Journal of Multidisciplinary Engineering Science and Technology (JMEST)
ISSN: 2458-9403
Vol. 5 Issue 12, December - 2018

Influence of Team Roles in Software Measurement: A Replicated Experiment

Aguilar, R.A.[1], Díaz, J.C.[1] & Ucán, J.P.[1]
[1] Autonomous University of Yucatan, Faculty of Mathematics
Mérida, México
{avera, julio.diaz, juan.ucan}@correo.uady.mx

***Abstract*—This article presents a replicated experiment in which the convenience of using the Belbin Role Theory for the integration of work teams with the task of measuring the software is studied. The study was developed in an academic environment with first-year students of the software engineering career and analyzes the differences between the metrics obtained with the Function Point Analysis Technique, by teams integrated with the Belbin Theory, and those obtained by teams integrated with students selected randomly.**

The results regarding the Function Points variable showed significant differences, its result contrasts with that obtained in the previous study, in which the general analysis of the FP did not show statistically significant differences. On the other hand, regarding the metric associated with the process, in this case, the time that the teams used to carry out the task, the results also contrast with the first experiment, since no statistically significant differences were found between the two types of team conformation (Belbin Teams and Traditional Teams).

The contrasting results in both experiments motivate us to continue the comparative analysis between the treatments for the factor related to the conformation of work teams.

Keywords—Belbin Roles; Experimentation in Software Engineering; Function Point Analysis; Software Measurement.

I. INTRODUCTION

Although the body of knowledge [1] of the Software Engineering (SE) began to be integrated at the late 60s, it was not until the 80s that the academic community began to adopt and use research approaches to study more rigorously the different aspects and problems involved in the software development process [2].

One of the most commonly used methodologies in the field of empirical software engineering, is the experimentation, particularly, experimentation in controlled environments. The emphasis on applying experimentation to SE dates back to the 80s [3] and generally aims to identify the causes by which certain results are produced, its application helps us to identify and understand the possible relationships between factors involved in software development processes.

A key element in the experimental paradigm is the development of replicas to previously performed experiments. The purpose of this task is to verify previously observed findings [4].

In this article we report a replication to a previous experiment [5] carried out by the same research group in which the conditions of the previous experiment are maintained, being the only variant in the new experiment, the maturity of the participants with whom the teams were formed of work —experimental subjects.

II. RELATED WORK

Some researchers [6-9] claim to have identified roles that describe the behavior of individuals in work teams and although there is no evidence that they are associated with any type of activity, particularly associated to the software development, their absence or presence is says that it has significant influence on the work and the achievements of the team [10].

In a previous study [5] the convenience of using the Belbin Role Theory for the integration of work teams for the task of measuring the software was explored with a controlled experiment. The study was development in an academic environment with students, and we analyzed the differences between the metrics obtained with the Function Point Technique, by teams integrated with the Belbin Theory, and those obtained by teams integrated with students selected randomly. The results obtained provided evidence regarding the significant differences in terms of the time spent for the task; we observed that teams integrated with the Belbin theory take more time. Regarding the five metrics obtained to measure the functionality of the software, we found differences only in the functionality linked to the external outputs.

III. METHODOLOGY

An experimental replication was developed from a previous study with the intention of deepening the analysis of the influence of the Belbin Role Theory in the integration of software development teams; its influence was evaluated using the result —product— of applying the technique “Function Points Analysis” in a Software Requirements Specification [11].

Additionally, the time used by the teams to carry out the task was used as a second dependent variable —a metric process.

A. Objective, Hypothesis and Variables.

With the aim of study whether the integrated work teams based on the Belbin Role Theory —which we will refer to as Belbin Teams (BT)— generate measurements significantly different from those obtained by the randomly integrated teams —which we will call Traditional Teams (TT)— two pairs of hypotheses were proposed, such as the following:

H01: The average of the FP obtained with the Function Point Analysis Technique by the TT is equal to the average of the ILF obtained by the BT.

H11: The average of the FP obtained with the Function Point Analysis Technique by TT differs from the average of the FP obtained by the BT.

