Ranking XP Estimation Methods Based on The ANP

Abdulmajeed Aljuhani*, Luigi Benedicenti†, Sultan Alshehri‡
*Software Systems Engineering Program, University of Regina, Canada
Email: aljuhana@uregina.ca
†Faculty of Computer Science, University of New Brunswick, Canada
Email: Luigi.benedicenti@uregina.ca
‡Computer Science and Information Technology College, Majmaah University, Saudi Arabia
Email: su.alshehri@mu.edu.sa

Abstract

The analytic network process (ANP) is considered one of the most powerful tools to facilitate decision-making in complex environments. The ANP allows decision makers to structure their problems mathematically using individual judgments. Researches suggest that ANP can be useful in software development, where complicated decisions happen routinely. This paper explains the ranking of extreme programming (XP) estimation methods using the analytic network process. A case study that were conducted in an academic environment is presented in this paper. The results of the case study show the benefits of applying the ANP in XP development cycle.

Keywords

Analytic Network Process, Extreme Programming, Planning Game, Estimation Techniques, User Stories.

I. INTRODUCTION

Developers face uncertainties when designing software projects. It is therefore important in the software process to minimize these uncertainties. In extreme programming practices, there are certain activities that aid this process, one being the planning game practice. In this practice, XP team members meet together to identify the system requirements. These requirements are written as user stories. According to Cohn user stories are "short descriptions of functionality told from the perspective of a user that are valuable to either a user of the software or the customer of the software"[1]. These user stories are significant because they make it easy to structure a general framework for the system. User stories structure the desired software product by testing the designed software against identified user stories. In addition, user stories contain different activities such as writing and breaking stories down into tasks. Also, they include prioritizing user stories according to customer business value. Moreover, one of the most important activities is estimating user story effort and cost.

During iteration planning meetings, estimates are used to establish the iteration plan. This occurs by assigning stories to each iteration based on their priorities. Using story points the development team evaluates user stories to specify the cost and complexity of the implementation. After that, developers break down the user stories into small tasks.

In this paper, the analytic network process (ANP) is used to assist the XP team in order to rank the estimation methods based on several criteria. These methods are Planning Poker, Expert Opinion, Analogy, and Disaggregation.

II. RELATED WORK

Agile methodology is based on collaborating among team members and therefore, estimations are not obtained by individuals. All team members share estimations, and as Cohn states, "estimates are best derived collaboratively by the team, which includes those who will do the work"[2]. It is important during estimations of user stories to create appropriate stories as these can be estimated clearly. For example, one could combine related stories into a group, which is called theme [2], in order to clearly estimate them as one item.

There are several methods that can be used to software estimation. Some methods mathematically obtain estimate by concentrating on historical data [3]. Other methods estimate effort by measuring the size of the tasks [3]. Expert opinion is one that works by asking an expert about each story. Based on his/her experience, the expert then gives an estimate. Another technique that can be used to estimate user stories is analogy. This technique can involve a triangulation process, which depends on comparing the story that is being estimated with two other stories [2]. Disaggregation is another method that is used to split large stories into smaller ones in order to estimate them.

Some studies suggest that the best technique that can be used to estimate stories is planning poker. This merges the three above-mentioned techniques of expert opinion, analogy, and disaggregation. All team members are involved in planning poker in order to estimate stories and this involves several steps and stages which are explained in more detail in [2].

In [3] the authors introduced several areas that are useful to estimate using expert opinion. These areas are:1) areas where it is difficult to find empirical data, and 2) when it is difficult to estimate because of the lack of understanding problems [3]. These mentions areas are considered reasons for widely applying of expert opinion method in software estimation. In [3] the authors conducted an industrial experiment in order to evaluate the reliability of using expert opinion method in cost estimation. The study aimed to enhance cost estimation in a medium size software organization. According to Faria and Miranda "the intention of the organization in supporting this study was to assess its cost estimation capability in the bid phase of a project and, if required, to use the results as a catalyst for change of their estimation practices"[3]. In this study, an on-line survey was distributed among thirty employees. 47% was the response rate of the survey, and 7.5 years was the average experience of the participants. The research questions of this study focused on two areas variability and calibration. The study contained three research questions asking about several issues; for example, the authors asked about the possibility of having the same estimate from various group of estimators. After presenting the results, the authors concluded with that expert opinion method obstacles are inconsistencies and overconfidence. This issue causes unreliable software estimation. However, the authors suggested integrating expert opinion method with another technique such as Wideband-Delphi in order to make expert opinion estimation more reliable.

