CONSISTENCY IN USE CASE TRANSACTION IDENTIFICATION METHODS

by

Tanveer Ikram | Visiting Faculty | ikram712000@yahoo.com | BITS Pilani, Rajasthan, India

ABSTRACT

Use case transactions are used in Use Case Point (UCP) method beside other functional size measurement methods. It is the smallest unit in functional size measurement while measuring the size of the system. For UCP many methods of use case transaction identification have been proposed. There can be a problem of consistency of functional size measurement due to various approaches of use case transaction identification and analysis of requirements using natural English language.

The aim of this study is to investigate consistency in identifying use case transaction of some popular use case identification methods used in industry.

A controlled experiment on a group of 27 students was performed to investigate whether the methods for transaction identification, known from the literature, provide similar results. In addition, a qualitative analysis of the experiment data was performed to investigate the potential problems related to transaction identification in use cases.

A significant difference in the median number of transactions was observed between groups using different methods of transaction identification. The ANOVA test was performed with the significance level set to 0.05. Although a large intra-method variability was not observed. The ratios between the maximum and minimum number of transactions identified by the participants using the same method were from 1.18 to 1.28.

There seems to be no difference between Gustav Karner and Ivar Jacobson methods of finding transactions while there is a considerable difference between these two and Robiolo Orosco. Difference in the values of transactions identification methods can have a visible impact on the values of the use-case-based Functional Size Measurement (FSM).

KEYWORDS: Transactions, Use Case, UCP, Ivar Jacobson, Karner, Robiolo Orosco.

INTRODUCTION

In the context of software development, the primary cost driver is effort [1]. Consequently, the total costs of a development project largely depend on the invested working time. Before estimating effort, however, the size of software must be determined. As stated above, software is intangible, and, consequently, it has no physical dimensions that describe its size. Therefore,
software engineering must provide techniques for software sizing which represents a pre-stage for effort estimation techniques. The most prominent size measures are Lines of Code as well as Function Points and its variants.

The Function Point (FP) estimation technique or Function Point Analysis (FPA) determines the software size by quantifying its information processing functionality [2]. This quantification approach was initially published by ALLAN ALBRECHT in 1979 [3]. Function Points are especially helpful for early estimations since the approach works with premature, vague information [2]. FPA was primarily designed for the domain of Management Information Systems (MIS), where it has been extensively analyzed and evaluated [4]. Since 1986, FPA is being maintained by the International Function Point Users Group (IFPUG), which provides certifications and estimation guidelines for software professionals [5]. FPA is still the most widely used formal estimation technique today [6]. The concept of “Use Cases” has been popularized by the Unified Modeling Language (UML) [7]. Before UML has been developed in the 1990s, IVAR JACOBSON, who was one of the contributors of UML, developed the technique of specifying Use Cases [8]. According to JONES, “A Use Case describes what happens to a software system when an actor (typically a user) sends a message or causes a software system to take some action [9].”

Use Case Points is probably the most recognizable and discussed use-case-based method for effort estimation [10]. The procedure of counting UCP proposed by Karner [11].

The notion of use-case transaction is used in UCP to calculate Unadjusted Use Case Weights (UUCW). Transactions are also a basis for other use-case-based Functional Size Measurement (FSM) methods, i.e., Transactions [12] and TTPoints [13].

The use-case transaction was defined, for the first time, by the inventor of use cases – Ivar Jacobson. He stated that each use-case transaction should consist of four types of actions [8]:

- actor’s request – the main actor sends request and data to the system;
- system validation – the system validates the given data;
- system internal-state change – the system performs operations leading to change of its internal-state;
- system response – the system responds to the actor with the operation result.

Therefore, the Jacobson’s use-case transaction can be understood as a “round trip”, where actor stimulates the system, and the system provides a response to that stimulus. This kind of operation is atomic from the actor’s point of view. The initial idea was that each step should constitute a transaction.

