

12<sup>th</sup>  
INTERNATIONAL CONFERENCE ON  
**EASE2008**  
DEPARTMENT OF INFORMATICS - UNIVERSITY OF BARI - ITALY



26-27 JUNE 2008  
UNIVERSITY OF BARI  
DEPARTMENT OF INFORMATICS.

12<sup>th</sup>  
INTERNATIONAL CONFERENCE ON  
**EASE2008**  
Evaluation and Assessment in Software Engineering



EASE 2008

# 12th International Conference on Evaluation and Assessment in Software Engineering

## Contents

[Introduction](#)

[Short Summary](#)

[Detailed Summary](#)

[Programme Committee](#)

**Thursday June 26<sup>th</sup> 2008**

### [Impact of Experience and Team Size on the Quality of Scenarios for Architecture Evaluation](#)

Stefan Biffi<sup>1</sup>, Muhammad Ali Babar<sup>2</sup>, Dietmar Winkler<sup>1</sup>

<sup>1</sup>Vienna University of Technology, Austria

<sup>2</sup>Lero, University of Limerick, Ireland

### [An Evaluation of Function Point Counting Based on Measurement- Oriented Models](#)

Vieri del Bianco<sup>1</sup>, Claudio Gentile<sup>1</sup>, Luigi Lavazza<sup>1,2</sup>

<sup>1</sup>University of Insubria

<sup>2</sup>CEFRIEL

### [Maximizing data retention from the ISBSG repository](#)

Kefu Deng and Stephen G. MacDonell

School of Computing and Mathematical Sciences, Auckland University of Technology

### [Using Controlled Experiments to Assess Inspection Methods](#)

Andrea De Lucia<sup>1</sup>, Fausto Fasano<sup>1</sup>, Giuseppe Scanniello<sup>2</sup>, Genoveffa Tortora<sup>1</sup>

<sup>1</sup> Dipartimento di Matematica e Informatica, University of Salerno

<sup>2</sup> Dipartimento di Matematica e Informatica, University of Basilicata

### [Using a Protocol Template for Case Study Planning](#)

Pearl Brereton<sup>1</sup>, Barbara Kitchenham<sup>1</sup>, David Budgen<sup>2</sup>, Zhi Li<sup>1</sup>

<sup>1</sup>School of Computing and Mathematics, Keele University, Keele UK

<sup>2</sup>Department of Computer Science, Durham University, Durham, Science Laboratories,

### [Flexibility in Research Designs in Empirical Software Engineering](#)

Vigdis By Kampenes<sup>1,2</sup>, Bente Anda<sup>1,2</sup>, Tore Dybå<sup>1,3</sup>

<sup>1</sup>Department of Software Engineering, Simula Research Laboratory

<sup>2</sup>Department of Informatics, University of Oslo

<sup>3</sup>SINTEF ICT, Norway

### [Empirical WebGen; A Web- Based Environment for the Automatic Generation of Surveys and Experiments](#)

Francisco Novillo, Fèlix Garcia, Elvira Rolòn, Francisco Ruiz, Mario Piattini

Alarcos Research Group, Department of Information Technologies and Systems

Escuela Superior de Informática University of Castilla la Mancha, Ciudad Real, Spain

### **Systematic Mapping Studies in Software Engineering**

Kai Petersen<sup>1,2</sup>, Robert Feldt<sup>1</sup>, Shahid Mujtaba<sup>1,2</sup>, Michael Mattsson<sup>1</sup>

<sup>1</sup>School of Engineering, Blekinge Institute of Technology

<sup>2</sup>Ericsson AB,

### **A follow-up empirical evaluation of evidence based software engineering by undergraduate students**

Austen Rainer and Sarah Beecham

School of Computer Science, University of Hertfordshire U.K

### **Lessons from a cross domain investigation of empirical practices**

David Budgen<sup>1</sup>, John Bailey<sup>1</sup>, Mark Turner<sup>2</sup>, Barbara Kitchenham<sup>2</sup>, Pearl Brereton<sup>2</sup>, Stuart Charters<sup>3</sup>

<sup>1</sup> Department of Computer Science Durham University, U.K.

<sup>2</sup> School of Computing & Maths, Keele University, U.K.

<sup>3</sup> Applied Computing Group, Lincoln University, N.Z.