Heemstra [4] conducted a survey in 364 organizations and found that only 51 estimated efforts used models. However, "the model users made no better estimate than the non-model users"[4]. The researcher concluded that expert opinion was better than the use of estimation models.

In [5] the authors investigated effort estimation in 32 software projects in different Iranian software companies. Questionnaire was distributed in order to collect data these companies and their project estimation. The study shows that most of these companies depend on small teams (Five people and less) because of the small scale of develop products, and the simplicity in managing small teams. Also, the most used development processes in product development are Rational Unified Process and Extreme Programming. The distributed questionnaire includes various parts such as the type of estimation method used in a project, and the degree of differences between the data estimated and real data [5]. Based on this study,expert opinion and analogy method are from the most common estimation methods. Among all these projects, the percentages of using analogy method and expert opinion are 29% and 25% respectively [5]. The study concluded with that these two methods are simple to apply to similar projects and provide rapidly estimation. However, the major defect is the dependence on expert judgment, which may be less reliable.

In [6] the authors conducted a study in order to evaluate the accuracy of planning poker estimation method. 13 students are formalized in teams in order to develop "a web-based student records information system"[6]. All students received the same user stories, and asked to implement them in three sprints. Also, by using planning poker method, students estimated the user stories and "the estimates provided by each team member during the first round were averaged to obtain the statistical combination for further comparison"[6]. In the same time, a number of experts were given the same user stories in order to provide their estimations. The results of this study show that planning poker method lead the students to have over optimistic estimation, while the experts estimation was closer to the actual effort [6]. The study addressed that planning poker method is less benefit when it is applied by less experience developers.

Williams *et al.* [7] investigated XP practice development in an IBM group. They concluded that the XP product has improved quality of pre-release and post release. The XP team noted enhancements in their effort estimation, schedule and productivity. Also, customer satisfaction was very high with the XP product since developers delivered more than what was planned.

Finnie and Wittig [8] estimated effort by applying artificial neural network (ANN) and case-based reasoning. They concluded that by "using a data set from the Australian Software Metrics Association, ANN was able to estimate development effort within 25% of the actual effort in more than 75% of the projects"[8].

In [9] the authors emphasized several challenges in current estimation methods. For example,"it cannot be easily related to the time duration because story points represent the amount of work, and the velocity differs from team to team"[9].

III. METHODOLOGY

The main objective in this research is to investigate how the analytic network process might be used to rank the estimation methods in order to deteremie the most suaitable one for the software project. The case study methodology, which is explained in [10], is the research methodology.

The following research questions provide more focus for the research case study:

1) How can the analytic network process assist the XP team in ranking the estimation methods in order to estimate each user story?
2) How does the ANP influence the development team's communication and productivity?

Moreover, the study propositions are as follows:

Proposition 1: *The ANP catches significant criteria and alternatives that have effect in ranking XP estimation methods. Also, the results of using the ANP display the order of alternatives and criteria based on their importance.*

Proposition 2: *The ANP includes creative debate and enhances team communication.*

Proposition 3: *The ANP clears up conflict perspectives between the development team within the ranking process.*

From the above questions, we derived the units of analysis for our study. The main objective is ranking various XP estimation methods that can be applied to estimate each user story. Appropriately, evaluating and ranking are two units of analysis. Another is the participants' perspective of the ANP benefits in planning game practice. Therefore, the design of this case study includes multiple cases, embedded with multiple units of analysis. The logic linking of the collected data to the study propositions is shown at the end of this paper.

IV. Data Collection and Sources

At the beginning of each use for the ANP in extreme programming, we investigate the ANP benefits and ability by introducing the related criteria and extreme programming areas. Data is collected from searching previous studies and literature review. As well, data triangulation is acquired in order to increase the validity of the study.