Another approach to transactions identification was proposed by Robiolo and Orosco [12]. They suggested counting the number of stimuli as the number of transactions. In that case, the actor is the subject of the sentence and the stimulus triggered by the actor is the verb.
The concept of use-case transaction was applied to FSM and effort estimation by Karner in his UCP method [11]. The definition of transaction used in UCP, states that use-case transaction is “a set of activities, which is either performed entirely, or not at all.”

In order to use transactions to measure the functional size of a system effectively, there has to be a standard procedure for their identification. In the case of FPA great efforts have been made to develop standards and manuals which cover also transaction identification process [15,16,17]. A much bigger problem could be observed for the UCP method, for which many different approaches to transactions identification have been proposed [18, 14,11, 12]

Keeping in mind different definitions of transactions there can be a problem of consistency of functional size measurement due to various approaches of use case transaction identification and analysis of requirements using natural English language.

The aim of this study is to investigate consistency in identifying use case transaction of above mentioned use case identification methods used in industry.

This paper is divided in five sections. In Section 2, the UCP method, a use-case-based method for effort estimation is discussed. Section 3 discussed Use Case Transaction methods. The experiment in which three methods for transaction identification were compared is presented in Section 4. In Section 5 finding of experiment is discussed.

USE CASE POINTS

The concept of “Use Cases” has been popularized by the Unified Modeling Language (UML). [7]. Before UML has been developed in the 1990s, IVAR JACOBSON, who was one of the contributors of UML, developed the technique of specifying Use Cases [8]. According to JONES,

“A Use Case describes what happens to a software system when an actor (typically a user) sends a message or causes a software system to take some action.” [9]

Use Case Points is probably the most recognizable and discussed use-case-based method for effort estimation [13]. The procedure of counting UCP proposed by Karner [11] consists of the following steps:

1. Calculate Unadjusted Actors Weights (UAW). Assign each of the actors to one of three categories based on the type of interface it uses to communicate with the system: simple – API, average – protocol or terminal, complex – GUI. For each category count the total number of assigned actors, and multiply it by the weight assigned to the category (simple = 1, average = 2, and complex = 3). Calculate UAW as a sum of products of number of actors assigned to categories and their weights.

2. Calculate Unadjusted Use Case Weights (UUCW). Count the number of transactions in each use case. Assign use cases into one of three categories based on the number of transactions: simple (T < 4), average (4 ≥T ≥ 7), and complex (T > 7). Calculate UUCW as a sum of products of the number of use cases assigned to categories and their weights (simple = 5, average = 10, and complex = 15).
3. Assess Technical Complexity Factors (TCF). Evaluate the influence of each Technical Complexity Factor, presented in Table 1, by assigning value between 0 and 5. Calculate TF_Prod as a sum of products of the factors weights and their influence. Calculate TCF according to Eq. (1).

\[
TCF = 0.6 + (0.01 \times TF_{Prod})
\] ...1

4. Assess Environmental Factors (EF). Evaluate the influence of each environmental factor, presented in Table 1, by assigning value between 0 and 5. Calculate EF_Prod as a sum of products of the factors weights and their influence. Calculate EF according to Eq. (2).

\[
EF = 1.4 + (-0.03 \times EF_{Prod})
\] ...2

<table>
<thead>
<tr>
<th>Technical Complexity Factors and environmental factors.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>T1</td>
</tr>
<tr>
<td>T2</td>
</tr>
<tr>
<td>T3</td>
</tr>
<tr>
<td>T4</td>
</tr>
<tr>
<td>T5</td>
</tr>
<tr>
<td>T6</td>
</tr>
<tr>
<td>T7</td>
</tr>
<tr>
<td>T8</td>
</tr>
<tr>
<td>T9</td>
</tr>
<tr>
<td>T10</td>
</tr>
<tr>
<td>T11</td>
</tr>
<tr>
<td>T12</td>
</tr>
<tr>
<td>T13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Environmental Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
</tr>
<tr>
<td>F2</td>
</tr>
<tr>
<td>F3</td>
</tr>
<tr>
<td>F4</td>
</tr>
<tr>
<td>F5</td>
</tr>
<tr>
<td>F6</td>
</tr>
<tr>
<td>F7</td>
</tr>
<tr>
<td>F8</td>
</tr>
</tbody>
</table>

