An Empirical Study on the Likelihood of Adoption in Practice of a Size Measurement Procedure for Requirements Specification

Nelly Condori-Fernández, Oscar Pastor
Department of Information Systems and Computation
Valencia University of Technology
Valencia-Spain
{nelly,opastor}@dsic.upv.es

Abstract

Software size is one of the essential parameters of the estimation models used for project management purposes, and therefore being able to measure the size of software at an early stage of the development lifecycle is in theory of crucial importance. However, although some proposals for functional size measurement in industry do currently exist, there is as yet little validating evidence for such proposals.

This paper describes an empirical study, based on the Method Evaluation Model (MEM), of users' perceptions resulting from actual experience of use of a measurement procedure called RmFFP. This procedure was designed in accordance with the COSMIC-FFP standard method for estimating the functional size of object-oriented systems from requirements specifications obtained in the context of the OO-Method approach.

In addition, an analysis of MEM relationships was also carried out using the regression analysis technique.

1. Introduction

We have based our work on the assumption that measuring system functionality in the early phases of a software production process is vital for optimum resource management. In recent years, we have been working on a software production method based on the transformation of models, called the OO-Method [20]. As an enrichment of this method, our intention is to facilitate functional size measurement for high-level specifications automatically. To do this, a measurement procedure [7] has been designed in accordance with the COSMIC-FFP standard method for estimating the functional size of object-oriented systems generated with the OO-Method approach.

In the specific area of functional size measurement of requirements specifications, we provide two main contributions. The first is a successful combination of well-known Requirements Engineering (RE) and Empirical Software Engineering (ESE) practices. The second is a concrete application of MEM to evaluate the RmFFP measurement procedure.

There are a number of current proposals that attempt to resolve the problem of measuring system functionality at the requirements level using an approach based on COSMIC-FFP [2],[4],[11]. A positive aspect of our proposal is that size estimated is much closer to size obtained in the final software product. This is possible due to the code generation features of our OO-Method approach, with semi-automated generation of a final software product from a software requirements specification. In this way, the traceability and consistency of the requirements specifications contribute to the measurement quality obtained using RmFFP.

This paper is structured as follows. In section 2, we briefly introduce the most relevant characteristics of both the OO-Method Requirements Modelling Process and the RmFFP procedure. Section 3 contextualizes our work by explaining the main aspects of the MEM. Section 4 describes how this model has been applied to evaluate the likelihood of adoption in practice of RmFFP. The paper ends with the presentation of conclusions and an indication of possible future work.

2. The Requirements Model and the RmFFP measurement procedure

RmFFP is a functional size measurement procedure designed on the basis of the COSMIC-FFP standard method, which has been approved by ISO/IEC 19761 [15]. The object to be measured is the functional requirements specification obtained using the OO-Method requirements model [14], which is described below.

* Supported by the DESTINO project, ref. TIN2004-03534, Spain

2.1. OO-Method Requirements Model

The purpose of this model is to gather user requirements to build a conceptual schema that is used to automatically generate a final software system. To do this, the requirements model includes a set of techniques such as the Mission Statement, the Function Refinement Tree, the Use Case Model, and the Sequence Diagram Model [14].

The Mission Statement describes the nature and purpose of the system. The Function Refinement Tree (FRT) is a hierarchical decomposition of the business functions of the system. The leaves of this tree represent the functions of the desired system and are the entry point for building the Use Cases Model, which models the system functional requirements. The leaf nodes of the FRT are considered to be primary use cases. It is also possible to have secondary use cases, which organize and manage complexity through EXTEND and INCLUDE relationships.

Each use case is then represented semi-automatically by one or more Sequence Diagrams [13]. These are interaction patterns that are represented by means of a set of messages among required classes to perform the system behaviour. These messages are labelled with different stereotypes (signal, service, query, connect), which allows us to later identify the different elements of the OO-Method conceptual schema. In this way, we assure consistency between the requirements specification and the subsequent phases of the development process [13], [14]. Furthermore, this approach is supported by a tool called RETO (Requirements Engineering Tool), which captures user requirements and semi-automatically generates elements of an OO-Method conceptual schema.

