A Quantitative Approach for Selecting a Software Cost Estimation Method

Nurol Husna Che Rose, Haslina Md Sarkan [*], Nurulhuda Firdaus Azmi, Siti Sophiayati Yuhaniz, Wan Adibah Hanis Wan Aziz

Advanced Informatics School, Universiti Teknologi Malaysia, Jalan Sultan Yahya Petra,, 54100 Kuala Lumpur, Malaysia.

Abstract

Cost estimation is crucial to a project's success. Realizing its importance, many companies are aware that, selecting the appropriate cost estimation method is vital to guarantee an accurate estimation. This paper was intended to investigate the existing approach used in cost estimation in a communication company and to select the most suitable cost estimation methods to improve their current practices. The cost estimation methods practiced were limited to analogy based and expert judgment only. On top of that, no guideline for performing cost estimation is available. A quantitative approach was conducted using questionnaire to gather information on the current cost estimation practices. The approach also investigated the respondents' knowledge and awareness about existing cost estimation methods. Analysis from this study will be used to determine the nature of the work and how the projects are managed in the company. Results obtained is hoped to be able to help to improve the current cost estimation practices.

Keywords: Software Cost Estimation, Cost Management Process, Non-Algorithmic

1. Introduction

Many organizations and businesses today rely heavily on software to showcase their organizations in this competitive world. The cost of developing software has increased significantly the last few decades. In order for software development project to run efficiently, it is very important to estimate costs accurately. The more complex the software, the more difficult it is for the project manager and the development team to have accurate estimation of software cost and effort.

This paper tries to investigate the existing approach used in cost estimation in a communication company, to study its effectiveness or its flaws, and to propose a more suitable cost estimation method to improve their current practices. From the pre investigation done in the company, it was found that the most common software cost estimation methods used were limited to only two methods which were expert judgment and estimation by analog. On top of that, there was no guideline to steer them towards practicing a proper software cost estimation process.

A quantitative approach was conducted using questionnaire to gather information on the current cost estimation practices. The approach also investigated the respondents' knowledge and their awareness about existing cost estimation methods. Analysis from this study will be used to determine the nature of the work and how the projects are managed in the company. The result obtained is hoped to be able to help to educate them and to improve their cost estimation practices. The survey is only involved with the respondents having previous or current experiences in cost estimation process in a project within the company. The data collection will only cover the business development branch, which was directly involved in cost estimation.

This paper will focus on the quantitative study of choosing the best cost estimation method(s) for

---
[*] Corresponding author. *E-mail address*: haslinams@utm.my

implementation at the company. Section 2 will present the study done on the factors affecting the current cost estimation activities. The various cost estimation methods are explained in section 3. The method to find the suitable cost estimation model is written in section 4. Section 5 will discuss the findings and the issues. Finally section 6 will conclude on our findings.

2. Literature Review

According to [1], software cost estimation consists of predicting the effort process and cost needed to develop a software system. It is associated with project duration and the total resource needed to complete the project. Cost estimation is based on an approximation of the possible program or product cost that was calculated using the information available. Software cost estimation basic input is a set of cost driver and the coding size. Effort is the software cost estimation output in paradigm of Person Months [2].

Realizing the importance of cost estimation, many companies are taking the effort practicing the cost estimation process in a software project. The significant factors that lead to the failure of software development project are mostly caused by wrong estimation of project's schedule and budget.

Starting with the proposal phase, software cost estimate continues until the last phase throughout the software development life cycle. The process of estimation involves effort estimation, size estimation, schedule development for initial project and total cost of the project estimation [3].

There are several methods to assist software developer to estimate better cost and effort required for software development projects. Accurate cost estimation of software development is very crucial. According to [4] accurate cost estimation process is essential for defining the software product resources. This element is also needed to control the software development activities and to sustain the product re-planning. By underestimating this feature, the project might encounter budget overrun and the expected payoff cannot be met. This could also lead to project cancellation before its completion. Future user cooperation could be discouraged and will indirectly compromise the company's management credibility.

The key to project success lies in the precision of project cost estimation. The imprecision in planning and estimation will lead to the biggest concern in software development which is software project failure. Different cost estimation approach applied might produce different accuracy of the estimated results.

This is coinciding with the study by [5] that poor estimation is the root factor of software development failure. Among the significant factors that we believe to be contributing to project cost are as follows:

A. Good Project Management

 Software project has to be completed on time, within budget and resources to achieve successful project delivery. Project manager involved in planning and estimation process must properly plan how to deliver the software product on time because schedule overrun is one of the main cause affecting software development cost. The quantity of resources needed must be defined early in the software development for more accurate cost estimation. Improper planning may lead to inaccuracy of cost estimation thus leading to over budget [6].

