A Replicated Study on the Evaluation of a Size Measurement Procedure for Web Applications[1]

Silvia Abrahão[1], Geert Poels[2], and Emilio Insfran[1]

[1]*ISSI Research Group, Department of Information Systems and Computation*
Universidad Politécnica de Valencia
Camino de Vera, s/n, 46022, Valencia, Spain
{sabrahao,einsfran}@dsic.upv.es

[2]*Faculty of Economics and Business Administration*
Ghent University
Tweekerkenstraat 2, 9000 Ghent, Belgium
Geert.Poels@UGent.be

Abstract

This paper presents a replication study that investigates the efficacy and likely adoption of a measurement procedure for sizing Web applications from conceptual models (OOmFPWeb). The goal of the replication was to provide evidence for the generalization of the results by repeating the experiment in a different environment, using different subjects. The results of the replica carried out in Austria have confirmed the results of the original experiment, which was carried out in Spain. OOmFPWeb is efficient when compared to current industry practices. It provides reproducible functional size measurements and is perceived as easy to use and useful by its users, who also expressed their intention to use OOmFPWeb in the future. The analysis further supports the validity and reliability of Moody's Method Evaluation Model for evaluating functional size measurement methods.

Keywords: *Empirical Web Engineering, Functional Size Measurement, Method Evaluation.*

1. Introduction

The rapid emergence of Web information systems in the last years presents a serious challenge for the skills of software project managers. There is a need for reliable approaches to help managers to deliver Web projects on time and within budget. Some approaches for sizing Web sites and applications have been proposed [5] [6] [7] [10] [18]. The main limitation of these approaches is that they cannot be used early in the Web development lifecycle as they rely on implementation decisions. In addition, little systematic evaluation of the proposed FSM alternatives for Web applications has been documented [5] [9] [16].

In previous work [1] [2], we developed a functional size measurement procedure for Web applications (OOmFPWeb). This procedure is intended to be used within the context of a model-driven development method for Web applications called OOWS (Object-Oriented Web Solutions) [17]. The procedure was designed to conform to the IFPUG (International Function Point Users Group) counting rules for FPA [11]. It redefines these rules in terms of the concepts used in OOWS, in order to enable and facilitate the application of this widely accepted and ISO-standard functional size measurement (FSM) method.

As OOmFPWeb is a new procedure, we decided to evaluate it in an artificial 'laboratory' setting using student participants rather than employing field studies with practitioners. The evaluation of OOmFPWeb [1] was based on an evaluation model that was obtained by operationalizing the Method Evaluation Model (MEM) [15]. However, as stated by Basili et al. [3] without confirming the results by replication studies, results in experimental software engineering should only provisionally be accepted.

Therefore, this paper presents a replication study that investigates the efficacy and likely adoption of OOmFPWeb for sizing Web applications from conceptual models. The goal of the replication was to provide evidence for the generalization of the results by repeating the experiment in a different environment.

This paper is organized as follows. Section 2 describes the model used to evaluate OOmFPWeb. Section 3 describes the design of the replica as well as the results of the original experiment. Section 4 describes the results of the replica and the limitations of the study. Finally, Section 5 presents our conclusions and future works.

[1] This work is supported by the Spanish Ministry of Science and Technology under the META Project (TIN2006-15175-C05-01).

978-0-7695-3261-5/08 $25.00 © 2008 IEEE
DOI 10.1109/ICWE.2008.45

2. Evaluation Model

The Method Evaluation Model (MEM) [15] provides a suitable basis for a multi-dimensional quality model of FSM measurement procedures (Figure 1). This model was originally proposed as an evaluation model for IS design methods. The main contribution of the MEM is that it incorporates two aspects of method "success": actual efficacy and perceived efficacy. Both aspects must be considered when evaluating FSM methods or procedures.