2.2. RmFFP: A size measurement procedure

In order to estimate the functional size of object-oriented systems from functional requirements specification obtained using the Requirements Model, the RmFFP procedure was proposed [7].

This procedure starts with the definition of the *measurement context*, which includes the purpose, the scope, and the measurement viewpoint. As state above, the scope of RmFFP comprises the functionality to be included in a particular measurement; this functionality is specified using the OO-Method requirements model [14]. The *measurement viewpoint* corresponds to the 'end-user' viewpoint, which will focus on a requirements specification (object of interest).

Then, RmFFP starts a *mapping phase* to identify the significant primitives of the Requirements Model that contribute to the system's functional size according to the concepts of the COSMIC-FFP metamodel. To do this, we defined sixteen mapping rules whose principal purpose is to reduce misinterpretation about the COSMIC-FFP generic concepts and to facilitate the automation of the RmFFP procedure. For instance, each use case is identified as a functional process, each message of the sequence diagram is identified as a data movement type, etc.

As a data movement is the fundamental component of the COSMIC-FFP method, we also defined four additional rules to eliminate duplicated data movements, which are explained in more detail in [8].

Once the data movements have been correctly identified, we proceed with the *measurement phase*, whose purpose is to produce a quantitative value that represents the software functional size of a requirements specification. To do this, we apply the *measurement function*, which consists of assigning a numerical value of 1 Cfsu (Cosmic Functional Size Unit) to each data movement. We defined four rules to add together these quantified data movements. To do this, we used the relationship types between use cases to calculate the size of the functional processes (use case) and the size of the entire system.

3. Overview of the Method Evaluation Model

The Method Evaluation Model (MEM) has been proposed by Moody [17] for evaluating IS design methods that we have used in the context of functional size measurement methods, and which has also been applied by Abrahao et al. to evaluate a measurement procedure called OOmFP [1].

MEM combines Rescher's theory of pragmatic justification [21], and Davis's Technology Acceptance Model (TAM) [9]. The core of the MEM consists of the same perception-based constructs as the TAM, but which are adapted to evaluate methods. These constructs are called the Method Adoption Model (MAM) [17]:

- *Perceived Ease of Use:* the extent to which a person believes that using a particular method would be free of effort.
- *Perceived Usefulness:* the extent to which a person believes that a particular method will be effective in achieving intended objectives.
- *Intention to Use:* the extent to which a person intends to use a particular method.

MAM is extended with additional constructs that provide inputs to this model and predict its ultimate output (i.e. whether the method will be used in practice). These additional constructs are:

- *Actual Efficiency:* the effort required to apply a method. This represents a MAM input variable.
- *Actual Effectiveness:* the extent to which a method achieves its objectives. This also represents a MAM input variable.
- *Actual Usage:* the extent to which a method is used in practice. This represents a MAM output variable.

The input variables relate to actual performance of users in employing a method. The assumption underlying of the MEM is that perceptions of efficiency and effectiveness are the result of actual experience with the method and that a user's performance in using a method has an impact on the user's perception. This, in turn, determines the user's intention to adopt the method. The actual usage of a method will be determined by this intention to use.

In the next section, we report upon how this model was applied in the evaluation of adoption in practice of the RmFFP procedure.

4. Evaluating the adoption in practice of the RmFFP procedure

To evaluate the likelihood of adoption in practice of RmFFP, we designed an experiment using the framework proposed by Wohlin et al. [23].

According to the Goal/Question/Metric template [3], the goal of the experiment is to **analyze** functional size measures **for the purpose** of evaluating RmFFP **with respect to** its adoption in practice **from the viewpoint** of the researcher **in the context** of Computer Science students measuring OO-Method requirements specifications.

The following research questions were addressed by this experiment:

RQ1: Is RmFFP perceived as easy to use and useful?
RQ2: Is there an intention to use RmFFP in the future?
RQ3: Are the perceptions really a result of the actual performance using RmFFP?
RQ4: Is the intention to use really a result of the perceptions experienced by the subjects using RmFFP?