Table 1: Technical Complexity Factors and environmental factors [11]

5. Calculate Unadjusted Use Case Points (UUCP) according to Eq. (3).

\[
UUCP = UAW + UUCW
\] ...3

6. Calculate Use Case Points (UCP) according to Eq. (4).

\[
UCP = UUCP \times TCF \times EF
\] ...4

7. To obtain the effort estimation measured in man-hours, one has to multiply UCP by Productivity Factor (PF). The default value for PF proposed by Karner was 20 h per UCP. Schneider and Winters [19] proposed rule for choosing PF. They suggested to count the
number of Environmental Factors F1–F6 which influence is predicted to be less than 3, and factors F7–F8 which influence is predicted to be greater than 3. If the counted total is equal to two or less, 20 [h/UCP] should be used; if the total is between 3 and 4, PF is set to 28 [h/UCP]; if the calculated number is greater than 4, the highest value of 36 [h/UCP] should be used.

EXPERIMENTAL COMPARISON

3.1 Experiment Design

The independent variable considered in the experiment was a method used to identify use-case transactions (three values considered). The dependent variable was a number of transactions identified in a use-case-based requirements specification.

Transaction Identification Methods. Three methods for transactions identification where investigated during the experiment:

M 1– identifying transactions based on the definition provided in Karner’s UCP [11];
M 2– counting stimuli-verbs (Robiolo and Orosco) [12];
M 3– identifying transactions based on Jacobson’s “round trip” transactions;

Software Requirements Specification. Number of identified transactions can depend both on the structure of use-cases and author’s writing style. To mitigate the risk of using a specific requirements specification (i.e. too easy, or too difficult to analyze), we decided to use 38 use cases taken randomly from 10 different projects from industry. It is an instance of a typical use-case-based requirements specification, derived based on the analysis of 321 use cases coming from 10 following projects.

Participants. Participants in the experiment were 27, 2nd year Masters Students, who were in the middle of the third semester of the four semester SE course. Each participant was having 2 years industry experience in IT field. Participants were familiar with the concept of use cases as they authored and reviewed use cases during that course.

Participants were randomly assigned to three groups: G1 that was asked to use the method M1; group G2 used the method M2; group G3 used the method M3.

3.2 Operation of the Experiment

The experiment was executed at a company on Saturday when students were having part time classes of their Masters course in software engineering in the month of May 2015.

Prepared Instrumentation: Each participant in the experiment was provided with the handouts that contained a description of the Use Case method, instructions, and tables for collecting the number of transactions for each use case.
Participants also had access to the presentation with the definition of the transactions identification method appropriate for their group (together with examples of its usage). During the execution of the experiment participants used only pen and paper to identify number of transaction in each use case. No participant has done this task before.

**Execution.** During the execution of the experiment students were supervised by lecturers, who were obliged to answer questions concerning the Use Case method, excluding those referring to the transaction identification process. Participants had 120 minutes to acquaint themselves with the transaction identification method appropriate for their group, and identify transactions in 38 use cases.

**ANALYSIS AND INTERPRETATION**

**Data verification.** After collecting and reviewing participants’ forms none of them was rejected and it was observed that students participated seriously in the experiment.

### 4.1 Quantitative Analysis

**Central tendency.** Methods M1, M2, and M3 could be treated as measurement instruments for counting transactions in use cases. By investigating central tendencies of the experiment samples, one can investigate whether the methods on-average provide similar results.

