort	S3, S4, S6, S8, S9, S10, S12, S13, S15, S16, S17, S18, S24, S25, S26	15	62.50
6	Team trust	S3, S5, S6, S7, S8, S9, S10, S12, S14, S15, S18, S24, S25, S26	14	58.33
7	Process model	S1, S3, S4, S6, S8, S9, S10, S15, S17, S18, S24, S25, S26	13	54.17
8	Knowledge management	S3, S4, S6, S7, S8, S10, S12, S13, S14,S16	10	41.67
9	Competence level	S3, S6, S8, S9, S10, S11, S13, S16, S21, S25	10	41.67
10	Team size	S3, S4, S5, S8, S10, S14, S16, S23, S26	9	37.50
11	Client involvement	S3, S4, S5, S8, S10, S14, S25, S26	8	33.33
12	Requirement legibility	S3, S6, S8, S10, S12, S13, S16, S25	8	33.33
13	Development productivity	S3, S4, S7,S8, S10, S11, S16, S26	8	33.33
14	Design and technology newness	S3, S4, S8, S10, S14, S16, S26	7	29.17
15	Reuse	S3, S4, S6, S8, S10, S12, S14, S16	7	29.17
16	Process maturity	S6, S7, S8, S9, S12, S13, S25	7	29.17
17	Process compliance	S3, S7, S8, S10, S12, S16, S25	7	29.17
18	Project management effort	S3, S4, S8, S10, S16, S25, S26	7	29.17
19	Travel cost	S3, S7, S8, S10, S14, S16	6	25.00
20	Rework	S3, S4, S8, S10, S12, S16	6	25.00
21	Shared resources	S3, S5, S6, S7, S8, S10, S12	6	25.00
22	Work pressure	S3, S5, S8, S10, S12, S26	6	25.00
23	Delay response	S3, S5, S8, S10, S14, S18	6	25.00
24	Task allocation	S3, S8, S10, S12, S13, S26	6	25.00
25	Project effort	S3, S4, S8, S10, S16, S26	6	25.00
26	Contract design	S6, S9, S16, S17, S24, S25	6	25.00
27	Organizational difference	S2, S6, S7, S8, S10	5	20.83
28	range of parallel sequential work handover	S3, S4, S8, S10, S16	5	20.83
29	Defect density	S3, S4, S8, S10, S16	5	20.83
30	Product architecture	S6, S8, S9, S10,S21	5	20.83
31	Product complexity	S8, S13, S23, S26	4	16.67
32	Buyer project manager	S6, S9, S18, S24	4	16.67
33	Outsourcer fit	S6, S9, S18, S24	4	16.67
34	Provider outsourcing Experience	S6, S9, S18, S24	4	16.67
35	Provider project managers	S6, S9, S18, S24	4	16.67
36	Change management activities	S6, S12, S13, S14	4	16.67
37	Code size	S4, S8, S10, S26	4	16.67
38	Intellectual property	S2, S7, S12, S14	4	16.67
39	Tool and infrastructure	S6, S9, S14, S21	4	16.67
40	Unrealistic milestones	S3, S5, S8, S10	4	16.67
41	Exchange rate fluctuation	S8, S9, S10	3	12.50

Table 7. Identified GSD-based cost drivers *(Continue)*

S. No	Cost driver	Studies	Frequency	Percentage
42	Information sharing constraints	S6, S9, S21	3	12.50
43	Novelty of collaboration model	S6, S9, S21	3	12.50
44	Platform volatility	S8, S10, S26	3	12.50
45	Unavailability of concerned person	S5, S8, S10	3	12.50
46	Common experience	S4,S6, S13	3	12.50
47	Social factors	S2, S8, S10	3	12.50
48	Coupling between tasks	S6, S7	2	8.33
49	Criticality	S6, S13	2	8.33
50	Formality of task description	S6, S13	2	8.33
51	Infrastructure	S9, S12	2	8.33
52	Local government	S7,S14	2	8.33
53	Number of sites	S6, S13	2	8.33
54	Overoptimism	S8,S11	2	8.33
55	Precedentedness	S6, S21	2	8.33
56	Resource cost	S7, S14	2	8.33
57	Rules/laws	S8, S10	2	8.33
58	Shared understanding	S9, S21	2	8.33
59	Training meeting/session	S8, S10	2	8.33
60	Vendor selection	S9, S25	2	8.33
61	Computer/application platform	S6, S23	2	8.33
62	Analyst ccapability	S6	1	4.17
63	Changing staff level	S11	1	4.17
64	Collaboration	S9	1	4.17
65	Concurrency	S23	1	4.17
66	Customer participation	S13	1	4.17
67	History of working together	S7	1	4.17
68	Intensity	S23	1	4.17
69	Lack of control	S2	1	4.17
70	Lack of transparency	S14	1	4.17
71	Lack of uniform processes among different team	S12	1	4.17
72	Learning curve	S11	1	4.17
73	Maturity level of vendor	S14	1	4.17
74	Mishandled dependencies	S14	1	4.17
75	Novelty of software to be developed	S9	1	4.17
76	People interest	S3	1	4.17
77	Personnel continuity	S6	1	4.17
78	Political risk related to international relations	S14	1	4.17
79	Poor visibility	S14	1	4.17
80	Process issues	S6	1	4.17
81	Reliability	S14	1	4.17
82	Slow decision making	S14	1	4.17
83	Tester risk factor	S16	1	4.17
84	Unsystematic handover	S14	1	4.17
85	Global team trust	S27	1	4.17
86	Global team composition	S27	1	4.17