4.1. Experiment Planning

4.1.1. Subjects. The subjects were 35 computer science students at the Valencia University of Technology who had similar backgrounds in the use of the OO-Method Requirements Model. These subjects were students enrolled in the "Software Development Environments" course from February until June of 2005. The experiment was organized as a mandatory part of this course. Two groups of students were formed because some students could not regularly attend class due to work commitments in companies. The first group was made up of 18 students who had no link (e.g. work experience) to companies and the second group was made up of 17 students who had some connection with companies.

Final-year students rather than practitioners were used as experimental subjects for the following reasons:

Availability: the possibility of involving practitioners is difficult given time and cost constraints. The costs and benefits of empirical studies with students are discussed by Carver et al. [6].

Similarity: final-year students are future practitioners [12]. These students will be the next generation of professionals and are therefore a close match to the population under study.

4.1.2. Independent and Dependent Variables. The *independent variable* was the FSM procedure being evaluated: RmFFP.

In order to answer the two first research questions, we considered three *dependent variables:* perceived ease of use, perceived usefulness, and intention to use.

- D1: Perceived Ease Of Use (PEOU): the extent to which a subject believes that using RmFFP will be effort-free.
- D2: Perceived Usefulness (PU): the extent to which a subject believes that RmFFP will be effective in achieving the objectives.
- D3: Intention To Use (ITU): the extent to which an individual intends to use RmFFP.

In order to answer the third research question, we need to consider two *dependent variables* that represent actual performance using RmFFP:

- D4: Measurement productivity: defined as the number of size units that can be measured per unit of time (e.g. per hour).
- D5: Reproducibility (REP): defined as the closeness between results of measurements of the same measurand carried out under changed conditions of measurement [16]. In our experiment, the "changed condition of measurement" was basically the different subjects used.

Accuracy, which is defined as the closeness between the result of a measurement and the true value of the measurand[1] [16], was not considered in this study. This is due to the fact that this "true" value is usually established through a consensus of experts. However, the issue of certification for COSMIC-FFP is currently being investigated by the COSMIC group.

[1] The particular quantity that is subject to measurement [16]

4.1.3. Hypotheses. We identified the following hypotheses:

- H1: RmFFP is perceived as easy to use.
- H2: RmFFP is perceived as useful.
- H3: There is intention to use RmFFP.

In addition, from the last two research questions defined, we identified the following hypotheses, which are relationships between dependent variables defined above:

- H4: Perceived ease of use is determined by productivity.
- H5: Usefulness perceived is determined by reproducibility.
- H6: Usefulness perceived is determined by perceived ease of use.
- H7: Intention to use is determined by perceived ease of use.
- H8: Intention to use is determined by usefulness perceived.
- H9: Usefulness perceived is determined by perceived ease of use and actual effectiveness.
- H10: Intention to use is determined by perceived ease of use and usefulness perceived.

4.1.4. Instrumentation. The instruments used in this experiment included the experimental object and training materials. The *experimental object* was an OO-Method requirements specification of a Car Rental application with thirty-five use cases. The *training materials* were the following: a set of instructional slides on the OO-Method Requirements Model and the RmFFP procedure; a case-study that describes an example of the application of RmFFP, a measurement guide, and another case study to verify the training carried out.

In addition, a survey instrument[2] was adapted to measure the three perception-based variables. It included thirteen closed questions that were based on the constructs of the MAM. The items that we used were formulated using a 5-point Likert scale with the opposing-statement question format. The order of the items was randomized and half the questions negated to avoid monotonous responses: Perceived Ease of Use was measured using 5 items on the survey (Questions 1, 3, 4, 6, and 9); Perceived Usefulness was measured using 5 items on the survey (Questions 2, 5, 8, 10, and 11); and finally, Intention to Use was measured using 3 items on the survey (Questions 7, 12, 13).

4.1.5. Experiment operation. Three groups of experimental tasks were carried out during the training task, measurement task, and the post-task survey.

With respect to the *training task*, we used the demonstration/practice method [10]. For the demonstration part, we considered the following tasks: (a) presentation of the OO-Method Requirements Model, (b) use of the RETO tool, (c) presentation of the RmFFP measurement procedure, and (d) illustration of the use of RmFFP with an illustrative example of a case study. For the practice part, we considered the following tasks: (a) application of the theory presented in the case study and guided by the instructor, (b)the application of RmFFP to a case study (the students could clarify their doubts), and (c) verification of knowledge learned by the student by working out an assigned case study. The time used for all the tasks included in this first session was eight hours distributed over four days.