**ANOVA-Comparing Multiple Samples of Numerical Data: One Factor**

We are here to compare means of multiple samples (beyond two) to determine if the means of the populations from which they were drawn are equal or not. Suppose we were interested in comparing means of three samples to determine if the means of the populations from which they were drawn are equal or not. The hypotheses statements are:

- **H₀:** Meanₐ = Meanₐ₂ = Meanₐ₃ or we can say that all methods are same.
  
- (H₀ : θₐ = θₐ₂ = θₐ₃)

- **H₁:** At least two Means are not equal or we can say that all methods are not same.
  
- (H₁ : Not H₀)

The parametric technique we use to conduct the analysis is called Analysis of Variance (ANOVA). The title Analysis of Variance may imply that we are going to compare variances, and not means. Actually we are going to compare means, but the way we do it is to compare the difference or variation **between the means** of the three groups to the variation **within** the groups. If the variation between the means is greater than the variation within the groups, we go with the alternative hypothesis.

The important thing to remember is that the between group variation simply measures how the sample means compare to the grand average of all of the data. Does at least one group have a sample mean much greater or much less than the grand average of all the data? If "yes"
(meaning the difference between the means is significantly greater than the within group variation), then we reject the null hypothesis and conclude at least two of the means are not equal. The within group variation I am referring to measures how the individual observations within a group differ from their group sample mean.

When the sample means are far from each other relative to the within group variation, "something" is going on. In the three methods scenario, the standard methods really have different ways that shifts the average number of transactions of one or more groups away from other another group or groups. Simply put, the average number of transaction is significantly different between the groups. When the sample means are close together relative to the within group variation, we say that any difference between sample means is due to chance - nothing "special" is going on to cause the means to be different.

We are going to use the ANOVA table, comparing between group variation to within group variation.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Count</th>
<th>Min</th>
<th>Max</th>
<th>Ratio of min and max</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>9</td>
<td>123</td>
<td>149</td>
<td>1.2</td>
<td>1232</td>
<td>136.8889</td>
<td>106.3611</td>
</tr>
<tr>
<td>Group B</td>
<td>9</td>
<td>151</td>
<td>179</td>
<td>1.18</td>
<td>1497</td>
<td>166.3333</td>
<td>77.6</td>
</tr>
<tr>
<td>Group C</td>
<td>9</td>
<td>125</td>
<td>151</td>
<td>1.28</td>
<td>1285</td>
<td>142.7778</td>
<td>55.4444</td>
</tr>
</tbody>
</table>

Table 2: Number of Transactions identified by participants of the experiment.

**The-Situation**

There were three groups of students and each group was given one method to identify number of transaction in 38 use cases taken from various projects from industry. Data is shown in table 2.
The objective is to compare the mean number of transactions of group A, with group B with group C to determine if there is or is not a difference between the means. The hypotheses statements are presented in the introduction above. We are going to use the F statistic because we are comparing the ratio of two variances (between group variation to within group variation).

The tool we use Excel and most of the work is done through manual calculations. My data is in columns, and I defaulted to an alpha of 0.05. **Single Factor** means we are studying one factor that may result in the means being different - that factor is the type of method used in finding use case transactions given to 9 students in each of the three groups. The results (output) of the ANOVA analysis are shown below:

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>F crit (2.24)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>4369.55</td>
<td>3-1=2</td>
<td>4369.775/2=2184.775</td>
<td>2184.775 /79.272=27.403</td>
<td>3.4028</td>
</tr>
<tr>
<td>Within Groups</td>
<td>1913.45</td>
<td>27-3=24</td>
<td>1913.45 / 24 = 79.727</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>6283.00</td>
<td>27-1=26</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 3: ANOVA Analysis Result**

The first thing to note in the ANOVA output are the descriptive statistics which include the means and the variances of the three samples. These are followed by the ANOVA Table which has the variation separated as Between Groups (variation between the three means and the grand mean) and Within Groups. Since the F crit (3.4028) is less than the F calculated value of 27.403, we reject the null hypothesis and conclude that there is a difference between the means. Thus, "something" special is going on - the methods do have a significant differential effect on the number of transactions.