ACKNOLWLEDGEMENT

The communication of this research is made possible through monetary assistance by Universiti Tun Hussein Onn Malaysia and the UTHM Publisher's Office via Publication Fund E15216. We are grateful for their financial assistance in this endeavor, which allowed us to pursue this important work.

REFERENCES

[1] A. A. Khan, S. Basri, and P. D. D. Dominc, "A proposed framework for communication risks during RCM in GSD," *Procedia - Social and Behavioral Sciences*, vol. 129, pp. 496–503, May 2014, doi: 10.1016/j.sbspro.2014.03.706.

[2] M. Ahmed, A. Ateeq, S. Khan, and M. Junaid, "Software cost estimation in global software development business: a review of models and cost drivers for economical business," *International Journal of Business and Economic Affairs*, vol. 6, no. 3, 2021, doi: 10.24088/ijbea-2021-63001.

[3] M. Ilyas and S. U. Khan, "Software integration in global software development: challenges for GSD vendors," *Journal of Software: Evolution and Process*, vol. 29, no. 8, p. e1875, Aug. 2017, doi: 10.1002/smr.1875.

[4] P. Sahoo and J. R. Mohanty, "Early test effort prediction using UML diagrams," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 5, no. 1, pp. 220–228, Jan. 2017, doi: 10.11591/ijeecs.v5.i1.pp220-228.

[5] D. Wickramaarachchi and R. Lai, "Effort estimation in global software development - a systematic review," *Computer Science and Information Systems*, vol. 14, no. 2, pp. 393–421, 2017, doi: 10.2298/CSIS160229007W.

[6] K. Usharani, V. V. Ananth, and D. Velmurugan, "A survey on software effort estimation," in *International Conference on Electrical, Electronics, and Optimization Techniques, ICEEOT 2016*, Mar. 2016, pp. 505–509, doi: 10.1109/ICEEOT.2016.7755665.

[7] J. Koskenkyla, "Cost estimation in global software development-review of estimation techniques," Aalto University School of Economics, 2012.

[8] S. Kraus, M. Breier, and S. Dasí-Rodríguez, "The art of crafting a systematic literature review in entrepreneurship research," *International Entrepreneurship and Management Journal*, vol. 16, no. 3, pp. 1023–1042, Sep. 2020, doi: 10.1007/s11365-020-00635-4.

[9] M. S. Farooq, S. Riaz, A. Abid, T. Umer, and Y. Zikria, "Role of IoT technology in agriculture: a systematic literature review," *Electronics (Switzerland)*, vol. 9, no. 2, p. 319, Feb. 2020, doi: 10.3390/electronics9020319.

[10] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering-a systematic literature review," *Information and Software Technology*, vol. 51, no. 1, pp. 7–15, Jan. 2009, doi: 10.1016/j.infsof.2008.09.009.

[11] Z. Stapić, G. Antonio, E. G. López Cabot, L. M. Ortega, and V. Strahonja, "Performing systematic literature review in software engineering," *In Central European Conference on Information and Intelligent Systems*, pp. 441–493, 2012 doi: 10.1145/1134285.1134500.

[12] B. L. Putro, Y. Rosmansyah, and Suhardi, "An intelligent agent model for learning group development in the digital learning environment: A systematic literature review," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 3, pp. 1159–1166, Jun. 2020, doi: 10.11591/eei.v9i3.2009.