The major data source of this research is an extreme programming project, conducted during the winter semester of 2016 at the University of Regina. The data sources in this research are:

- Questionnaires given to the students during the development of the XP project.
- Archival records, such as study plans, from the students.
- Comments from the customer.
- Open-ended interviews with the students.

V. Case Study

The case study was conducted during a 12-week Winter 2016 semester at the University of Regina. Several studies, like [11], [12] and [13], addressed that the suitable XP team size is between three and seven members. Moreover, Ambler [14] emphasized that the success of agile project is 83 % with team size less than eleven members, and the percentage goes lower with increasing the team size for more than eleven people [14]. The major cause of this reducing in the success percentage is regarding to communication lack or misunderstanding with the large team size. Therefore, we had 12 graduate students from the University of Regina, and one additional participant, a client, who were included in this case study. These students had intermediate knowledge of extreme programming process and practices, and different programming levels. The majority of these students was part of a professional program, meaning that their graduate degree was part of their professional development and that they had previous employment experience in the software industry. Some of these students were continuing to work part-time. The participants' backgrounds included various programming languages such as C++, Java, and PHP. The participants were organized into two teams, the first team used the ANP method in order to make their decisions in ranking the estimation methods, and the second team followed the traditional XP method for their decisions. Both teams were asked to develop a project called "Professors' Availability Managing System" complete with a set of requirements. The project was developed in 5 iterations, allowing two weeks for each. At the end of the project, the two teams implemented all system requirements. The participants were asked to evaluate all user stories in each estimation technique before using the ANP or the traditional way in order to rank them. Assistance materials that focused on planning game practices were given to the participants in order to ensure their understanding. These materials involved estimating user stories, writing user stories, and making programming commitments. The ANP team was given white papers, several presentations, and other important materials about the ANP in order to allow them to apply it in their development. Team 1 practiced on several pairwise comparisons and increased their understandings of the ANP structure. At the end, the researcher handed out a survey to the participants in order to collect more data about the participants' perspectives.

VI. The ANP

According to Saaty [15] "the Analytic Network Process (ANP) is a multi-criteria theory of measurement used to derive relative priority scales of absolute numbers from individual judgments (or from actual measurements normalized to a relative form) that also belong to a fundamental scale of absolute numbers"[15]. The ANP provides a structure to present a solution for a certain problem, which leads to a decision for that problem. In the ANP method, dependencies among various criteria are considered making it different from the Analytic Hierarchy Process (AHP) [15]. Saaty states [15] "in fact the ANP uses a network without the need to specify levels. As in, the AHP, dominance or the relative importance of influence is a central concept. In the ANP, one forms a judgment from the fundamental scale of the AHP by answering two kinds of questions with regard to strength of dominance:

1) Given a criterion, which of two elements is more dominant with respect to that criterion,
2) Which of two elements influences a third element more, with respect to a criterion"[15]?

In pairwise comparisons, entered values reflect the relative effect among elements with respect to a control criterion. These entered values are based on the importance of each criterion. As such, "the ANP is a useful tool for prediction and for representing a variety of competitors with their explicitly known and implicitly assumed interactions and the relative strengths with which they wield their influence in making a decision. It is also

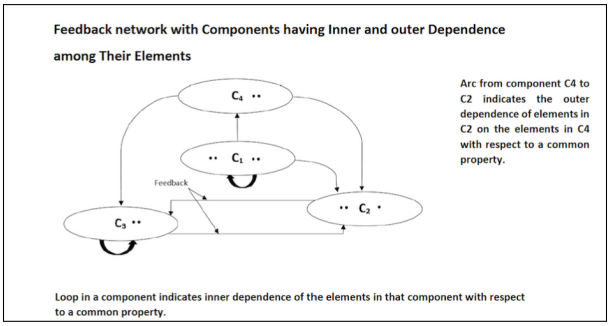

Figure 1: The analytic network process structure [16]

Table I: ANP fundamental scale developed by Saaty [17]