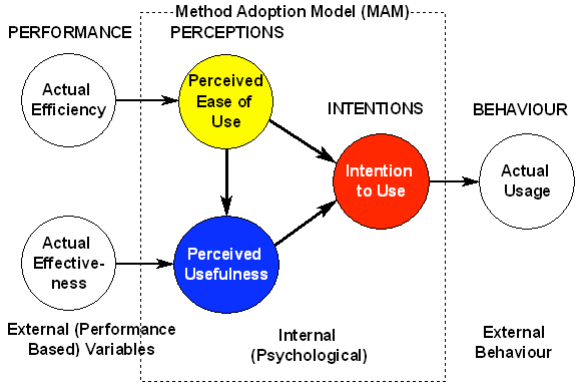

Figure 1. The Method Evaluation Model

In the MEM, efficacy is defined as the *efficiency* and *effectiveness* to which a method achieves its objectives. Thus, the evaluation of the efficacy of a method requires measurement of both effort required (efficiency) and the quality of the results (effectiveness). The core of the MEM, called the Method Adoption Model (MAM), is based on the Technology Acceptance Model (TAM) [8], a well-known validated model for evaluating information technologies. Thus, the constructs of the model are:

- *Actual Efficacy*, which consists of two parts:
  - *Actual Efficiency:* the effort required to apply a method.
  - *Actual Effectiveness:* the degree to which a method achieves its objectives.
- *Perceptions*, which consists of two variables:
  - *Perceived Ease of Use (PEOU):* the degree to *which* a person believes that using a particular method would reduce the effort.
  - *Perceived Usefulness (PU):* the degree to *which* a person believes that a particular method will achieve its intended objectives.
- *Intention to Use (ITU):* the extent to which a person intends to use a particular method.
- *Actual Usage:* the extent to which a method is used in practice.

To evaluate FSM procedures, the constructs of the MEM must be operationalized for use with this kind of method. In the next section, we show how the MEM was used to evaluate OOmFPWeb.

3. A Controlled Experiment and its Replica

According to the Goal-Question-Metric (GQM) template [4], the goal of the original experiment and the replica was to **analyze** functional size measurement **with the purpose of** evaluating OOmFPWeb **with respect to its** efficacy and likely adoption in practice **from the point of view** of the researchers **in the context of** postgraduate students in Computer Science.

The broad research questions addressed were:

- RQ1: Is OOmFPWeb efficacious?
- RQ2: Is OOmFPWeb likely to be adopted in practice?
- RQ3: Is the MEM a valid evaluation model for evaluating OOmFPWeb?

As there is currently no standard FSM procedure for sizing Web applications, we cannot evaluate OOmFPWeb against a control method. Hence, instead of using a control group or letting each subject be its own control, we decided to evaluate the performance-based variables in a qualitative way by comparing the data collected for OOmFPWeb against similar performance data reported in industry or in other empirical studies. To evaluate the perception/intention-based variables a more quantitative analysis is used.

The replica was conducted under the same conditions of the original experiment (strict replication) [3]. The same experimental materials were used, except that they were translated from Spanish to English.

3.1 Planning

3.1.1 Subjects selection. The subjects that participated in the original experiment were 15 students in the PhD Program in Software Engineering at the Valencia University of Technology in Spain. These students were chosen for convenience, i.e., they were students enrolled in a Web Engineering course during the period of March until July of 2004. The subjects of the replica were 27 postgraduate students enrolled in a course on Web Engineering at the University of Klagenfurt (Austria) during November of 2004. This course was selected because the necessary preparation and training, and the experimental task itself fitted well into the scope of this course. The experiment was organized as a mandatory part of the course.

3.1.2 Variables Selection. The independent variable is OOmFPWeb. Two types of dependent variables were selected: performance-based and perception/intention-based variables. We distinguish between two performance-based variables:

- *Productivity (PROD)*: the size of the conceptual model per unit of time (i.e., hour).
- *Reproducibility (REPR)*: the agreement between the measurement results of different subjects using OOmFPWeb (for the same application).

The ISO/IEC 14143-3 [12] also suggests other performance criteria such as repeatability and accuracy. Repeatability refers to the agreement between the measurement results of the same subject, for the same system and procedure (OOmFPWeb), but taken at different moments in time. As OOmFPWeb is currently a manual procedure, there is the risk that subjects will remember their previously obtained results when asked to measure the conceptual model again. So this measure might not be reliable in experimental settings employing human participants.