B. Conformance to Requirements

A research done by [7] stated that insufficient requirement, poor project planning, abrupt changes at the early project phase and imprecise estimation were the main reasons that causes software development to fail. Software project needs to undergo requirement changes to ensure that it is parallel to business process needed. Inconsistency and unspecified requirements may upset software development cost. Requirement is subject to changes. The development team must ensure that all requirements are well defined and stated at the beginning of the project to avoid the requirement from changing too frequently later on. If the changes of the requirement made cannot be controlled the project development will encounter difficulties and this will stretch the project duration that will later escalate the development cost [8].

C. Product Complexity

The complexity of the software to be built can influence the cost of the development. The more complex the software is, the more cost and resources will be needed to complete the software development within the estimated schedule. The increase in software complexity makes it more difficult to have an accurate estimation of software cost and effort. Most large projects went over budget, did not deliver the project on time and some of them could not be completed due to misleading estimation of software development cost [9].

There are two approaches in cost estimation which are algorithmic and non-algorithmic. Algorithmic models take into account arithmetic formula. Means and standard deviation are the straightforward mathematics equation that by such summary for this algorithmic model [10]. As for others, the methods are based on regression models and differential [11].

3. Software Cost Estimation Methods

There are two approaches in estimating cost for a software project:

A. Non Algorithmic Method

Among the non-algorithmic cost estimation methods mentioned by [12] are as follows:

i. Estimation by Analogy

Analogy can be defined as a comparison between things that have similar features, on which a comparison may be based. It is a cognitive process that transfers information from a certain subject to another. From a source to a target. According to [13], analogy refers to some similar completed software project and estimation of effort and cost, are based on that actual software. As for the result, the cost and effort of a similar project type can be estimated [14] also state that the central parts of the analogy based estimation is the similarity function. The similarity level between two diverse projects was measured using this estimation method. A completed project is needed in analogy costing so that the sensibility by analogy using the actual cost of earlier project with the new project are similar as a compulsory requirement. The basis estimation of current project will be derived using the key metrics from a similar type of previous project that has been conducted.

ii. Expert Judgement

This method engages discussing with one or more professionals using the knowledge gained from past project management, own methods and experiences, according to [15].

The estimation process can be established from a non-explicit and non-recoverable reasoning method for example by intuition [16]. Through the work of [17], the author has introduced the Delphi technique as a solution in minimizing the inconsistencies during estimation process.

iii.   Parkinson

The Parkinson method presented a method where the cost is determined by the available resources instead of depending solely on the objective assessment estimation. Example of the calculation is if a software has to be delivered in 18 months with 5 available people, the effort is estimated to be 90 person month. According to [12], this method might provide estimates that are not realistic even though at certain times it can estimate well. On top of that the method is not promoting good software engineering practice.

iv.   Price to Win

The software cost is estimated to be the best price to win the project. The estimation is based on the customer's budget instead of the software functionality. This is not a practical solution as there may be delays or hiccups during the project that would severely affect the project. [18] believed that "price-to-win" method would not be declared by the managers even though it can sometimes work. The managers looked at this method lightly and they were oblivious of the client's expectations on effort estimates.

v.   Bottom up

This method analyses all individual work packages and activities with the greatest level of detail [19]. Later the individual estimates will be accumulated together to derive a higher level estimates. If the individual work packages and activities are defined in detail, it will be benefit in optimizing the accuracy of bottom up approach. Using this approach, each of the software system's component is estimated individually. Then the estimation of all the components are added up to obtain the complete system's estimate. The initial design must be presented as a benchmark to the system's decomposition.

B.   Algorithmic Method

Algorithmic models vary widely in mathematical sophistication. These methods are based on arithmetic models that derive expenditure estimates as a purpose of a digit of variables which are measured to be key rate factors. The Effort parameter depends on the cost factor used for the calculation.

$$Effort = f(x_{(1,)} x_{(2,)} \cdots x_{(n)})$$

where { x1, …, xn } represents the expenditure factors. The difference of these method lies on two aspects which are the choice of expenditure factor and the form of the function.

The expenditure or cost factor that personalized the model are as follows:

There are many other cost factor other than software size. The most complete set of cost factors was proposed and used by Boehm et al [20] in the COCOMO II model. There are four types of cost factors as follow:

i. Product factors
This factor requires consistency; invention density; database range used; necessary reusability; records equivalent to life-cycle requirements.

ii. Computer Factor
This factor requires implementation time limit; key storage limitation; computer revolve constraints and stage precariousness;

iii. Personnel factors
This factor requires analyst ability; purpose occurrence; programming competence; platform knowledge; language and tool occurrence and personnel stability;

iv. Project Factors
This factor requires multisite expansion; apply of software tool and obligatory expansion plan.