Scale	Numerical rating	Reciprocal
Equal importance	1	1
Moderate importance of one over other	3	$\frac{1}{3}$
Very strong or demonstrated importance	7	$\frac{1}{7}$
Extreme importance	9	$\frac{1}{9}$
Intermediate values	2,4,6,8	$\frac{1}{2}, \frac{1}{4}, \frac{1}{6}, \frac{1}{8}$

Table II: Random Index [16]

Order	1	2	3	4	5	6	7	8	9	10
R.I	0	0	0.52	0.89	1.11	1.25	1.35	1.4	1.45	1.49

useful in conflict resolution where there can be many opposing influences"[15]. The network structure consists of different clusters, and these clusters contain various nodes or elements. These clusters are connected to each other based on the relative influences among the nodes. The links can either have external relative influence, which means elements in cluster X affect element in cluster Y, or internal relative influence, which means elements in the same cluster (e.g.X) affect each other. In this case, the external relative influence is named outer-dependence, and the internal relative influence is named inner-dependence [15]. The network structure allows feedback models through the idea of cycle connection, and the ANP provides different types of nodes such as source, intermediate, and sink. Again, according to Saaty [16] "a source node is an origin of paths of influence (importance) and never a destination of such paths. A sink node is a destination of paths of influence and never an origin of such paths. A full network can include source nodes; intermediate nodes that fall on paths from source nodes, lie on cycles, or fall on paths to sink nodes; and finally sink nodes"[16]. Figure 1 gives a general idea of the ANP structure [16].

Another aspect of the ANP structure is the prioritizing of different alternatives in order to make an appropriate decision. This starts by making pairwise comparisons, based on a fundamental scale, as shown in table I. Following this, "the vector of priorities is the principal eigenvector of the matrix. This vector gives the relative priority of the criteria measured on a ratio scale. That is, these priorities are unique within multiplication by a positive constant. If one ensures that they sum to one they are then unique and belong to a scale of absolute numbers"[16]. "The consistency index of a matrix is given by C.I. (max n)/(n-1), where n is the number of alternatives. The consistency ratio (C.R.) is obtained by forming the ratio of C.I. The appropriate set of numbers is shown in table II, each of which is an average random consistency index computed for n 10 for very large samples. They create randomly generated reciprocal matrices using the scale $\frac{1}{9}$, $\frac{1}{8}$, $\frac{1}{2}$, 1, 2, 8, 9 and calculate the average of their eigenvalues. This average is used to form the Random Consistency Index R .I"[16]. The consistency ratio (C.R) should be lower than 0.10 (or 0.20), otherwise, the entered judgements need to be enhanced.

After obtaining all priorities from the pairwise comparisons, these priorities are placed in a supermatrix. According to Saaty [16] "the supermatrix represents the influence priority of an element on the left of the matrix on an element at the top of the matrix with respect to a particular control criterion. A supermatrix along with an example of one of its general entry matrices is shown in figure 2. The component C1 in the supermatrix includes all priority vectors derived for nodes that are parent nodes in the C1 cluster"[16].

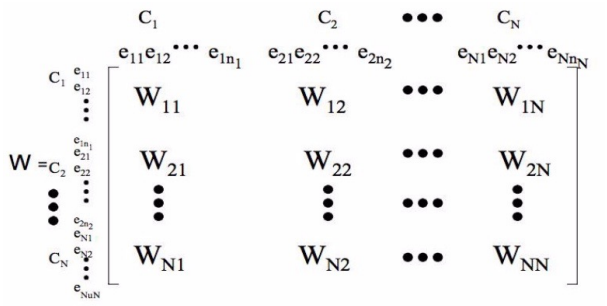

Figure 2: The Super-matrix of a network [16]