Accuracy is defined as the agreement between the measurement value and its 'true value'. The evaluation of accuracy assumes that there is another, supposedly right way of finding the 'true value' of functional size. However, in our case, no generally accepted measurement procedure exists to size Web applications according to the IFPUG 4.1.1 method. So, there is no independent way of obtaining the 'true value' of functional size. Therefore, effectiveness was measured in terms of productivity and reproducibility.

To evaluate the perceived efficacy and intention to use OOmFPWeb, the three perception/intention-based variables of the MEM were selected:

- *Perceived Ease of Use (PEOU):* the degree to which a subject believes that using OOmFPWeb would be efficient.
- *Perceived Usefulness (PU):* the degree to which a subject believes that OOmFPWeb will be effective in achieving its intended objectives.
- *Intention to Use (ITU):* the degree to which an individual intends to use OOmFPWeb.

3.1.3 Hypotheses formulation. The following hypotheses were formulated to test the first two research questions:

- *H1:* OOmFPWeb is efficient when compared to current industry practices
- *H2:* OOmFPWeb is effective when compared to similar studies reported in literature
- *H3:* OOmFPWeb is perceived as easy to use
- *H4:* OOmFPWeb is perceived as useful
- *H5:* There is an intention to use OOmFPWeb

These hypotheses relate to a direct relationship between the use of OOmFPWeb and the users' performance, perceptions and intentions. The MEM also proposes a number of relationships that indicate causal links between dependent variables. To test the predictive power of the MEM (research question 3) the following hypotheses were formulated:

- *H6:* PEOU is determined by Productivity
- *H7:* PU is determined by Effectiveness
- *H8:* PU is determined by PEOU
- *H9:* ITU is determined by PEOU
- *H10:* ITU is determined by PU

3.1.4 Instrumentation. The material prepared for the replica was composed of an experimental object including training materials and a survey instrument. The experimental object was the OOWS conceptual model for a photography agency Web application (it contains an Object Model with 10 classes and a Navigational Model with one navigational map and 14 navigational contexts). The following training materials were prepared: a set of instructional slides describing OOmFPWeb; a case study that describes an example application of OOmFPWeb; and a guideline with the rules of the procedure.

The survey instrument[2] included 13 closed questions, which were based on the items used to measure the constructs of the MAM [15]. The questions were formulated using a 5-point Likert scale, using the opposing statements question format. PEOU is measured using 5 items on the survey (Questions 1, 3, 4, 6, and 9). PU is measured using 5 items on the survey (Questions 2, 5, 8, 10, and 11). Finally, ITU is measured using 3 items on the survey (Questions 7, 12, 13). For instance, the first question of survey is as follows: *I found the procedure for applying the method simple and easy to follow*.

The experiment includes two tasks: a measurement task and a post-task survey. In the measurement task, each subject used the OOmFPWeb measurement rules for measuring an OOWS conceptual model. This task was used to collect data for evaluating the performance-based variables. Next, in the post-task survey task students were asked to complete the survey instrument in order to collect data for evaluating the perception/intention-based variables. The same experimental object and training materials as used in the original experiment was employed in the replica. However, they were translated to English. As the diffusion of the experimental data is important to external replication, the materials are available at: http://ww.dsic.upv.es/~sabrahao/OOmFPWeb.

[2] http://www.dsic.upv.es/~sabrahao/FSM/survey.html

3.2 Operation

Subjects were given an intensive training session before the experiment took place. However, the subjects were not aware of what aspects we intended to study. The training session consisted of four hours. In the first two hours, we explained the OOmFPWeb measurement rules and demonstrated their application using some toy examples. In the other two hours, the subjects used the measurement rules to size a complete case study.

The experiment took place in a single room. We gave the subjects all the experimental materials. The experiment execution was controlled. Therefore, no interaction between subjects occurred. To avoid a possible ceiling effect, there was no time limit on sizing the OOWS conceptual model. After they finished the measurement task the subjects were asked to perform the post task survey. To avoid a potential bias in subject responses, the subjects were told their answers would be treated anonymously. Before filling the survey, the students were also informed that their grade on the course would not be affected by their performance in the experiment.

The performance-based dependent variables were measured using a data collection form. This form records the outputs of the OOmFPWeb functional size measurement and the time spent for sizing the OOWS conceptual model. We called this time measurement time, expressed in hours. Once the data were collected, we verified whether the tests were complete. As all tests were completed, we took into account the responses of all subjects.