Likewise, with the objective of identifying whether the time spent by the integrated software development teams based on the Belbin Role Theory, in the aforementioned software measurement task, differs from the time that the integrated teams invest randomly, we proposed the sixth hypotheses:

H02: The average time recorded by the TT in the measurement task is equal to that reported by the BT.

H12: The average time recorded by TT in the measurement task differs from that reported by the BT.

The factor to be controlled is the integration mechanism of the software development teams, which has two alternatives: (1) Belbin Teams (BT), and (2) Traditional Teams (TT). On the other hand, response variables (Function Points and Time) obtained by applying the technique of Function Points Analysis in the ERS are recorded in an instrument that the work teams deliver at the end of the activity.

Aspects such as the complexity of the problem to be solved, the time available for the task, the instruction received, are considered parameters that do not affect or skew the results of the study, because they are homogeneous parameters for all the teams of development. As for the replication, these aspects were kept the same by the research group.

In relation to the expertise of the participants regarding the measurement technique used, we can affirm that it remained homogeneous since the subjects had not yet studied it. As for the reply, this aspect had a slight variation that is worth commenting; in the first study, third-year students participated, while in the present study, the teams were composed of first-year students.

B. Participants/Subjects

The participants in the experiment were seventeen students of a Bachelor in Software Engineering from the Autonomous University of Yucatan, who were studying the course “Fundamentals of Software Engineering” during the summer of 2018, subject located in the first year. With the students enrolled in the course, five software development teams of three members each were formed —a sixth team consisting of only two students was not considered for the analysis. We used the information obtained —primary roles in students— after the administration of the self-perception inventory of Belbin, and we integrate three teams with compatible roles (Belbin Teams: BT) and other two with students assigned in a random way (Traditional Teams: TT). Given that the measurements would be obtained on the products generated by the development teams, the experimental subjects in this case were the five work teams integrated by the researchers.

The conformation of the three teams (with three members) based on the Belbin Role Theory are illustrated in the figure 1.

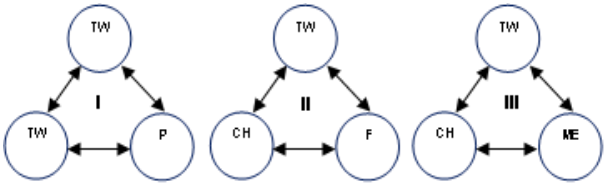

Fig. 1. *Integrated teams with Belbin Theory (BT).*

C. Experimental Design

A Factorial Design with "one variation at a time" was used, the independent variable corresponds to the way of integrating the work teams, in the study there are two experimental treatment factor levels: BT and TT. Table 1 illustrates the assignment of the five teams to each of the treatments.

TABLE I. EXPERIMENTAL SUBJECTS BY TREATMENT

Treatment	Teams
Belbin Teams (BT)	I, II, III
Traditional Teams (TT)	IV, V

D. Execution of the Study

The experimental replica was carried out in three work sessions and it was executed in similar conditions to those of the first experiment. In the first one, the self-perception study was administered to the students, this session was carried out in the last half hour of a class session of the subject; subsequently, in a second session that lasted two hours, participants received instruction on the technique function point analysis; finally, in the third session, the experiment was executed.

In the experimental session, the teams were integrated —according to the experimental design— and they were provided with the ERS of a case study, as well as a report sheet; later, a brief description of the activity was provided and they were asked to identify the team, as well as record the start and end time of the activity.

IV. Analysis and Results

This section presents both the descriptive statistical analysis of the measurements collected and the inferential statistical analysis.

A. Descriptive Analysis

For the descriptive analysis, table II presents the main statistics for FP variable and the boxplot illustrates its behavior (see fig. 2).