The above factors are not essentially independent. They are used depending on the equation characterizing each model.

4. Methodology

A quantitative approach has been selected to conduct the work. Questionnaires have been chosen as it is the most widely used data collection method in educational research. The principles for designing a questionnaire are straight forward. However, careful considerations are also necessary. To design a systematic questionnaire, the content and its transformation into question need to be very well-crafted to minimize error.
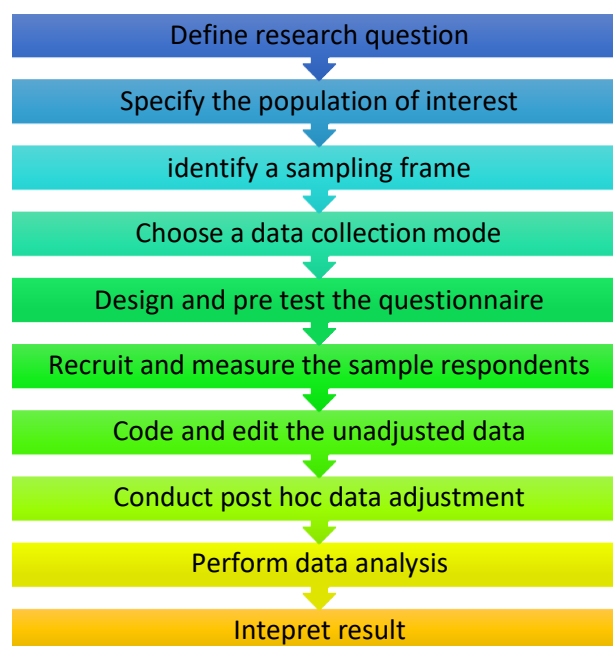
Figure 1: Research Steps

Two research questions have been initially derived. These research questions are marked as constructs. The research question was concentrated on the cost estimation method practiced in the company. It also investigated on whether the estimation process used at the company followed any form of guideline.   The relationship between all constructs and question objectives is shown in Table 1.

This survey will provide the existing condition of cost estimation practices and furthermore to provide the personnel with the best practices in cost estimation process for their further use.   For this study, we selected a focused group in the company. This group is limited to the personnel having experience and those who are directly involved in the cost estimation process within any projects.

In this study, the data collection mode has been categorized into three sections: A, B and C. For Section A, the respondents were asked questions based on demographic information. Their background and personal details were retrieved from this section. While for Section B, the questionnaire is more focused on the respondents' knowledge on cost estimation. For the last section, the data on guidelines in cost estimation method will be retrieved.

After finalizing the data collection, the questionnaire was designed and a pre-test was conducted. The process towards performing data analysis then continued by recruiting and measuring the respondent samples. The overall process included codes and editing the unadjusted data thus conducting post hoc data adjustment. The data analysis was then performed and we interpreted the results. These research steps can be viewed in Figure 1.

Table 1: Relationship between constructs and question objectives.

No	Project Question	Project Objective	Project Step	Deliverables

| 1 | What is the cost estimation method practices in Company X? | To investigate the current method used in cost estimation using survey questionnaire | 1. Research question definition
2. Population of interest specification
3. Sampling frame identification
4. Data collection mode chosen
5. Questionnaire designed and pre-tested
6. Sample respondents recruited and measured
7. Unadjusted data coded and edited
8. Conduct post hoc data adjustment
9. Perform data analysis
10. Interpret result | Questionnaire |
| 2 | Is there any guidelines for cost estimation in Company X? | To develop a cost estimation guidelines | Clarify all the elements into the guidelines as follows:
1. Introduction
2. Purpose
3. Cost Estimation Types
4. Cost Estimation Method
5. Cost Estimation Principle
6. Classification level of estimates
7. Framework for Cost Estimation project submission. | Project Cost Management Guidelines |

The following approach was taken in order to complete our quantitative study on the above issue:

A. Designing the Questionnaire

A pilot test was conducted to the focus group in the initial phases of constructing the questionnaire to understand how people think about an issue or comprehend a question. The questionnaire was entitled Software Cost Estimation Process within Company X. The questionnaire was divided into three sections which are Section A: Demographic Information, and Section B: Knowledge of Cost Estimation Methods.

The demographic part of the questionnaire chosen include the gender, the age – which is spread out into 4 categories (20 to 30 years, 31-40 years, 41-50 years and 51 years and above), the working experience which is divided into 5 categories (1 to 5 years, 6 to 10 years, 11 to 15 years, 16 to 20 years and more than 20 years), their departments, job classification in terms of job title or position and highest level of education.