3.3 Experimental Results

Detailed results about the efficacy and likely adoption of OOmFPWeb are reported in [1]. Table 1 summarizes the results of the original experiment.

Out of ten hypotheses, seven were supported. Hypothesis H6 was not confirmed since the relationship was not statistically significant. This suggests that effects of productivity on perceived ease of use are not significant. One plausible explanation may be that this variable is determined by other factors such as the counter experience.

Similarly, H8 and H10, which are related to the MAM constructors, were not confirmed. A possible reason could be the low correlation existing among some items of the survey used. This issues need to be investigated in the replication study. More important than the actual results obtained was the evaluation of the experimental materials and process. No particular deficiencies were found in the materials.

Table 1. Results of the Original Experiment

Hypotheses	Sig.	Confirmed
H1: Productivity	-	Yes
H2: Reproducibility	-	Yes
H3: Perceived Ease of Use	.000	Yes
H4: Perceived Usefulness	.000	Yes
H5: Intention to Use	.000	Yes
H6 Productivity → PEOU	.740	No
H7: Reproducibility → PU	.017	Yes
H8: Perceived Ease of Use → PU	.292	No
H9: Perceived Ease of Use → ITU	.001	No
H10: Perceived Usefulness → ITU	.028	Yes

Table 2 shows descriptive statistics for the variables in the original experiment.

Table 2. Descriptive statistics for variables in the experiment

Variables	Min.	Max.	Mean	Std. Dev.
PROD	85.50	140.00	108.79	18.52
REPR	0.01	0.19	0.06	0.04
PEOU	3.00	5.00	3.86	0.58
PU	2.20	4.60	3.80	0.67
ITU	1.67	5.00	3.73	1.01

4 Analysis and Interpretation

The results for the replica are presented according to the research questions stated. The data were analyzed using the following levels of significance: not significant ($p > 0.1$), low significance ($p < 0.1$), medium significance ($p < 0.05$), high significance ($p < 0.01$) and very high significance ($p < 0.001$).

4.1 Analysis of the Actual Efficacy

The efficiency of OOmFPWeb was evaluated by comparing the measurement productivity of subjects using OOmFPWeb against reported industry averages. We are aware that the productivity of people in sizing a specification can vary considerably. It depends on many factors such as experience, the quality of the specifications, the use of tools, etc. Notwithstanding these limitations, a comparison of the productivity observed in the experiment with what is considered as being acceptable in industry, gives us some basis to assess the performance of people using the procedure.

According to industry experience[3] we can expect "counting rates" of 300 FP per day (FP/day) by first-time counters with one day's training. The IBM[4] expects someone to be able to count 100 FP/day with one week of counting assistance. However, the 100

[3] http://www.functionpoints.com/faq.asp#a14

[4] http://ourworld.compuserve.com/homepages/softcomp/fpfaq.htm

FP/day counting rate may be right if it includes the preparation and possible presentation of a complete project review. As further evidence, Total Metrics published a document describing different counting levels [20]. According to these levels, the productivity of an estimator can vary between 200-750 FP/day.

Based on the performance data reported in industry, we can consider that the lowest productivity for first-time counters is 200-300 FP/day. As a day is assumed to have 8 working hours, the productivity rate is approximately 25-37.5 FP/hour. In this study, we adopted 37.5 FP/hour as the benchmark industry rate for the productivity of novice function point counters.

To calculate the productivity of a subject, we divided the subject assessment by the measurement time. As can be seen in Table 3, the mean productivity that was observed for the Austria dataset is 102.80, which is almost three times the size of the benchmark. This provides evidence for our hypothesis H1. Although the replica provides empirical evidence of the subjects' productivity using OOmFPWeb, it must be noted that the industry values that are reported are not specific to Web projects.

Table 3. Descriptive statistics for productivity and reproducibility

OOmFPWeb	PROD Austria	REPR Austria
Minimum	85.49	0.00
Maximum	146.00	0.09
Mean	102.80	0.039
Standard deviation	12.77	0.027

Next, the effectiveness of OOmFPWeb was evaluated in terms of reproducibility (also referred in literature as inter-rater reliability). We assume that the closer the measurements obtained by different raters, the more effective the FSM procedure is.