With respect to the *measurement task*, each subject used the RmFFP measurement guideline to measure an OO-Method requirements specification (rent a car). This task was used to collect data to evaluate the performance-based variables.

With respect to the *post-task survey*, each subject was asked to complete a survey to evaluate RmFFP. This survey was used to collect data to evaluate the perception-based variables.

Before the subjects took the test, the designed experiment was conducted with another small group of people in order to improve it and ensure that the documentation was well-designed. No changes were necessary as a result of this pre-test.

4.2. Data analysis and interpretation

The analysis and interpretation of the results were divided into two phases:

4.2.1. Adoption in Practice. To evaluate the perceived efficiency, the perceived efficacy, and the intention to use RmFFP, hypotheses H1, H2, and H3 were formally tested by verifying whether the scores that the students assigned to the constructs of the MAM were significantly better than the middle score on the Likert scale for an item.

Firstly, the scores of each subject were averaged over the different items that are relevant for a construct. We obtained three mean values for each subject (see Appendix, Table A1). Descriptive statistics for the three variables are presented in Table 1. Note that the highest mean value corresponded to PEOU, and the value range oscillated between 3.2 and 5.0. However, although the mean value of the ITU variable was acceptable, certain values were lower than 3 because the values range oscillated between 2.67 and 4.67.

[2] http://www.dsic.upv.es/~nelly/survey.pdf

Table 1. Descriptive statistics for perception variables

Statistic	PEOU	PU	ITU
Mean	4.2424	3.9394	3.8586
Standard dev.	0.5190	0.4108	0.4932
Minimum	3.20	3.00	2.67
Maximum	5.00	4.60	4.67

Secondly, the Kolmogorov-Smirnov test was applied for all the PEOU, PU, and ITU values of each subject. As this data distribution was also normal, we again used the one-tailed sample t-test to check the difference between the mean PEOU, PU, ITU and the value 3 (middle score). The statistical test was applied with a significance level of 5 %, i.e., alpha = 0.05.

The results of the t-tests (shown in Table 2) allow the rejection of the null hypotheses (with a very high significance level) meaning that we empirically corroborated that RmFFP is perceived as easy to use and useful, and that there is an intention to use it in the future.

Table 2. One Sample t-test for perception variables

Statistic	PEOU	PU	ITU
Mean Difference	1.2424	0.9394	0.8586
95% Conf. Interval for the diff.	1.0584 (lower) 1.4265 (upper)	0.7937 (lower) 1.0850 (upper)	0.6837 (lower) 1.0335 (upper)
T	13.7510	13.1380	10.0000
1-tailed p-value	0.000	0.000	0.000

4.2.2. MEM relationships. As the intention to use RmFFP is influenced by perceptions of actual efficiency and effectiveness, we identified several relationships, which were defined above in the last seven hypotheses (H4-H10). In this phase, we validated them by applying the regression analysis technique.

Actual efficiency using RmFFP has been evaluated in terms of productivity, and actual effectiveness in terms of reproducibility. The values obtained for these two performance variables are shown in the Appendix.

H4: Productivity → Perceived ease of use. The regression equation resulting from the analysis is: *PEOU = 2.13 + 0.02*Productivity.* The regression was found to be significant, with $P < 0.001$, which means that H4 was strongly confirmed. The determination coefficient ($r^2 = 0.57$) showed that 57% of the total variation in perceived ease of use can be explained by variation in productivity.

H5: Reproducibility → Perceived usefulness. The regression equation resulting from the analysis is: *PU = 3.79 + 3.03*Reproducibility.* The regression had a low significance level ($P < 0.1$), which means that H5 was not confirmed.

H6: Perceived ease of use → Perceived usefulness. The regression equation resulting from the analysis is: *PU = 3.46 + 0.11* PEOU.* This regression did not have a significance level ($P > 0.1$), which means that H6 was not confirmed.

H7: Perceived ease of use → Intention to use. The regression equation resulting from the analysis is: *ITU = 2.58 + 0.301* PEOU.* The regression had a low significance level ($P < 0.1$), which means that H7 was not confirmed.