In the knowledge section of cost estimation, among the questions asked are the respondent's

familiarity with the process of cost management, his involvement in the project, his knowledge on the algorithmic and non-algorithmic estimation methods. They are also asked about the frequency of using cost estimation methods during a given project and whether the method used is algorithmic or non-algorithmic.

B.   Developing the Questionnaire

There are 6 principles in the questionnaire development that we followed:

i.   Deciding what to include

This questionnaire tries to figure out the cost estimation method practices applied in Company X and any existing guidelines to assist on selecting the suitable cost estimation method.

ii.   Selection of areas to ask question about

For cost estimation method study, the area of investigation focused on the clearly defined measurable entities that are algorithmic method and non algorithmic method.

iii. Construction of question
During the construction of question, the 4 stages of Lazarfeld's scheme for measuring concept were applied. Lazarsfeld Scheme for measuring concept involves these main steps which are imagery, concept specification, choice of indicators and formation of indices [21].

iv.   Types of question

An open-ended question gives the opportunity to respondents to express their opinions in a free-flowing manner. The questions do not offer predetermined set of response and the respondents were free to answer whatever the respondent feels right in their own words. By doing this, a true insightful and even unexpected suggestion could be retrieved. This is the approach for our study.

v.   Questionnaire layout

The layout for the questionnaire have been standardized into a two column view. As, for the questions, it was organized in tabular form with a landscape layout to provide respondent a better view. The question was specified into three section and the choices of answer was put in box for best arrangement.

vi.   Increasing response rates

For increasing response rate, it will depend on the presentations of the questionnaire itself. The font type, the font size and the colour of the questionnaire do play an important part in for this matter.

C. Construct Operational Definition

According to [22], construct in questionnaire define the flow of the questionnaire by describing questionnaire as an art in administration of the questionnaire, wording and ordering decision, formatting and response categories, construct is the key that represent another form of indirect research question. Table 2 shows the construct operational definition for software cost estimation practices survey in the company.

Table 2: Construct Operational Definition

Terms	Definition of Terms	References	Scope
Familiarity	Closed connection with **knowledge** or something in terms of **awareness**, **experience** and **understanding** In this context, the respondent familiarity of cost estimation method usage are argued In terms of their knowledge, awareness, experience and understanding of software cost estimation method	(Van de Walle, Turner, & Davenport, 2003) [23]	Software cost estimation method , Software project management
Involvement	**Participation**, action and involvement in an activity. Also refer as **attachmen**t to something. In this context, the respondent participation and attachment in the software cost estimation process through cost estimation method.	(Bullock, 2010) [24]	
Non Algorithmic	Method **with no formula**. Contain specific steps and standard recognition. Example : Analogy, bottom up, Top Down, Price to win, vendor bid and etc	(Pushkar & Kumari, 2013) [3] (Kumar, 2014) [1]	
Algorithmic	Method **with equation.** The mathematical equation are based on research and historical data. Example: SLOC, COCOMO, COCOMO II. Function point, SEER SEM, Putnam and etc.	(Wu, 2014) [19] (Pushkar & Kumari, 2013) [3] (Kumar, 2014) [1]	
Frequency	**Repeatable** over a particular period of time.		
Guidelines	A **manual** of general rule principle, **guidance**, suggestion and **instruction**.		

D. Conceptual Framework

Conceptual framework shows the relationship between all constructs and question objectives. Figure 2 depicts the framework for software cost estimation.
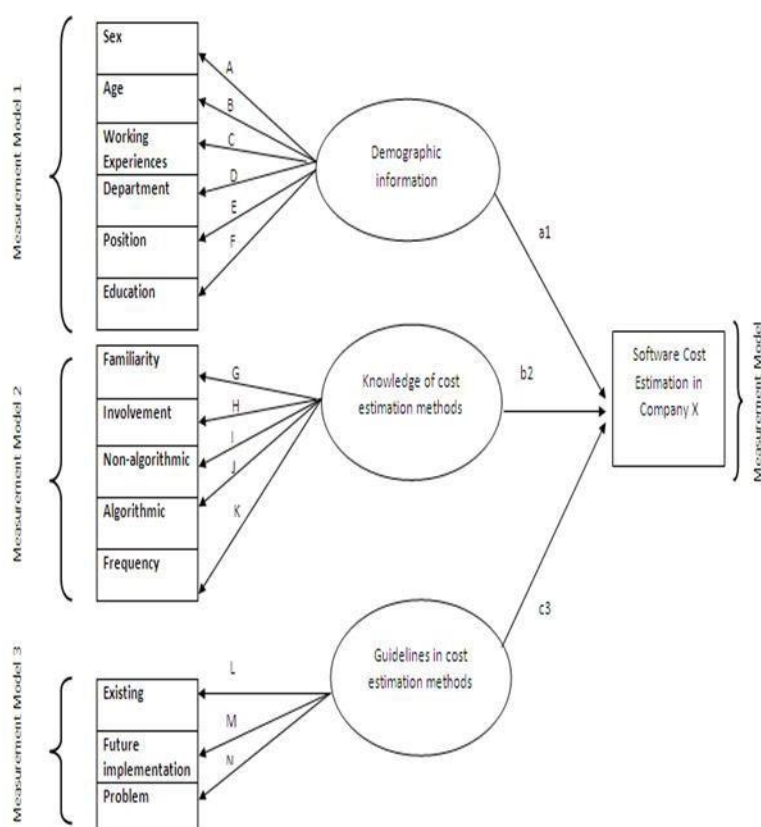
Figure 2: Cost Estimation Conceptual Framework