There are some published studies that address the inter-rater reliability question. In a first study reported by Rudolph [19], 20 subjects calculated the Function Point (FP) value for a system based on its requirements specification. Values within the range of ±30% of the average FP value were observed. This is consistent with the findings of Low and Jeffery [14] who observed an inter-rater consistency within of ±30%, with an error rate of 42%. In the study of Kemerer [13], twenty-seven actual applications were sized using two FSM methods: the IFPUG method [11] and the ER method. The mean inter-rater reliability obtained using the IFPUG method was 26.53%, whereas the mean value obtained for the ER method was 20.66%.

To measure the degree of variation between assessments produced by different subjects using OOmFPWeb, we proposed a practical statistic similar to that proposed by Kemerer [13]. This statistic is calculated as the difference in absolute value between the count produced by a subject and the average count (for the same FSM method) produced by the other subjects in the sample, relative to this average count. This means that lower values indicate higher reproducibility. Reproducibility measurements (REP) were thus obtained by applying the following equation:

$$REP_i = \left| \frac{\frac{\sum_{k=1,k\neq i}^{n} FPValues_k}{n-1} - FPValue_i}{\frac{\sum_{k=1,k\neq i}^{n} FPValues_k}{n-1}} \right|$$

Table 3 shows descriptive statistics for reproducibility. Compared to the results reported by Kemerer the reproducibility of measurements was high (mean REPi was 3.9%). The variation around the mean subject assessment (i.e. range of values divided by mean value) was 15%. These results compare well with Rudolph's study [19]. Finally, the standard deviation in subject assessments divided by the mean subject assessment was 4.6%. This value corresponds to a fraction of what was reported by Low and Jeffery [14].

4.2 Analysis of the Perceived Efficacy and Likelihood of Adoption

To evaluate the perceived efficacy and likely acceptance in practice of OOmFPWeb, we tested hypotheses H3, H4 and H5. These hypotheses were tested by verifying whether the scores that the students assign to the constructs of the MAM were significantly better than the neutral score (i.e. the score 3), on the Likert scale for an item. Table 4 shows descriptive statistics for PEOU, PU and ITU.

Table 4. Descriptive statistics for perception-based variables

Variable	Min.	Max.	Mean	Std. deviation
PEOU	2.40	5.00	3.97	0.75
PU	2.20	4.80	3.69	0.61
ITU	2.33	4.67	3.59	0.70

The Kolmogorov-Smirnov test for normality was applied to the PEOU, PU and ITU data. As the distributions were normal, we used the One-tailed sample t-test to check for a difference in mean PEOU, PU, and ITU score for OOmFPWeb and the value 3. The results (see Table 5) allowed us to empirically demonstrate that participants perceived OOmFPWeb to be easy to use, useful, and that there is an intention to use OOmFPWeb in the future. Thus, H3, H4, and H5 were re-confirmed. The statistical significance of the results was very high for all hypotheses ($p < 0.001$).

Table 5. 1-tailed t-test rank for perception-based variables

Variable	Std. error mean	t	p-value
PEOU	0.977	6.746	0.000
PU	0.696	5.921	0.000
ITU	0.592	4.399	0.000

The construct validity of the survey instrument was evaluated using an inter-item correlation analysis. All items in the survey were found to be valid. We also evaluated reliability using Chronbach's alpha. All constructs were found to be reliable, i.e., they have an alpha value equal to or greater than 0.7 (PEOU=0.75, PU=0.81, and ITU=0.75).

4.3 Analysis of the Causal Relationships

According to MEM, there are a number of hypothesized causal relationships among the dependent variables in our study (H6 to H10). To test these hypotheses, we used regression analysis since they are causal relationships between continuous variables.

H6: Productivity → Perceived Ease of Use. It verifies if perceptions of efficiency are determined by actual efficiency. The regression equation for the Austria dataset resulting from the analysis is:

*PEOU = 1.86 + 0.02 * Productivity*

The regression was found to be highly significant, with $p < 0.01$ (see Table 6). The r2 statistic showed that Productivity accounts for 22% of the variance in PEOU. This means that H6 was confirmed.