H8: Perceived usefulness → Intention to use. The regression equation resulting from the analysis is: *ITU = 2.119 + 0.442* PU.* The regression had a medium significance level ($P < 0.05$), which means that H8 was confirmed. The determination coefficient ($r^2 = 0.14$) showed that 14% of the total variation in intention to use can be explained by variation in perceived usefulness.

H9: Perceived ease of use + Actual effectiveness → Perceived usefulness. The regression equation resulting from the analysis is: *PU = 3.3 + 3.04*Reproducibility + 0.12* PEOU.* This regression was not significant ($P > 0.1$), which means that H9 was not confirmed.

H10: Perceived ease of use + Perceived usefulness → Intention to use. The regression equation resulting from the analysis is: *ITU = 1.21 + 0.4* PU+ 0.26*PEOU.* The regression had a medium significance level ($P < 0.05$), which means that H10 was confirmed. The determination coefficient ($r^2 = 0.21$) showed that 21% of the total variation in intention to use can be explained by variation in perceived usefulness and perceived ease of use.

Table 3 below summarizes the regression analysis results in terms of the predictive power (r^2) and significance level of the model, and the confirmation of the casual relationships.

Table 3. Regression analysis results

MEM hypotheses	Predictive power	Significance level*	Con-firmed?
H4: D4 → D1	57%	Very high	Yes
H5: D5 → D2	9%	Low	No
H6: D1→ D2	2%	Null	No
H7: D1→ D3	10%	Low	No
H8: D2 → D3	14%	Medium	Yes
H9:D1+D5→D2	12%	Null	No
H10:D1+D2→D3	21%	Medium	Yes

Note that two hypotheses out of five were confirmed using a simple regression analysis (H4 and

* Null: $\alpha > 0.1$, Low : $\alpha < 0.1$, Medium: $\alpha < 0.05$, High: $\alpha < 0.01$, Very high: $\alpha < 0.001$

H8). This means that the perceived ease of use is determined by the actual efficiency of the measurers, and the intention to use RmFFP is determined by the perceived usefulness. However, this perceived usefulness is not determined by the measurement reproducibility (H5), nor by the perceived ease of use (H6). The perception of usefulness may be determined by other variables, such as the tangibility of the results [18], and the experience of the measurer [22].

With respect to the last two hypotheses (H9 and H10), validated with a multiple regression model, we confirmed only that the intention to use RmFFP is stronger determined by perceived usefulness than by perceived ease of use (H10).

In addition, hypotheses H6 and H7, which include only the MAM constructors, have a null and low significance level, respectively. An explanation for this could lie in the reliability analysis of the survey used, which is described in the next section.

4.3. Validity evaluation

It is important to consider validity in order to ensure that the experimental results are valid for the target population. There are various threats to the validity of results of an experiment [23]. In this section, we discuss the threats identified in our experiment.

4.3.1. Conclusion validity. Threats to conclusion validity are concerned with issues that affect the ability to draw the correct conclusion about relations between the treatment and the outcome of an experiment. The following threats were considered:

Reliability of the application of treatments to subjects: There is a risk that the RmFFP application is not the same for different people applying the treatment on different occasions. In our experiment, RmFFP (treatment) was applied following a prescribed procedure for the two defined groups. Hence, the risk of obtaining dissimilar applications for different subjects and occasions was low.

Random heterogeneity of subjects: All the subjects in the experiment had approximately the same level of experience working with the OO-Method Requirements Model. However, this homogeneity reduces the external validity of the experiment.

4.3.2. Internal validity. Threats to internal validity are influences that can affect the independent variable (RmFFP) with respect to causality, without the researcher's knowledge. The following threats were considered:

Selection: This is the effect of natural variation in human performance. In the experiment, the subjects were selected for convenience, i.e., they were students enrolled in the "Software Development Environments" course. This course was selected because it was a specialized teaching unit on Requirements Engineering. Furthermore, the experimental task itself fitted well into the scope of this course.