Figure 2: Cost Estimation Conceptual Framework

E. Variables and Measurement

The variables for the cost estimation were identified. From the construct and item measurements, the question are developed and level of measurement was clarified

A quantitative approach among business development personnel in Company X on their project cost estimation practices within company has been done. The studies are accomplished by using the quantitative method through customized questionnaire. The survey questionnaire objectives are as follows:
  i. To identify about the cost estimation method practices in Company X
  ii. To identify respondent level of knowledge about Project Cost Management process
The survey questionnaire is divided into two sections which are:

  i. Section A: Demography Information
  ii. Section B: Knowledge of Cost Estimation Method

The result is based on survey questionnaire result that have been distributed among business development personnel in the company during a month. Among 100 sets of questionnaire that have been distributed, there are only 70 set of questionnaire that are competent to be analysed. 17 of them are return incomplete while the rest are return unanswered.

5. Discussions

The results of the questionnaire were collected and counted accordingly in order to produce the precise analysis. 2D and 3D pie charts were selected to represent the results. The data were analyzed and converted to percentage value for better presentation and comparison purposes.

A. Section A: Demography Information

There are 6 questions here to identify the demographic information of the respondents. The questions cover the demographic value from sex, age, working experience, department, job title and education background. The involvement of male respondents is higher than the female by 2%. This can be visualized in Figure 3, which combines the respondents' age 21-40 years old.
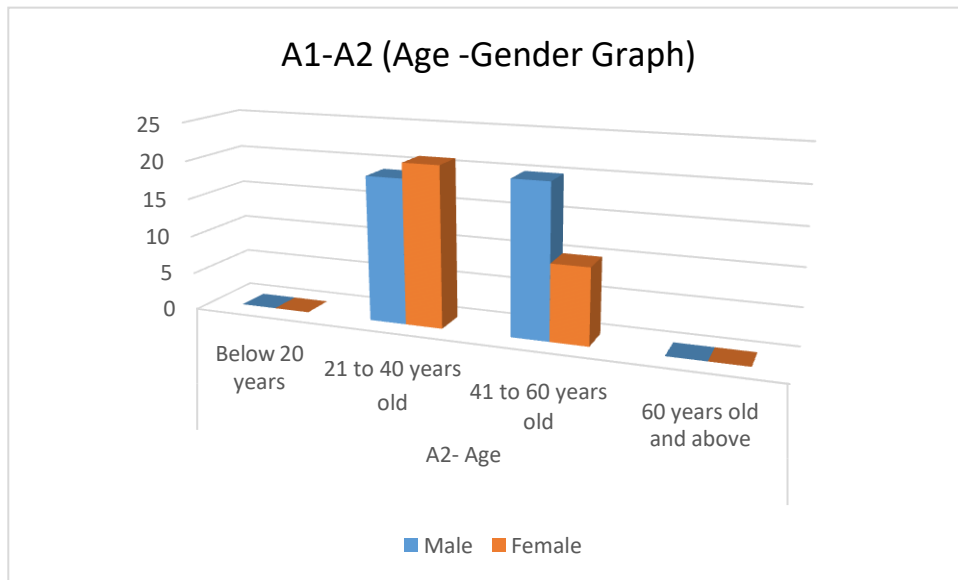
Figure 3: Age-Gender Graph