Therefore, we concluded that there is a significant difference between the median numbers of transactions of at least two groups.

However in this particular case, the interpretation seems fairly straightforward: There seems to be no difference between G1 (Gustav Karner ) and G3 (Ivar Jacobson) methods of finding transactions while there is a considerable difference between G1 (Gustav Karner ) and G2 (Robiolo and Orosco) & also between G3 (Ivar Jacobson) and G2 (Robiolo and Orosco).

**Variability.** Another important aspect is the inner variability in the results obtained by participants using the same method. The lesser the variability observed for a group, the more reliable is the procedure for transaction identification. The observed inner variability in the number of transactions for all groups was low. The ratios between the maximum and minimum number of transactions were equal to 1.2, 1.18, and 1.28.
4.2 Qualitative Analysis

Data collected allowed us to make some observations concerning the usage of methods for transactions identification. The biggest problem for participants, who used the method M2, was a choice of verbs that are stimuli. It appeared that different people could choose different verbs as stimuli. While verbs which describe typical interaction with the system (e.g., “submit”, “select”, or “enter”) were rather equally marked as stimuli, verbs describing more general activities (e.g., “find”, “browse”, “go”) were misleading, and often lead to different interpretations.

Two methods M1 and M3 showed approximate same understanding and gave almost same results and seem to be less confusing as compared to M2.

Threats to validity

In case of threats to internal validity the potential problem could be a difference in the student’s level of knowledge regarding use cases. Study has not identified level of knowledge of participants having previous experience with Use Case Transaction Method.

Another factor that could influence the outcomes of the experiment was the quality of the instrumentation provided to the participants. To mitigate that risk we provided definitions, and examples as they were published in the original papers. (If there was insufficient number of examples available for a method in the original papers, additional examples were prepared by authors of this study.) However, in case of the experiment regarding methods for transactions identification, it was important that participants had no previous experience in that area. (Previous experience could interfere with the provided transaction definitions.)

On the other hand, in case of study by Mirosław Ochodek case study [22] with six experienced estimators (who were not forced to use any specific definition of transaction) the difference in maximum and minimum numbers of identified transactions was also high (a factor of 2.1).

CONCLUSION

There are many definitions of use case transactions. In this study we tried to find the cause – effect relationship between independent variable considered in the experiment was a method used to identify use-case transactions (three values considered) and the dependent variable was a number of transactions identified in a use-case-based requirements specification.

Experiment showed that there is a considerable difference between the median of transaction identified by round trip method by Ivar Jacobson method and Robiolo Orosco method of counting the number of stimuli.
Experiment also found that median of transaction identified by Gustav Karner is almost same as transaction identified by Ivar Jacobson method.

The notion of use-case transaction is vague, as there are many different definitions of this concept. The first one comes from the Ivar Jacobson – the inventor of use cases. It states that use-case transaction consists of actor’s request and the system response. The second one, introduced by Gustav Karner, which states that transaction is a set of activities in use-case-scenarios (that is meaningful to the actor), which is either performed entirely, or not at all, and leaves the business of application in consistent state. In addition, third definition, Robiolo Orosco suggested to count the number of stimuli as the number of transactions.

Results confirms a similar study by Miroslaw Ochodek [13] but a high intra-method variability was not observed as it was observed in experiment by Miroslaw Ochodek. A high intra-method variability could be a threat to the reliability of use-case based FSM methods (the observed ratios between maximum and minimum number of identified transactions by participants using the same method, ranged from 1.96 to 3.83 in case of Miroslaw Ochodek while in our case it was only 1.18 to 1.28). However, one should keep in mind that the participants were students who had no previous experience in counting transactions in use cases in case of Miroslaw Ochodek while in our case all students were having 2 years industry experience.

The final conclusion is that the method chosen for transaction identification can have a visible impact on the values of the use-case-based FSM. There is a need to experiment that beside method, better understanding of method, experience of staff, automation effect transaction identification.

REFERENCES