Instrumentation: This is the effect caused by the artefacts used in the experiment execution. The instruments used for the experiment were verified in advance. Firstly, the specifications of the case studies (objects) were reviewed by the model's author; secondly, the measurement guideline and the survey were verified in advance with a small group of people in order to improve their clarity.

4.3.3. Construct validity. Threats to construct validity concern generalizing the result of the experiment to the concept or theory behind the experiment. The following threat was considered:

Inadequate pre-operational explanation of constructs: This threat means that the constructs are not sufficiently defined, and, hence, the experiment cannot be sufficiently clear. We used an inter-item correlation analysis to evaluate the construct validity of the perception-based variables (PEOU, PU, and ITU). To do this, we used two criteria proposed by Campbell and Fiske [5]: *Convergent validity (CV),* which refers to the convergence among different indicators used to measure a particular construct; and *Discriminant validity (DV), which* refers to the divergence of indicators used to measure different constructs. This average DV should be lower than the average CV. The results of the validity analysis for each construct show that the CV value was higher than the DV value for all PEOU, PU, and ITU items (see Appendix, Table A2).

In addition, we also conducted a *reliability analysis* on the survey. The reliability was conducted using the Chronbach alpha. The generic value obtained was 0.70, indicating that the items on the survey are reliable, as claimed by Nunally [19]. However, the corresponding Cronbach alpha value for each MAM construct was: PEOU = 0.73, PU = 0.5 and ITU = 0.5. A design adjustment of the questions corresponding to the constructs PU and ITU would therefore be advisable for further empirical studies.

4.3.4. External validity. Threats to external validity concern generalization of the results to industrial practice. Here the following threats were considered:

Interaction of selection and treatment: This is the effect of not having a representative population in the experiment with which to generalize. In our case, we are aware that more experiments with a larger number of subjects (e.g., students and professionals) will be necessary to reconfirm the results obtained.

Interaction of setting and treatment: This is the effect of not having representative material. In the experiment, we tried to use a representative requirement specification of a real case study. However, more empirical studies with other user requirement specifications will be necessary in the future.

5. Conclusions and Future Work

This paper describes an evaluation of the adoption in practice of a measurement procedure for functional requirements specifications (RmFFP). This procedure is based on the COSMIC-FFP standard method, which is recognized by the ISBSG (International Software Benchmarking Standards Group).

This evaluation was carried out by applying a theoretical model that is based on a range of performance-based and perception-based variables; and the MEM relationships between them.

With respect to the *perception-based variables*, our findings were that RmFFP is perceived as easy to use and useful. Moreover, there is a clear intention to use in practice.

With respect to the *MEM relationships*, we noted that the perceived ease of use is influenced by the productivity of the persons using RmFFP. Intention to use RmFFP can be influenced more strongly by perceived usefulness than by perceived ease of use. We also noted that perceived usefulness is not determined by perceived ease of use. Furthermore, we found that reproducibility is not a significant influence on perceived usefulness.

Further research might include other variables that may help to improve the relationships that were not validated. We are aware that more experimentation with industry professionals is needed in order to reconfirm these initial results. In a future study, we plan to adjust the questions on the survey instrument used and verify the learning level for using RmFFP. In addition, RmFFP is currently being automated to be incorporated in the RETO tool[3].

References

[1] Abrahao S., Poels G., Pastor O. A Functional Size Measurement Method for Object-Oriented Conceptual Schemas: Design and Evaluation Issues. Software & System Modelling, Volume 5, Issue 1, Springer Verlag, 2005, pp. 48 -71.

[2] Azzouz S., Abran A., “A Proposed Measurement Role in the Rational Unified Process and its Implementation with ISO 19761: COSMIC-FFP” in Software Measurement European Forum, Rome, Italy, 2004.

[3] Basili V. R. and Rombach H. D., The TAME Project: Towards Improvement-Oriented Software Environments, IEEE Transactions on Software Engineering, 1988.

[4] Bévo V., Lévesque G., and Abran A. UML Notation for Functional Size Measurement Method. In Proc. 9th International Workshop on Software Measurement, Canada, September 8-10, 1999, pp. 230-242.

[5] Campbell D. T. and Fiske D. W., "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix", in Psychological Bulletin, vol. 56, 1959, pp. 81-105.