Since the company practices recruitment by contract basis, most of the respondents have a minimum working experience of 1 to 5 years as shown in Figure 4 with 42.9% result where most of them have a bachelor's degree.
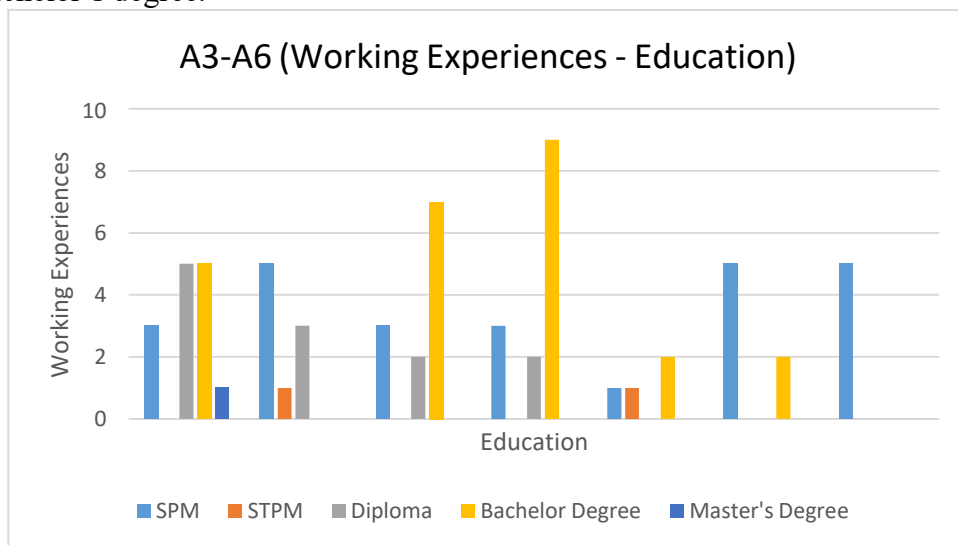

Figure 4: Working Experience-Education Graph

B. Section B: Knowledge of Cost Estimation Method

8 questions have been chosen to identify the respondent's knowledge on cost estimation method. The question's objectives are to investigate the familiarity of the respondent with cost management process, the respondent's involvement in the project, his knowledge on algorithmic and non-algorithmic cost estimation method, the frequency of usage of algorithmic and non-algorithmic cost estimation models and lastly assessing the fundamental information of algorithmic and non-algorithmic cost estimation models.

Figure 5 depicts the respondents' familiarity with the cost management process are moderately equal among all terms. 24% of respondent are familiar with the terms of estimate cost compared to other cost management process terms.
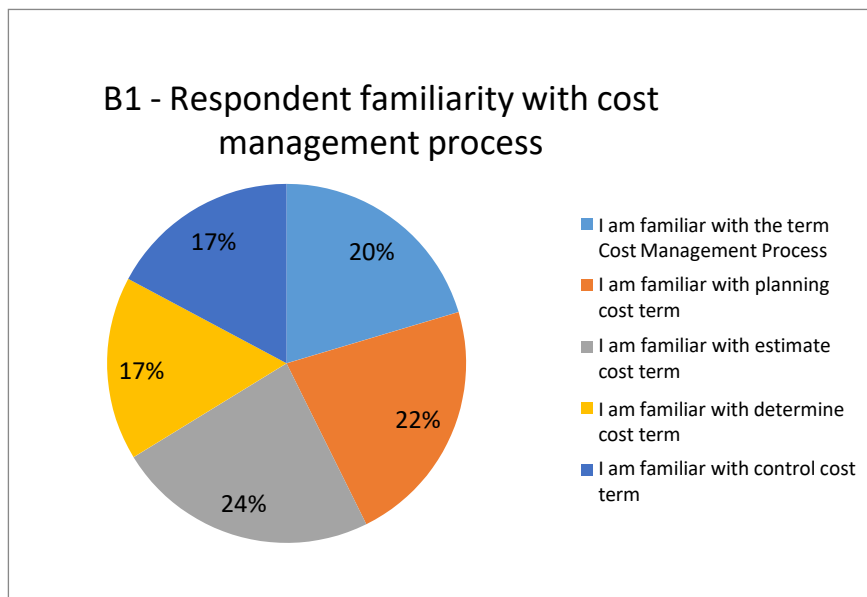