**Friday June 27<sup>th</sup> 2008**

### **Patterns of Evolution in the Practice of Distributed Software Development: Quantitative Results from a Systematic Review**

Rafael Prikladnicki<sup>1</sup>, Daniela Damian<sup>2</sup>, Jorge Lis Nicolas Audy<sup>1</sup>

<sup>1</sup> PUCRS, Brazil

<sup>2</sup> Univ. of Victoria, Canada

### **Lessons learnt Undertaking a Large-scale Systematic Literature Review**

Mark Turner<sup>1</sup>, Barbara Kitchenham<sup>1</sup>, David Budgen<sup>2</sup>, Pearl Brereton<sup>1</sup>

<sup>1</sup> School of Computing and Mathematics, Keele University, UK

<sup>2</sup> Department of Computer Science, Durham University, UK

### **Adopting Curvilinear Component Analysis to Improve Software Cost Estimation Accuracy Model, Application Strategy, and an Experimental Verification**

Salvatore A. Sarcia<sup>1</sup>, Giovanni Cantone<sup>1</sup> and Victor R. Basili<sup>2,3</sup>

<sup>1</sup>DISP, Università di Roma Tor Vergata, Rome, Italy

<sup>2</sup>Dept. of Computer Science, University of Maryland, USA

<sup>3</sup>Fraunhofer Center for Experimental Software Engineering Maryland, USA

### **A Preliminary Study of Sequence Effects in Judgment-based Software Development Work- Effort Estimation**

Stein Grimstad, Magne Jørgensen

Simula Research Laboratory, Norway

### **Investigating the Use of Chronological Splitting to Compare Software Cross- company and Single-company Effort Predictions**

Chris Lokan<sup>1</sup>, Emilia Mendes<sup>2</sup>

<sup>1</sup> School of IT&EE, Australia

<sup>2</sup> Computer Science Department, The University of Auckland, New Zealand

### **Evaluating Structural Criteria for Testing Mobile Code**

Márcio E. Delamaro<sup>1</sup>, Auri Marcelo Rizzo Vincenzi<sup>2</sup>, José Carlos Maldonado<sup>3</sup>

<sup>1</sup> Centro Universitário Eurípides de Marília (UNIVEM) São Paulo, Brazil

<sup>2</sup> Universidade Federal de Goiás (UFG) Goiás, Brazil

<sup>3</sup> Universidade de São Paulo (USP), São Paulo, Brazil

**Assessment of a Framework for Designing and Evaluating Security Sensitive Architecture**

Muhammad Ali Babar  
Lero, University of Limerick, Ireland

**A Correlational Study on Four Measures of Requirements Volatility**

Annabella Loconsole  
Department of Computing Science, Umea University, Sweden

## Introduction

The International Conference on Evaluation and Assessment in Software Engineering provides a forum for discussion and exchange of experiences among researchers who present their research results and discuss issues related to empirical and evaluation studies in a workshop-like atmosphere in which papers are presented and time is allowed for constructive discussion of their results and processes. The 12<sup>th</sup> edition (EASE 2008) has been hosted by the Department of Informatics at the University of Bari, Italy.

This year's papers have addressed topics including systematic literature reviews, case study planning and execution, formal experiments, effort estimation, lessons learnt, architecture evaluation and empirical software engineering.

We received 38 submissions from countries all over the world. Each paper was reviewed by three members of the Program Committee and 18 were accepted for presentation at the Conference and inclusion in the Proceedings.

In addition to research papers we have two keynote speeches, one for each day of the conference: the first by Prof. Roel Wieringa, Chair of Information Systems at the University of Twente Netherlands, is on the roles of empirical research in software engineering; the second, by Prof. Mario Piattini, full professor at the Information Systems and Technologies Department of the Universidad de Castilla-La Mancha Spain, is on the evaluation and assessment of UML models.

We want to thank all of those who have contributed to the set up and running of this year's conference: the authors for submitting their papers, the Program Committee members for their valuable work in reviewing and selecting the papers and in promoting the conference, the organizing committee together with all the people that helped in arranging the conference. We also would like to thank all the organizations that have sponsored the event and the Department of Informatics, University of Bari for hosting the event.

We finally thank all the attendees of the conference for making it a successful event, and hope you find the program interesting and enjoy your stay in Bari.