[6] Carver J., Jaccheri L., Morasca S., and Shull F., Issues in Using Students in Empirical Studies in Software Engineering Education, 9th International Software Metrics Symposium (Metrics'03), Sydney, Australia, 2003.

[7] Condori-Fernández N., Abrahao S., Pastor O., Towards a Functional Size Measure for Object-Oriented Systems from Requirements Specifications Abstract Functional Size Measurement. Quality Software International Conference, Germany, 2004, pp. 94-101.

[8] Condori-Fernández N., Abrahao S., Pastor O., El Problema de la Duplicidad de Movimientos de Datos en un Procedimiento de Medición, In Proc. of 9th Iberoamerican Workshop on Requirements Engineering and Software Environments, La Plata-Argentina, 2006.

[9] Davis F. D., "Perceived Usefulness, Perceived Ease of Use and User Acceptance of Information Technology", MIS Quarterly, vol. 3, no. 3, 1989.

[10] DOE Handbook, Alternative Systematic Approaches to Training, 1074-95, January 1995.

[11] Habela P., Glowacki E., Serafinski T., Adapting Use Case Model for COSMIC-FFP based Measurement, in the 15th International Workshop on Software Measurement, Montreal, Canada, Shaker-Verlag, 2005.

[12] Höst M., Regnell B., and Wohlin C., Using Students are Subjects – A Comparative Study of Students and Professionals in Lead-Time Impact Assessment, Empirical Software Engineering, pp. 201-214, 2000.

[13] I. Diaz, J. Sanchez, A. Matteo. Conceptual Modelling based on Transformation Linguistic Patterns, 24th International Conference on Conceptual Modelling, Klagenfurt-Austria, 2005.

[14] Insfran E., Pastor O. and Wieringa R., “Requirements Engineering-Based Conceptual Modelling”. Journal Requirements Engineering, Springer-Verlag, 2002.

[15] ISO, ISO/IEC 19761 Software Engineering-COSMIC-FFP-A Functional Size Measurement Method, International Organization for Standardization_ISO, Geneva, 2003.

[16] ISO, “ISO/IEC 14143-3: Information technology - Functional size measurement - Part 3: Verification of functional size measurement methods”, 2003.

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

[18] Moore G., Benbasat I., Development of an instrument to measure the perceptions of adopting an information technology innovation. Inform. Sytems, pp 192-222, 1991.

[19] Nunally J., Psychometric Theory, McGraw-Hill, 2nd ed., New York, NY1978.

[3] http://reto.dsic.upv.es/reto/

[20] Pastor O., Gomez J., Insfran E., Pelechano V., "The OO-Method approach for information systems modelling: from object-oriented conceptual modelling to automated programming", Information Systems 26, pp. 507-534, 2001.

[21] Rescher, N., Methodological Pragmatism: Systems-Theoretic Approach to the Theory of Knowledge, Oxford: Basili Blackwell, 1977.

[22] Venkatesh V., Davis D., A Theoretical Extension of the Technology acceptance Model: Four Longitudinal Field Studies, Management Science, Vol. 46, pp 186-204, 2000.

[23] Wohlin C., Runeson P., Höst M., M. C. Ohlsson, B. Regnell, and A. Wesslén, Experimentation in Software Engineering: An Introduction, 2000.