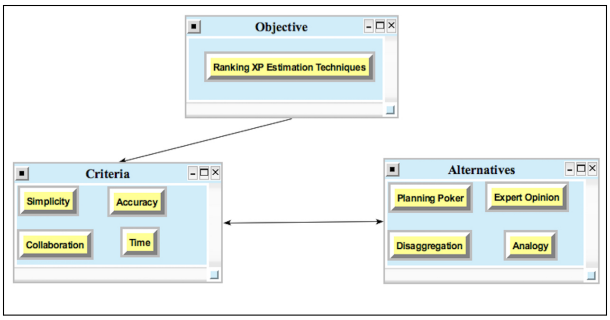

Figure 3: ANP structure for ranking XP estimation methods

VII. Proposed Criteria for Ranking

To rank the selected estimation methods, it is significant to identify the criteria that influence the estimation process. These criteria are compared with each other to show the interdependences, and also, compared with respect to each alternative or estimation method. The estimation methods are compared with respect to the criteria in order to show the feedback relation in the selection process. In this case study, four estimation criteria were proposed. These criteria are:

1) Accuracy: Which estimation technique gives the most accurate estimation?
2) Simplicity: What is the simplest estimation method to understand and to apply?
3) Collaboration: Which estimation method has the highest degree of collaboration between the team members?
4) Time: Which estimation method saves the time when estimating the user stories?

VIII. ANP in Practice

Structuring the problem as a network is the first step in the analytic network process. The ANP network contains criteria cluster, alternative cluster, and the goal. The main cluster is the alternative cluster, which contains the four estimation methods wanted to select the best among them. The three network components are: the goal cluster - selecting the best estimation method; the criteria cluster - accuracy, simplicity, collaboration, and time; the alternatives cluster - analogy, expert opinion, disaggregation, and planning poker. Figure 3 illustrates the ANP structure in this paper.

Next, the suitable ANP tables were generated, and all ANP team members received the tables. The ANP team was asked to fill out the pairwise comparisons based on the ANP fundamental scale that was described previously. General information, such as member's experience and programming level, was collected in each cover page. The ANP participants were also asked to compare the criteria among each other with respect to each estimation method.

Appropriately, the participants were asked to use the estimation techniques during the whole project development in order to practice the advantages and disadvantages of each technique. After that, the participants evaluated each estimation technique based on the four criteria. This was achieved, by giving the participants the suitable ANP tables and other supporting materials that mentioned above. Examples of the participants questions are:

- With respect to Planning Poker: which criterion is more important, simplicity or accuracy?
- With respect to Expert Opinion: which criterion is more important, accuracy or time?
- With respect to Disaggregation which criterion is more important, collaboration or simplicity?

Table III: Estimation methods ranking for Team 1

Methods	Scores (%)
Expert Opinion	35.36%
Planning Poker	33.64%
Analogy	25.50%
Disaggregation	5.48%

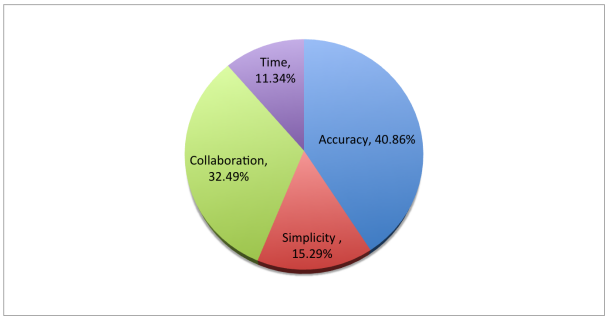

Figure 4: The importance of the criteria by Team 1

Table IV: Estimation methods ranking by Team 2

Ranking	Methods
1	Analogy
2	Expert Opinion
3	Planning Poker
4	Disaggregation

- With respect to Analogy: which criterion is more important, time or collaboration?

The participants then compared the estimation methods with respect to each criterion. Examples of questions for the participants are:

- With respect to simplicity: which method do you prefer, analogy or expert opinion?
- With respect to accuracy: which method do you prefer, planning poker or analogy?
- With respect to time: which method do you prefer, disaggregation or expert opinion?
- With respect to simplicity: which method do you prefer, planning poker or disaggregation?
- With respect to collaboration: which method do you prefer, expert opinion or planning poker?

IX. Finding and Results

In Team 1, each participant individually evaluated the estimation techniques using the ANP pairwise comparisons. Super Decisions Software [**?**] was used in order to determine the aggregation judgements for Team 1.