[13] M. Petticrew and H. Roberts, *Systematic reviews in the social sciences: a practical guide*. Wiley, 2008.

[14] L. Chen, M. A. Babar, and H. Zhang, "Towards an evidence-based understanding of electronic data sources," *In 14th International conference on evaluation and assessment in software engineering (EASE)*, 2010, pp. 1-4, doi: 10.14236/ewic/ease2010.17.

[15] D. Azhar, E. Mendes, and P. Riddle, "A systematic review of web resource estimation," in *ACM International Conference Proceeding Series*, Sep. 2012, pp. 49–58, doi: 10.1145/2365324.2365332.

[16] J. A. Khan, S. U. R. Khan, J. Iqbal, and I. U. Rehman, "Empirical investigation about the factors affecting the cost estimation in global software development context," *IEEE Access*, vol. 9, pp. 22274–22294, 2021, doi: 10.1109/ACCESS.2021.3055858.

[17] M. Usman and R. Britto, "Effort estimation in co-located and globally distributed agile software development: a comparative study," in *Proceedings - 26th International Workshop on Software Measurement, IWSM 2016 and the 11th International Conference on Software Process and Product Measurement, Mensura 2016*, Oct. 2017, pp. 219–224, doi: 10.1109/IWSM-Mensura.2016.042.

[18] T. F. C. Tait and E. H. M. Huzita, "Software project management in distributed software development context," in *ICEIS 2013 - Proceedings of the 15th International Conference on Enterprise Information Systems*, 2013, vol. 2, pp. 216–222, doi: 10.5220/0004442402160222.

[19] R. Britto, V. Freitas, E. Mendes, and M. Usman, "Effort estimation in global software development: a systematic literature review," in *2014 IEEE 9th International Conference on Global Software Engineering*, Aug. 2014, pp. 135–144, doi: 10.1109/ICGSE.2014.11.

[20] N. Ramasubbu and R. K. Balan, "Overcoming the challenges in cost estimation for distributed software projects," in *Proceedings - International Conference on Software Engineering*, Jun. 2012, pp. 91–101, doi: 10.1109/ICSE.2012.6227203.

[21] M. Muhairat, S. Aldaajeh, and R. E. Al-Qutaish, "The impact of global software development factors on effort estimation methods," *European Journal of Scientific Research*, vol. 46, no. 2, pp. 221–232, 2010.

[22] A. Lamersdorf and J. Münch, "Studying the impact of global software development characteristics on project goals: a causal model," *The Open Software Engineering Journal*, vol. 4, no. 1, pp. 2–13, 2010, doi: 10.2174/1874107X0100401000.

[23] M. El Bajta *et al.*, "Software project management approaches for global software development: a systematic mapping study," *Tsinghua Science and Technology*, vol. 23, no. 6, pp. 690–714, Dec. 2018, doi: 10.26599/TST.2018.9010029.

[24] M. Niazi *et al.*, "Challenges of project management in global software development: a client-vendor analysis," *Information and Software Technology*, vol. 80, pp. 1–19, Dec. 2016, doi: 10.1016/j.infsof.2016.08.002.

[25] A. Lamersdorf, J. Münch, A. F.-D. V. Torre, C. R. Sánchez, and D. Rombach, "Estimating the effort overhead in global software development," in *Proceedings - 5th International Conference on Global Software Engineering, ICGSE 2010*, Aug. 2010, pp. 267–276, doi: 10.1109/ICGSE.2010.38.

[26] R. Jain and U. Suman, "A project management framework for global software development," *ACM SIGSOFT Software Engineering Notes*, vol. 43, no. 1, pp. 1–10, Mar. 2018, doi: 10.1145/3178315.3178329.

[27] S. Ramacharan and K. V. G. Rao, "Scheduling based cost estimation model: an effective empirical approach for GSD project," in *IFIP International Conference on Wireless and Optical Communications Networks, WOCN*, Jul. 2016, vol. 2016-November, pp. 1–5, doi: 10.1109/WOCN.2016.7759881.

[28] R. Britto, M. Usman, and E. Mendes, "Effort estimation in agile global software development context," in *Lecture Notes in Business Information Processing*, vol. 199, 2014, pp. 182–192.

[29] S. Ramacharan and K. V. G. Rao, "Software effort estimation of GSD projects using calibrated parametric estimation models," in *ACM International Conference Proceeding Series*, Mar. 2016, vol. 04-05-March-2016, pp. 1–8, doi: 10.1145/2905055.2905177.