TABLE II. Statistical Summary for FP variable

Treatment	f	$\bar{x}$	s
BT	3	91.7	5.1
TT	2	177.5	0.7

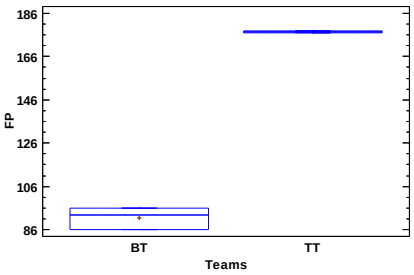

Fig. 2. *Boxplot for dependent variable FP*

We can observe that the treatments do not present overlapping, which induces us to think that they are different; also we can observe that the BT teams presents higher variability than the TT and that the BT teams has a lower average.

On the other hand, table III presents the statistics for Time variable and the figure 3 presents the boxplot.

TABLE III. Statistical Summary for Time variable

Treatment	f	$\bar{x}$	s
BT	3	111.3	23.2
TT	2	116.5	22.6

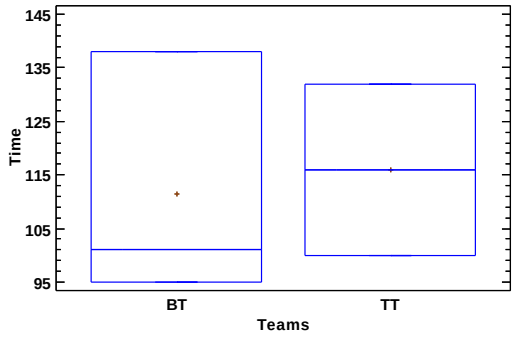

Fig. 3. *Boxplot for dependent variable Time*

We can see that the treatments overlap so it is very likely that these treatments do not show statistically significant differences; also we can observe that the BT teams presents slightly higher variability than the TT.

B. Inference Analysis

In order to evaluate the differences observed in the FP variable, discard the one related to the Time variable, and determine if they are significant from the statistical perspective, the following statistical hypotheses were raised:

$$H0_{FP}: \mu_{BT} = \mu_{TT} \quad ; \quad H1_{FP}: \mu_{BT} <> \mu_{TT} \qquad (I)$$
$$H0_{Time}: \mu_{BT} = \mu_{TT} \quad ; \quad H1_{Time}: \mu_{BT} <> \mu_{TT} \qquad (II)$$

We proceeded to use the one-way ANOVA; the associated linear statistical model is the following:

$$Y_{ij} = \mu + \beta_i + \varepsilon_{ij} \qquad (III)$$

where Y_{ij} is the ij-th observation (value of the j-th replica under treatment i), μ is a parameter common to all treatments called general or global mean, βi is a parameter associated with the i-th treatment called effect of the i-th treatment and ε_{ij} is the random component of the error. Tables IV and V present the results of these analyzes.

TABLE IV. ANOVA table for FP

Source	SS	Df	MS	F	P-Value
Between groups	8840.83	1	8840.83	498.8	0.0002
Within groups	53.1667	3	17.7222		
Total (Corr.)	8894.0	4			

TABLE V. ANOVA table for Time

Source	SS	Df	MS	F	P-Value
Between groups	26.133	1	26.1333	0.05	0.8389
Within groups	1596.67	3	532.222		
Total (Corr.)	1622.8	4			

The ANOVA table decomposes the variance of the variable under study into two components: an inter-group component and an in-group component.

Since the P-value of the F-test is less than 0.05 for the FP variable, there is a statistically significant difference between the average of FP between one level of treatments and another. Therefore, we can reject the null hypothesis and affirm the average of the FP obtained with the Function Point Analysis Technique by TT differs from the average of the FP obtained by the BT with a 5% level of statistical significance. In the case of the variable Time, the P-value of the F-test is bigger than 0.05, we can´t refuse the null hypothesis.

$$H1_{FP}: \mu_{BT} <> \mu_{TT} \quad \& \quad H0_{Time}: \mu_{BT} = \mu_{TT}$$

In order to deepen the analysis of the differences between the components of the variable FP, we analyze the variance for each of the five metrics that integrate — EI, EO, EQ, ILF, EIF. In table VI we can identify that the metric ILF represents a significant difference with a p-value of 0.0049, likewise the component EI is in the rejection threshold.