Team 1 results show that expert opinion method was given the highest rank among the four alternatives. Planning poker came second followed by analogy, then disaggregation. Table III exhibits the relative weight of each one as a percentage. In addition, in using the software, we are able to examine the significance of each criterion. The accuracy criterion received the highest importance among the criteria, followed by collaboration, simplicity, and time. Figure 4 shows the criteria importance scores.

For Team 2, the participants were asked to follow the traditional method in their decisions and therefore were asked to document each step in their process in terms of how and why the decision was made. Most of their decisions were made based on deep discussions and voting. Team 2 results show that the analogy method was given the highest rank among the others estimation methods. Table IV displays the estimation methods ranking by Team 2. By asking team 2 what the most important factor was for ranking the estimation methods, they gave accuracy the top score. Table V shows the ranking of the criteria by Team 2.

X. Observations

A. *ANP Ranking Results*

By considering all criteria together, the expert opinion method was ranked the highest estimation method by Team 1 and analogy was ranked as the highest estimation method by Team 2. Planning poker method was ranked in the second position by Team 1, while it was ranked at the third position by Team 2. Both teams ranked disaggregation in the last position.

Table V: The importance of the criteria by Team 2

Ranking	Criteria
1	Accuracy
2	Time
3	Collaboration
4	Simplicity

Both teams considered accuracy the most significant criterion. Collaboration criterion was considered to be the second important criterion by Team 1 while Team 2 ranked collaboration at third position. Team 1 found simplicity and time to be considered the third and fourth highest important criteria respectively. Team 2 ranked time criterion in the second position, while simplicity was the less important criterion.

Team 1 ranked the estimation methods by considering each criterion individually. The findings were: the planning poker method was ranked the highest in terms of accuracy criterion, expert opinion ranked the highest in terms of time and simplicity criteria, while analogy was ranked the highest in terms of collaboration criterion. Similar to Team 1, Team 2 ranked expert opinion technique as the best estimation technique in term of time criterion.

These results show options that were made by each team. Rankings were completed individually, however, the group act was consistent in the consistency rates.

B. Interview Results

After completing the project, the results of the ANP evaluation for ranking the estimation methods were shown to the participants in order to conduct the interviews. Not all results were as expected and some findings were surprising. The interviews involved open-ended questions in order to collect the participants' perspectives about the ANP, their perspectives on its benefits and disadvantages in XP, as well to collect their opinions about the best application for ANP in XP among all mentioned practices. The collected data was comprised of handwritten notes from the interviews.

The interview results show positive comments from the participants regarding the ANP. The ANP was a helpful tool in solving conflict perspectives, and encouraged each team member to participate in making decisions. The main concern was the time it took during the ANP evaluation, and the number of pairwise comparisons. Another recommendation was applying the ANP in more XP practices and studying the effects. All ANP team members recommended using ANP in their future XP projects.

On the other hand, Team 2 was not completely satisfied with the process of their decisions. Some of the team members complained about that the most experience member had more voting weight than others, which lead them to follow decisions that they may not like. Another issue is that the ANP allowed us to know the difference between each ranking position in a percentage; however, Team 2 could not specified the amount of difference between each ranked technique and criterion.

C. Questionnaires

Questionnaires were distributed among the participants in order to collect their experiences and viewpoints. The given questionnaires consisted of two sections. The first section included questions about ANP as a ranking and decision tool, such as capturing the needed information, goodness of the decision structure, clarity of criteria involved, and clarity of alternatives involved. The second section included questions about the benefits of each extreme programming practice, and the students' satisfaction, such as enhancing the team communication, clarifying the ranking problem, creating positive discussion and learning chances, team performance, and satisfaction of the final results of the ANP. In this study, a seven-point Likert scale was used in order to determine the acceptability level of the ANP tool as follows:

1) Totally unacceptable.
2) Unacceptable.
3) Slightly unacceptable.
4) Neutral.
5) Slightly acceptable.
6) Acceptable.
7) Perfectly Acceptable.