[30] S. Betz and J. Mäkiö, "Amplification of the COCOMO II regarding offshore software projects," in *Offshoring of Software Development: Methods and Tools for Risk Management; 27th-30th August 2007, Munich, Germany; OUTSHORE; Proceedings*, 2008, p. 33, doi: 10.5445/KSP/1000007358.

[31] M. El-Bajta, "Analogy-based software development effort estimation in global software development," in *Proceedings - 2015 IEEE 10th International Conference on Global Software Engineering Workshops, ICGSEW 2015*, Jul. 2015, pp. 51–54, doi: 10.1109/ICGSEW.2015.19.

[32] J. Ahmad, A. W. Khan, and I. Qasim, "Software outsourcing cost estimation model (SOCEM). a systematic literature review protocol," *Journal of Information and Communication Technology*, vol. 2, no. 1, pp. 25–30, 2018.

[33] P. Keil, D. J. Paulish, and R. S. Sangwan, "Cost estimation for global software development," in *Proceedings of the 2006 international workshop on Economics driven software engineering research*, May 2006, pp. 7–10, doi: 10.1145/1139113.1139117.

[34] R. Madachy, "Distributed global development parametric cost modeling," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 4470 LNCS, pp. 159–168, 2007, doi: 10.1007/978-3-540-72426-1_14.

[35] M. Humayun and C. Gang, "Estimating effort in global software development projects using machine learning techniques," *International Journal of Information and Education Technology*, pp. 208–211, 2012, doi: 10.7763/ijiet.2012.v2.111.

[36] S. Ramacharan and K. V. G. Rao, "Parametric models for effort estimation for global software development," *Lecture Notes on Software Engineering*, pp. 178–182, 2013, doi: 10.7763/lnse.2013.v1.40.
[37] J. A. Khan, S. U. R. Khan, T. A. Khan, and I. U. R. Khan, "An amplified COCOMO-II based cost estimation model in global software development context," *IEEE Access*, vol. 9, pp. 88602–88620, 2021, doi: 10.1109/ACCESS.2021.3089870.
[38] M. El Bajta, A. Idri, J. L. Fernández-Alemán, J. N. Ros, and A. Toval, "Software cost estimation for global software development: a systematic map and review study," in *ENASE 2015 - Proceedings of the 10th International Conference on Evaluation of Novel Approaches to Software Engineering*, 2015, pp. 197–206, doi: 10.5220/0005371501970206.
[39] M. Azzeh, "Software cost estimation based on use case points for global software development," in *2013 5th International Conference on Computer Science and Information Technology, CSIT 2013 - Proceedings*, Mar. 2013, pp. 214–218, doi: 10.1109/CSIT.2013.6588782.

BIOGRAPHIES OF AUTHORS

Mehmood Ahmed received the MS (CS) degree from Iqra University, Islamabad Campus, Pakistan, in 2008, whereas he is currently pursuing the Ph.D. degree in Information Technology from Universiti Tun Hussein Onn Malaysia (UTHM) Malaysia. Mehmood Ahmed has published several papers in various journals and conference proceedings. His research interests include software requirement engineering, software project management, software process improvement, and their application in GSD context. He can be contacted at email: gi180002@student.uthm.edu.my.

Dr. Noraini Ibrahim received Doctor of Philosophy in Information Technology, University Tun Hussein Onn Malaysia (2013). Bachelor of Computer Science-Real Time Software Engineering, Technological University of Malaysia (2003) Bachelor of Computer Science, National University of Malaysia (2000). She can be contacted at email: noraini@uthm.edu.my.

Prof. Dr. Wasif Nisar received his Doctor of Philosophy in computer science in 2008 from Institute of Software, GUCAS China. He received his M. Sc. degree in 2000 and B.Sc. from The University Peshawtar. His research interest includes software effort estimation, software process improvement, software project management, software distributed systems, databases e management system. He can be contacted at email: wasif@ciitwah.edu.pk.

Adeel Ahmed (Graduate Student Member, IEEE) received the Ph.D. degree in computer science from Quaid-i-Azam University, Islamabad, Pakistan, in 2022. He has software industry experience for few years and his research interests include social network analysis, recommender systems, machine learning, and swarm intelligence. Currently he is serving as a Faculty member at Department of IT, University of Haripur, KPK, Pakistan. He can be contacted at email: adeel@uoh.edu.pk.