Figure 5: Respondent familiarity with cost management process

The involvement of the respondent in a project and cost estimation application for the project is also suited with each other by 44% versus 56% shown in Figure 6.
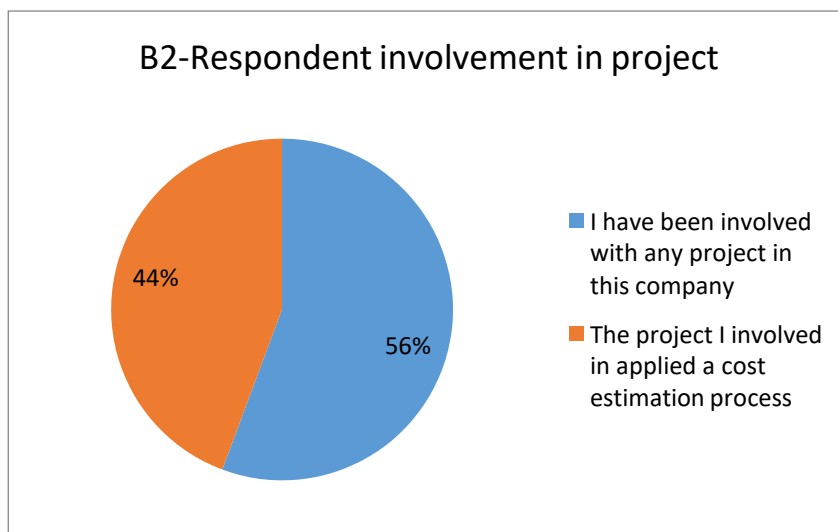

Figure 6: Respondent involvement in project

From the respondent knowledge of cost estimation scope, the results are moderately at par for non-algorithmic cost estimation method. However, there is a big difference between COCOMO and Seer Seem algorithmic cost estimation method with 39% gap.

The survey also obtained feedback from the respondents about the frequently used algorithmic and non-algorithmic cost estimation in the company. From Figure 7 and Figure 8, it was discovered that 31% of price to win method are used, followed by bottom up and estimation by analogy methods. Both at 22%. As for the algorithmic method, majority of 67% of the respondents used COCOMO more frequently compared to other algorithmic method.
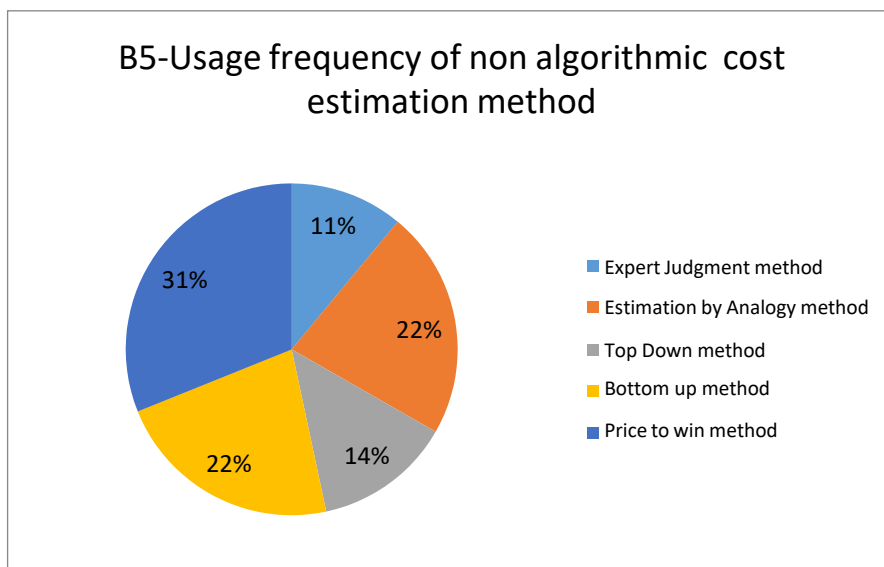

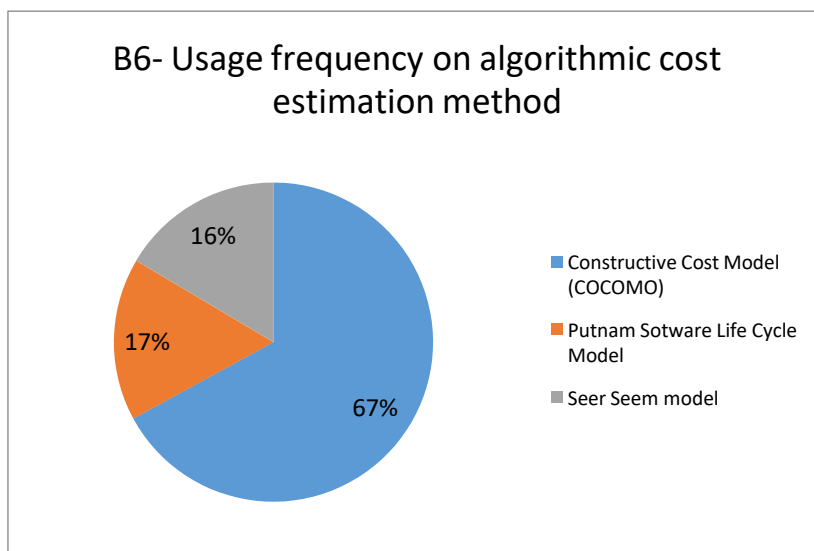Figure 7: Usage frequency of non-algorithmic cost estimation method

Figure 8: Usage frequency of algorithmic cost estimation method

Majority of the respondents have clear view of the fundamental information of cost estimation method especially for Price to win and algorithmic methods with 22% for Price to win and 42% for algorithmic. However there is still lacking of respondent knowledge on Delphi method and Work Breakdown Structure for expert judgment base to estimate, with the lowest percentage of 3% and 2 % consecutively.