After completing the questionnaire, the same steps were followed as in [18] in order to aggregate the collected data and display the total acceptability percentage. The total acceptability percentage can be obtained as follows:
The total acceptability percentage (TAP)= the average score x $\frac{100}{7}$.
Where the average score = the sum of all scores given by team members / number of the team members.

The following percentages show the level of acceptability for the ANP as a ranking and decision tool:

- Enhancing team communication: 82%.
- Maximizing team performance: 87%.

- Supporting positive discussion and learning chances: 72%.
- Clearing up conflict perspectives among the team members: 87%.
- Defining the ranking problem: 91%.
- Satisfaction of the ANP final results 71%.

From different data sources, the data was collected. By comparing the collected data with the study propositions based on the interpretation of the criteria that were mentioned above, we will analysis this collected data. The followings are the study propositions and their answers:

- For the first proposition, we can see that both the alternatives and criteria are structured sufficiently, and considered in figure 3. Also, the accomplish results and objectives of the ANP use in ranking the estimation methods can be seen in table III, which exhibited the ranking of the ANP team for the XP estimation techniques, and expert opinion was ranked as the highest.
- The questionnaire statement 'satisfaction of the ANP final results' supported the second proposition, and the feedback of this was positive, which is 71 %. Moreover, the statement ' clearing up conflict perspectives among the team members' supported the third initial proposition, and the score was 87 %.

XI. Validity

In this section, related threats to the validity are explained. These threats are construct validity, external validity, internal validity, and reliability. Several researchers emphasized that case studies are difficult to analyze due to biases and validity threats as described in [19] "empirical studies in general and case studies in particular are prone to biases and validity threats that make it difficult to control the quality of the study to generalize its results"[19].

A. Construct Validity

Construct validity ensures that "the treatment reflects the construct of the cause well, and the outcome reflects the construct of the effect well"[20]. It deals with matching the concept being researched and studied, to the specific measurements. The small number of participants is the main threat to this case study.

Using various methods to ensure the validity of the results reduced this threat. Some of these methods are:

- Data triangulation: a major advantage of case study is the opportunity to use several sources of evidence [21]. An evidence chain is built through using interviews and surveys with various types of participants with different skills and experience levels, and the use of participants' comments and many observations. Therefore, a valid conclusion can be reached.
- Methodological triangulation: engaging a combination of research methods such as conducting an XP project to serve the study purpose, surveys, results of ANP pairwise comparisons, researchers' observations, and interviews.
- Member checking: showing the findings to the participants is recommended. This concern was addressed by presenting the final findings to all students in order to guarantee the accuracy of the study and to avoid researcher bias.

B. Internal Validity

Internal validity is about making sure the outcome is caused by the treatment (the effect). This type of validity is only related to explanatory case study. This issue may be addressed by linking all data sources regarding the research questions, and linking the research questions to research propositions.

C. External Validity

External validity ensures the relationship between the construct and the effect in order to guarantee that the experiment will be generalized to a different scope [20]. In this study, additional case study will be need to be conducted in different environments such as industry in order to involve more experts from the field. Conducting such a case study will help in comparing the various results and findings from different environments. Future work will add to increased external validity.

D. Reliability

Reliability deals with the procedure of data collection and findings. Similar conclusions and results should be arrived by other researchers when following the same procedure. This can be done through the availability of same research questions, data collection, and case studies designed by other researchers.

XII. Conclusion

After applying the analytic network process to rank the estimation methods used in extreme programming, ANP was an appropriate and beneficial tool that gave the development team a good understanding for determining the suitable estimation method. The participants evaluated the four estimation methods with respect to four criteria, which were accuracy, simplicity, time, and collaboration. The expert opinion method was ranked as the top estimation technique by Team 1 regarding to the four criteria. The ANP helped the team in evaluating each estimation method from different perspectives. Team 2 followed the traditional XP approach in ranking their estimation methods. Team 2 members deeply discussed their choices of estimation methods, and ultimately chose the analogy method. All team 2 members participated in the discussion; however, the most experienced members were more powerful in convincing other team members of their preferences.