TABLE VI. ANOVA SUMMARY FOR EACH METRIC

Metric	F	P-Value
EO	3.90	0.4128
EI	9,82	0.0519
EQ	3.37	0.1636
ILF	56.59	0.0049
EIF	3.75	0.1482

The ANOVA Model has associated three assumptions that it is necessary to validate before using the information it offers us; the assumptions of the model are: (1) The experimental errors of your data are normally distributed, (2) Equal variances between treatments (Homoscedasticity) and (3) Independence of samples.

To validate the first assumption, we will use the normal probability graph of residuals. It is a graphical technique for assessing whether or not a data residuals set is approximately normally distributed.

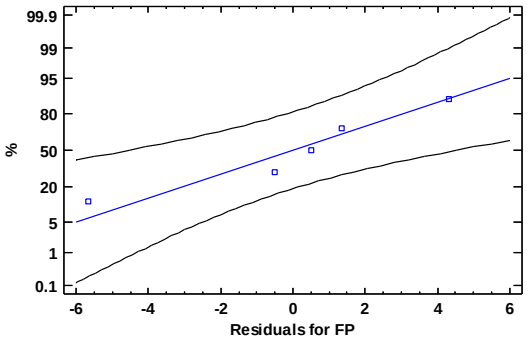

Fig. 4. *Normal probability plot for FP variable*

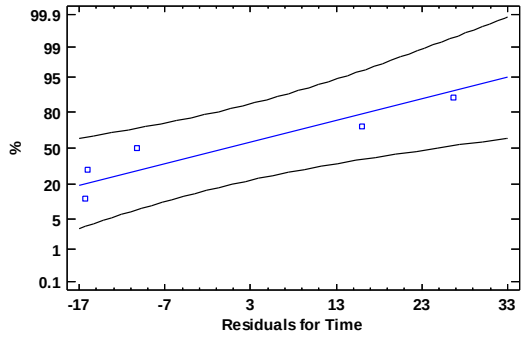

Fig. 5. *Normal probability plot for Time variable*

As can be seen in the graph of Figures 4 and 5, the points, in both graphs, do not show deviations from the diagonal, so it is possible to assume that the residuals have a normal distribution in both cases.

For the assumption about the equal variances between treatments —traditionally known as Homoscedasticity— we apply the levene's hypothesis test. As we can see in table VII, in both cases the p value is greater than 0.5 so it is not possible to reject the hypothesis of nullity that indicates that the means of the variances are equal, that is, in both cases the second assumption of the model is met.

TABLE VII. STATISTICAL SUMMARY FOR TIME VARIABLE

Variable	Levene's Test	P-Value
FP	3.92594	0.1419
Time	0.0833424	0.7916

Finally, to validate the assumption of data independence, we generate a residuals vs. Order data plot. In this case, we observe if it is possible to detect any tendency to have gusts with positive and negative residuals.

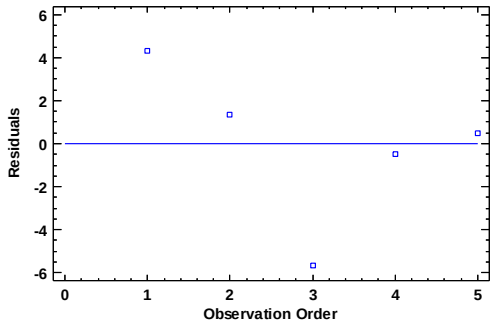

Fig. 6. *Residuals vs. Order data Plot for FP*

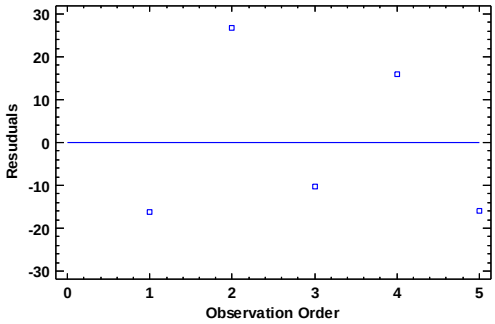

Fig. 7. *Residuals Vs. Order data Plot for Time*

In the case of our analysis, we can see in figures 6 and 7, that in both cases no trend are identified, so it is possible to assume that the data come from independent populations.