6. Conclusion

From the questionnaire result and analysis, it can be concluded that majority of respondent have knowledge and aware of the cost management process especially in cost estimation. Also, the result revealed that most frequent used cost estimation method is Price to win method. This is due to nature of business for Company X which is contract based jobs that depends on winning the contract to develop and make profit.

Also, majority of respondent do aware of the importance of proper guidelines in order to lead the cost estimation process in a matured and standardize basis. The instructed guidelines really bother the respondent in manually think of the solution rather than referring to guidelines for any cost estimation related issue. A proper general and structured guideline should be developed and proposed as a cost estimation process manual in the future for better process and implementation.

References

[1] Y. Kumar. "Comparative Analysis of Different Software Cost Estimation Methods", 3(6), 547–557. 2014

[2] S. Kumari. "Comparison and Analysis of Different Software Cost Estimation Methods" 4(1), 153–157. 2013.

[3] S. Pushkar, and S. Kumari. "Cost Estimation Methods: A Review. Performance,". International Journal of Advanced Research in Performance Analysis of the Software. 3(7), 229–238. 2013.

[4] L. Lederer and U. Kentucky. "Causes of Inaccurate Cost Estimates Software Development", 1212(94), 125–134. 1995.

[5] M. Jorgensen, B. Boehm, and S. Rifkin. "Software Development Effort Estimation: Formal Models or Expert Judgment?" IEEE Software, 26(2), 1–8. http://doi.org/10.1109/MS.2009.47. 2009

[6] Z. Mansor, S. Yahya, N. H. H. Arshad, "Success factors in cost estimation for software development project". 2011.

[7] D. Galorath, and M. W. Evans. "Software Sizing, Estimation, and Risk Management: When Performance is Measured Performance Improves". 2006.

[8] R. Lagerstrom, L. Marcks von Wurtemberg, H. Holm and O. Luczak "Identifying factors affecting software development cost" Industrial Information and Control Systems, The Royal Institute of Technology, Stockholm, Sweden. 2012.

[9] P. Pandey, "Analysis of the techniques for software cost estimation" Third International Conference on Advance Computing & Communication Technologies , 2013.

[10] W. S., Donelson. "Project Planning and Control", Datamation, 22(6), 73-80. 1976.

[11] L. H. Putnam. "A General Empirical Solution to the Macro Software Sizing and Estimating Problem," IEEE Trans. Software Eng. 345-361. 1978.

[12] S. Lee, L. Titchkosky and S. Bowen. "Software Cost Estimation." 2010.

[13] V. Khatibi and D. N. Jawawi. "Software Cost Estimation Methods : A Review". Journal of Emerging Trends in Computing and Information Sciences, 2(1), 21–29. 2010.

[14] Y. F. Li, M. Xie, and T. N. Goh. "A study of project selection and feature weighting for analogy based software cost estimation", The Journal of Systems and Software 82, 241–252. http://doi.org/10.1016/j.jss.2008.06.001. 2009.

[15] C. Rush and R. Roy. "Expert Judgement in Cost Estimating". Modelling Process, 271–284. 2015.

[16] P. Faria and E. Miranda. "Expert Judgment in Software Estimation During the Bid Phase of a Project – An Exploratory Survey", Software Measurement and the 2012 Seventh International Conference on Software Process and Product Measurement, 2012.

[17] B. W. Boehm. Software Engineering Economics. 1983.

[18] K. Moløkken-Østvold, M. Jørgensen, S. Tanilkan, H. Gallis, A. Lien, S. Hove. "A survey on software estimation in the norwegian industry". Proceedings of Metrics '04, 2004 pp. 208–219. 2004.

[19] L. Wu. The Comparison of the Software Cost Estimating Methods. 2014.

[20] B. Boehm. "COCOMO II Model Definition Manual. The American journal of tropical medicine and hygiene", 87, p.i. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23632284. 2012.

[21] P.F. Lazarsfeld. "Evidence and Inference in Social Research." In. Lazarsfeld, P.F & Rosenberg, M. (Eds.). Language of Social Research, pp. 99-128. Free Press. 1955.

[22] E. Martin. "Survey Questionnaire Construction." Available at : http://www.census.gov/srd/papers/pdf/rsm2006-13.pdf. 2006.