Using the ANP tool, the XP team was able to evaluate each estimation method with respect to different aspects. Moreover, the ANP allowed us to specify the difference between each element in our model by a percentage, while the traditional XP team were not be able to do that. Furthermore, the traditional team ranked the estimation methods by considering only accuracy criterion without considering the other criteria in their decisions. However, the ANP allowed Team 1 to rank the alternatives based on a multi criteria decision making approach, which helped the team to rank the alternatives with considering various factors. The ANP helped the team members resolve conflicts based on a structured approach grounded in scientific principles. The ANP ended up simplifying decision making, which maximized the effect of the software being developed. Team 1 members reconciled their conflicts of perspectives based on a mathematical approach. This maximized their satisfaction with the teams decisions.

XIII. Acknowledgement

Aljuhani's research is supported by the Saudi Cultural Bureau in Canada and Taibah University.

REFERENCES

[1] M. Cohn, "Advantages of user stories for requirements," *InformIT Network*, 2004.

[2] ——, *Agile estimating and planning*. Pearson Education, 2005.

[3] P. Faria and E. Miranda, "Expert judgment in software estimation during the bid phase of a project–an exploratory survey," in *Software Measurement and the 2012 Seventh International Conference on Software Process and Product Measurement (IWSM-MENSURA), 2012 Joint Conference of the 22nd International Workshop on*. IEEE, 2012, pp. 126–131.

[4] F. J. Heemstra, "Software cost estimation," *Information and Software Technology*, vol. 34, no. 10, pp. 627–639, 1992.

[5] Y. Yavari and H. Bashiri, "Effort estimation of software development: Depth view on iran case," *American Journal of Software Engineering*, vol. 1, no. 1, pp. 5–11, 2013.

[6] V. Mahnič and T. Hovelja, "On using planning poker for estimating user stories," *Journal of Systems and Software*, vol. 85, no. 9, pp. 2086–2095, 2012.

[7] L. Layman, L. Williams, and L. Cunningham, "Motivations and measurements in an agile case study," *Journal of Systems Architecture*, vol. 52, no. 11, pp. 654–667, 2006.

[8] G. R. Finnie and G. E. Wittig, "Ai tools for software development effort estimation," in *Software Engineering: Education and Practice, 1996. Proceedings. International Conference*. IEEE, 1996, pp. 346–353.

[9] R. Popli and N. Chauhan, "Research challenges of agile estimation," *Journal of Intelligent Computing and Applications(July-December2012)*, 2012.

[10] R. K. Yin, *Case study research: Design and methods*. Sage publications, 2013.

[11] A. Bustamante and R. Sawhney, "Agile xxl: Scaling agile for project teams, seapine software, inc," 2015.

[12] V. Lalsing, S. Kishnah, and S. Pudaruth, "People factors in agile software development and project management," *International Journal of Software Engineering & Applications*, vol. 3, no. 1, p. 117, 2012.

[13] B. Rumpe and P. Scholz, "Scaling the management of extreme programming projects," *arXiv preprint arXiv:1409.6604*, 2014.

[14] S. Ambler, "Agile teams making decisions: Decision making tools," http://www.ambysoft.com/surveys/success2010.html, accessed: 2015-02-24.

[15] T. L. Saaty, "Fundamentals of the analytic network processdependence and feedback in decision-making with a single network," *Journal of Systems science and Systems engineering*, vol. 13, no. 2, pp. 129–157, 2004.

[16] T. L Saaty, "The analytic network process," *Iranian Journal of Operations Research*, vol. 1, no. 1, pp. 1–27, 2008.

[17] T. L. Saaty, "Decision making with the analytic hierarchy process," *International journal of services sciences*, vol. 1, no. 1, pp. 83–98, 2008.

[18] S. Alshehri and L. Benedicenti, "Ranking approach for the user story prioritization methods," *J Commun Comput*, vol. 10, pp. 1465–1474, 2013.

[19] R. Lincke, M. Höst, and P. Runeson, "How do phd students plan and follow-up their work?–a case study," *School of Mathematics and Systems Engineering, University Sweden*, 2007.

[20] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in software engineering*. Springer Science & Business Media, 2012.

[21] R. Yin, "Case study research design and methods 3rd ed sage publications," 2002.