Table 6. Simple regression between PEOU and Productivity

Model Variables	Constant	Productivity
Unstd. coef. (b)	1.859	0.016
Std. Error	0.808	0.006
Std. coef. (beta)		0.469
t	2.302	2.659
Sig.	0.015	0.006

H7: Reproducibility → Perceived Usefulness. It verifies if Perceived Usefulness (PU) is determined by Reproducibility. The regression equation resulting from the analysis is:

*PU = 3.84 - 3.50 * Reproducibility*

As expected, the regression coefficient for reproducibility is negative, meaning that the higher the reproducibility value the lower the PU value is. Therefore, if the reproducibility value decreases towards zero (which actually means higher reproducibility), then perceived usefulness increases.

However, the result of the regression analysis (see Table 7) does not allow us to empirically corroborate that PU is determined by reproducibility. The regression coefficient for reproducibility was not significant ($p > 0.1$). Therefore, H7 was not confirmed in the Austria dataset.

Table 7. Simple regression between PU and Reproducibility

Model Variables	Constant	REPR
Unstd. coef. (b)	3.835	-3.500
Std. Error	0.208	4.318
Std. coef. (beta)		-0.160
t	18.430	0.811
Sig.	0.000	0.212

H8: Perceived Ease of Use → Perceived Usefulness. It verifies if perceived usefulness is determined by perceived ease of use. The regression equation resulting from the analysis is:

*PU = 2.55 + 0.29 * Perceived Ease of Use*

As shown in Table 8, the regression coefficient was found to be medium significant ($p < 0.05$). With respect to the predictive power of the model, PEOU explains 13% of the variance in PU. This means that H8 was re-confirmed.

Table 8. Simple regression between PEOU and Productivity

Model Variables	Constant	PEOU
Unstd. coef. (b)	2.548	0.289
Std. Error	0.614	0.15
Std. coef. (beta)		0.356
t	4.152	1.904
Sig.	0.000	0.034

H9 and H10: Perceived Usefulness + Perceived Ease of Use → Intention to Use. The multiple regression equation resulting from the analysis is:

*Intention to Use = -1.27 + 0.28 * PEOU + 0.33 * PU*

The result of the regression summarized in Table 9 allows us to empirically corroborate that intention to use is determined by perceived ease of use and perceived usefulness. The regression coefficients for PEOU and PU were found to be significant. This means that H9 and H10 were confirmed. This also indicates that perceptions in intention to use are partially determined by perceptions in PEOU and PU. With respect to the predictive power of the model, PEOU and PU together explain 23% of the variance in Intention to Use, as indicated by r^2.

Table 9. Multiple regression between PEOU, PU and ITU

Model Variables	Constant	PEOU	PU
Unstd. coef. (b)	1.268	0.277	0.331
Std. Error	0.874	0.178	0.219
Std. coef. (beta)		0.298	0.289
t	1.452	1.556	1.511
Sig.	0.079	0.066	0.072

5. Conclusions and Further Work

This paper has presented a replication of an experiment to evaluate the efficacy and likely adoption of OOmFPWeb. The results show that:

- OOmFPWeb is efficient when compared to current industry practices. We found support for hypothesis H1 in both experiments.
- OOmFPWeb is effective when compared to similar studies reported in the literature. Hypothesis H2 was supported in both experiments.
- OOmFPWeb was perceived to be easy to use and useful. There also is an intention to use the procedure in the future. Hypotheses H3, H4 and H5 were confirmed in both experiments.
- The MEM seems to be useful to evaluate FSM procedures such as OOmFPWeb.

Hypotheses H6, H8, and H9 could not be confirmed in the experiment but were confirmed in the replica. However, H7, which was confirmed in the original experiment, could not be re-confirmed in the replica. A possible explanation is that the participants did not know the results of their measurement. Thus, they did not have the perception of usefulness of the method they applied. An improvement in our experiment procedure could be to present the results to the students prior to getting them to do the surveys. This issue will be investigated in further experiments.