V. Conclusions And Future Work

In this paper, we present an experiment that aims to contrast the results obtained in a previous experiment. In both controlled experiments, we compare the performance —metrics linked to the product and the process— of work teams in tasks related to software measurement using the Function Point Analysis Technique. The treatments to be compared, were linked with the way of integrating the work teams; firstly, the proposal to integrate teams using the Belbin Role Theory, and on the other, the traditional way of randomly assigning its members. According to [12] the experimental replica is of similar internal type, with a slight variant with respect to the maturity of the students who integrated the experimental teams.

The results regarding the process metrics showed significant differences in the metric FP, this result contrasts with that obtained in the previous study, in which the general analysis of the FP did not show statistically significant differences. It is worth mentioning that in the first experiment only the EO component —one of the five metrics that additively make up the FP metric— presented statistically significant differences, while in the second, the ILF and EI metrics show such differences. We also observed a contrast with the real value of the case study, the TT team had on average a value closer to the real value.

On the other hand, regarding the metric associated with the process, in this case, the time that the teams used to carry out the task, the result also contrasts with that obtained in the first experiment.

The contrasting results in both experiments motivate to the research group to continue the comparative analysis between the treatments for the factor related to the conformation of work teams. The alternatives identified by the research group are: 1) the realization of future experimental replicas, 2) perform a new statistical analysis that integrates both datasets using a blocking factor —the group.

Acknowledgment

Authors from the Autonomous University of Yucatan (UADY) acknowledge the financial support received from the Mexican Secretary of Public Education (SEP) through project P/PFC-2016-31MSU0098J-13. We also acknowledge the support provided by UADY students participating in this study.

References

[1] Bourque, P & Fairley, R.: Guide to the Software Engineering Body of Knowledge (SWEBOK V3.0). IEEE Computer Society, 2014.

[2] Glass, R., Vessey, I., Ramesh, V. Research in software engineering: an analysis of the literature. Information and Software Technology, 44 (8), 2002, pp. 491 – 506.

[3] Basili, V., Selby, R. y Hutchens, D. Experimentation in software engineering. IEEE Trans. Softw. Eng., 12 (7), 1986, pp. 733–743.

[4] Gómez, O., Juristo, N., & Vegas, S. Understanding replication of experiments in software engineering: A Classification, *Inf. Softw. Technol.*, vol. 56, no. 8, 2014, pp. 1033-1048, 2014.

[5] Aguilar, R., Díaz, J., Ucán, J. & Quiñones, Y. Exploring the Influence of Belbin´s Roles in Software Measurement: A controled Experiment with Students. In J. Mejia et al (Eds.) Advances in Intelligent Systems and Computing, vol 865. Springer, Cham, 2019, pp. 69-79.

[6] Humphrey, W.: Introduction to the Team Software Procesess. Reading, USA: Addison Wesley Longman Inc., 2000.

[7] Belbin, M.: Team roles at Work. Oxford, USA: Elsevier Butterworth Heinemann, 1993.

[8] Mumma, F.S.: Team-work & team-roles: what makes your team tick? King of Prusia, PA: HRDQ, 2005.

[9] Margerison, C.J. & McCann, D.J.: Team Management Profiles: Their use in Managerial Development, Journal of Management Development, vol. 4 no. 2, 1985, pp. 34-37.

[10] Senior, B.: Team roles and Team performance: Is there ‘really’ a link?, Journal of Occupational and Organizational Psychology, 70, 1997, pp. 85-94.

[11] Garmus, A. & Herron, D.: Measuring the Software Process. Prentice Hall, 1996.

[12] Almqvist, J. Replication of Controlled Experiments in Empirical Software Engineering– A Survey. Final Project, Department of Computer Science, Faculty of Science, Lund University, 2006.