Appendix

Table A1. Data set used in the experiment

Subject	Size (Cfsu)	Measurement time (Hour)	Productivity (Cfsu/hour)	REP	PEOU	PU	ITU
1	144	1,08	133,33	0,11	4,80	3,60	4,33
2	160	1,48	108,11	0,01	3,80	3,40	2,67
3	153	1,15	133,04	0,05	4,80	4,00	4,33
4	142	1,50	94,67	0,12	3,80	4,00	3,67
5	159	1,50	106,00	0,01	3,80	3,00	3,33
6	154	1,25	123,20	0,05	4,60	3,20	3,67
7	155	1,00	155,00	0,04	5,00	3,80	4,33
8	154	1,45	106,21	0,05	4,00	4,40	4,33
9	149	1,20	124,17	0,08	4,20	4,40	4,33
10	156	1,17	133,33	0,03	4,80	4,40	3,33
11	160	1,58	101,27	0,01	3,20	3,20	3,33
12	168	1,20	140,00	0,04	4,60	4,00	4,33
13	161	1,50	107,33	0,00	3,60	4,20	4,00
14	181	1,55	116,77	0,13	3,60	3,80	3,67
15	162	1,05	154,29	0,01	4,60	3,40	3,67
16	159	1,15	138,26	0,01	4,60	4,20	3,00
17	176	1,60	110,00	0,10	3,20	4,00	3,33
18	166	1,15	144,35	0,03	4,40	4,20	4,00
19	184	1,00	184,00	0,15	4,60	4,40	4,00
20	180	1,45	124,14	0,12	3,80	4,60	3,67
21	175	1,00	175,00	0,09	5,00	4,40	4,67
22	148	1,00	148,00	0,08	4,60	4,00	3,67
23	152	1,25	121,60	0,06	4,20	3,80	3,33
24	160	1,50	106,67	0,01	3,60	3,80	4,33
25	149	1,30	114,62	0,08	4,20	4,00	4,00
26	166	1,25	132,80	0,03	4,20	4,20	4,00
27	158	0,83	190,36	0,02	4,60	3,60	3,00
28	168	1,33	126,32	0,04	4,40	4,00	4,00
29	164	1,00	164,00	0,02	4,40	4,40	4,00
30	168	1,45	115,86	0,04	3,80	3,40	4,00
31	166	1,05	158,10	0,03	5,00	3,80	4,67
32	161	1,17	137,61	0,00	4,60	4,00	4,00
33	157	1,42	110,56	0,03	3,60	4,40	4,33

Table A2. Correlation between Survey Items (Construct Validity)

		Perceived Ease of Use					Perceived Usefulness					Intention to Use			Mean		
		I1	I3	I4	I6	I9	I2	I5	I8	I10	I11	I7	I12	I13	CV	DV	VALID
PEOU	I1	1,00	0,52	0,45	0,61	0,31	0,17	0,26	-0,08	0,12	0,15	0,27	0,33	-0,02	0,58	0,15	YES
	I3	0,52	1,00	0,22	0,39	0,20	-0,03	0,06	-0,17	-0,12	0,06	0,11	0,25	0,10	0,46	0,03	YES
	I4	0,45	0,22	1,00	0,49	0,25	0,01	0,10	0,01	-0,12	0,32	0,19	0,22	0,09	0,48	0,10	YES
	I6	0,61	0,39	0,49	1,00	0,23	-0,06	-0,09	-0,29	-0,18	0,06	0,04	0,06	-0,03	0,54	-0,06	YES
	I9	0,31	0,20	0,25	0,23	1,00	0,04	0,22	0,13	0,27	0,20	0,25	0,15	0,18	0,40	0,18	YES
PU	I2	0,17	-0,03	0,01	-0,06	0,04	1,00	0,22	-0,09	-0,02	0,12	0,30	0,18	-0,02	0,25	0,07	YES
	I5	0,26	0,06	0,10	-0,09	0,22	0,22	1,00	0,26	0,26	0,44	0,21	0,00	0,05	0,44	0,10	YES
	I8	-0,08	-0,17	0,01	-0,29	0,13	-0,09	0,26	1,00	0,47	0,13	0,31	-0,15	0,38	0,35	0,02	YES
	I10	0,12	-0,12	-0,12	-0,18	0,27	-0,02	0,26	0,47	1,00	0,07	0,19	0,13	0,08	0,36	0,05	YES
	I11	0,15	0,06	0,32	0,06	0,20	0,12	0,44	0,13	0,07	1,00	0,09	0,21	0,26	0,35	0,17	YES
ITU	I7	0,27	0,11	0,19	0,04	0,25	0,30	0,21	0,31	0,19	0,09	1,00	0,22	0,43	0,55	0,20	YES
	I12	0,33	0,25	0,22	0,06	0,15	0,18	0,00	-0,15	0,13	0,21	0,22	1,00	0,02	0,41	0,14	YES
	I13	-0,02	0,10	0,09	-0,03	0,18	-0,02	0,05	0,38	0,08	0,26	0,43	0,02	1,00	0,49	0,11	YES