In general, our results support the three research questions stated. Running replicated experiments instead of a single experiment provides more evidence of the external validity of the results. The same hypotheses were tested and confirmed (with few exceptions) in a different environment. The replication provides further evidence of the hypotheses confirmation. Thus, we can conclude that the general goal of the empirical evaluation has been achieved. We believe that such evaluation, prior to the actual technology transfer to the intended user community, is desired to obtain feedback on our proposal and adjust or fine-tune it before promoting its use in industry.

The main limitation of the experiments was the use of student as participants. However, they were PhD and postgraduate students. So at least they can be considered as representative of novice users of functional size measurement methods. To increase external validity, the current study needs to be replicated using practitioners experienced in FSM.

However, our study adds new insights into the problem of how to evaluate alternative FSM methods or procedures. The work also contributes to the body of knowledge about experimentation in the field of Web Engineering. Future work includes the replication of the experiment with practitioners.

References

[1] Abrahão, S., Poels, G., Pastor, O. 2004. Evaluating a Functional Size Measurement Method for Web Applications: An Empirical Analysis. In: Proc. of the 10th IEEE Software Metrics Symposium (METRICS'04), USA, pp. 358–369.

[2] Abrahão, S. and Pastor, O. 2003. Measuring the Functional Size of Web Applications. International Journal of Web Engineering and Technology, Inderscience Enterprises Ltd., England, 1 (1), 5–16.

[3] Basili, V., Shull, F., Lanubile, F., 1999. Building knowledge through families of experiments. IEEE Transactions on Software Engineering 25 (4), 435–437.

[4] Basili, V. R., Rombach, H. D., 1988. The TAME Project: Towards Improvement-Oriented Software Environments, IEEE Trans. on Software Engineering, 14 (6), 758–773.

[5] Cândido, E. J. D., Sanchez, R., 2004. Estimating the size of web applications by using a simplified function point method. In: Proc. of WebMedia/LA-WEB 2004, pp. 98- 105.

[6] Cleary, D., 2000. Web-Based Development and Functional Size Measurement. In: Proceedings of IFPUG Annual Conference, San Diego, USA.

[7] Cost Xpert Group, Estimating Internet Development, http://www.costxpert.com/Reviews_Articles/SoftDev/.

[8] Davis, F.D., 1989. Perceived Usefulness, Perceived Ease of Use and User Acceptance of Information Technology, MIS Quarterly, 13(3), 319–340.

[9] Fraternali, P., Tisi, M., Bongio, A. Automating Function Point Analysis with Model Driven Development, In Proc. of CASCON 2006, Markham, Ontario, Canada.

[10] IFPUG. Hints to Counting Web Sites: IFPUG White Paper, 1998.

[11] ISO/IEC, 2003. ISO/IEC 20926: 2003, Software engineering - IFPUG 4.1 Unadjusted functional size measurement method - Counting practices manual.

[12] ISO/IEC, 2003. ISO/IEC TR 14143-3: 2003, Information technology - Software measurement - Functional size measurement - Part 3: Verification of FSM methods.

[13] Kemerer, C. F., 1993. Reliability of Function Points Measurement, A Field Experiment. Communications of the ACM, 36(2), 85–87.

[14] Low, G. C., Jeffery, D. R. 1990. Function Points in the estimation and evaluation of the software process, IEEE Transactions on Software Engineering, 16(1), 64–71.

[15] Moody, D. L., 2001. Dealing with Complexity: A Practical Method for Representing Large Entity Relationship Models, PhD Thesis, University of Melbourne, Australia.

[16] Ochoa, S. F., Bastarrica, M. C., Parra, G., 2003. Estimating the Development Effort of Web Projects in Chile, Proc. LA-WEB'03, Chile, pp. 114-124.

[17] Pastor, O. Fons, J. Pelechano, V., Abrahão, S. 2005. Conceptual Modelling of Web Applications: the OOWS approach, Web Engineering, Springer.

[18] Reifer, D. 2000. Web Development: Estimating Quick-to-Market Software, IEEE Software, 17 (6), 57–64.

[19] Rudolph, E., 1983. Productivity in computer application development, Working paper 9, University of Auckland, Department of Management Studies, New Zealand.

[20] Total Metrics, 2001. Total Metrics - Levels of Counting, Australia August 2001.