Giuseppe Visaggio, General Chair  
Maria Teresa Baldassarre & Steve Linkman, Program Chairs  
Mark Turner, Short Papers Chair



## Detailed Summary



The International Conference on Evaluation & Assessment in Software Engineering (EASE) offers a forum for discussion and exchange of experiences among researchers who present their research results and discuss issues related to empirical and evaluation studies. The conference aims to provide a workshop-like atmosphere in which papers can be presented and then time is allowed for constructive discussion of their results and processes. The 12<sup>th</sup> EASE conference in 2008 will provide an opportunity to present research results and a forum to discuss, consolidate and extend the existing research. We solicit contributions on the above themes, or general matters of interest to the community.

The 12<sup>th</sup> edition is hosted at the Department of Informatics of the University of Bari in Italy (<http://serlab.di.uniba.it/ease2008/>)

Research papers face topics related to:

- experiments (laboratory and field)
- replications of empirical studies
- case studies
- surveys
- observational studies
- field studies
- action research
- evaluation methodology
- systematic reviews and meta-analysis

In addition to research papers the conference will have two keynote speakers, one for each day of the conference: the first by Prof. Roel Wieringa, Chair of Information Systems at the University of Twente Netherlands, is on the roles of empirical research in software engineering; the second, by Prof. Mario Piattini, full professor at the Information Systems and Technologies Department of the Universidad de Castilla-La Mancha Spain, is on the evaluation and assessment of UML models.

## **EASE 2008**

### **Programme Committee**

Giuseppe Visaggio (chair)	University of Bari, Italy
Muhammad Ali Babar	Lero, Ireland
Teresa Baldassarre	Bari University, Italy
Anita Bente	Simula, Norway
Pearl Brereton	Keele University, UK
David Budgen	University of Durham, UK
Danilo Caivano	Bari University, Italy
Stuart Charters	Lincoln University, New Zealand
Steve Counsell	Brunel University, UK
Ray Dawson	Loughborough University, UK
Tore Dybå	Simula and SINTEF Norway
Marcela Genero	University of Castilla La Mancha, Spain
Tracy Hall	Brunel University, UK
Rachel Harrison	University of Reading, UK
Andreas Jedlitschka	IESE Fraunhofer, Germany
Magne Jørgensen	Simula, Norway
Barbara Kitchenham	Keele University, UK
Shari Lawrence Pfleeger	Rand Corporation, USA
Stephen Linkman	Keele University, UK
Emilia Mendes	University of Auckland, New Zealand
Matthias Müller	EnBW AG, Germany
Mahmood Niazi	Keele University, UK
Gordon Rugg	Keele University, UK
Per Runeson	Lund University, Sweden
Helen Sharp	The Open University, UK
Martin Shepperd	Brunel University, UK
Dag Sjøberg	University of Oslo, Norway
Mikael Svahnberg	Blekinge Institute of Technology, Sweden
Walter Tichy	University of Karlsruhe, Germany
Mark Turner	Keele University, UK
June Verner	National ICT Australia Ltd, Australia

# Empirical-WebGen; a Web-based Environment for the Automatic Generation of Surveys and Experiments

Francisco Novillo, Félix García, Elvira Rolón, Francisco Ruiz, Mario Piattini  
Alarcos Research Group, Department of Information Technologies and Systems  
Escuela Superior de Informática, University of Castilla-La Mancha  
Paseo de la Universidad, 4, 13071 Ciudad Real, Spain  
{francisco.novillo, felix.garcia, elvira.rolon, francisco.ruizg, mario.piattini}@uclm.es

Empirical studies in Software Engineering are essential for the validation of various methods, techniques, tools, etc. and human resources play a fundamental role in carrying out these studies successfully. These studies have captured the attention of the scientific community in Software Engineering over the last years, and are those which are most commonly used in controlled experiments, case studies and surveys. Traditionally, these studies have been undertaken in laboratories, bringing their experimental subjects together in one time and space. Due to the limitations caused by this during experimentation, the necessity of providing an open and distributed environment arises. Such an environment would permit the configuration of the parameters of an experiment or survey, depending on the characteristics of the people to whom they are directed and, furthermore resolving the problems of time and space thanks to the possibility of their being carried out via the Web. This paper describes the Empirical-WebGen software tool. This tool makes the design of surveys and experiments possible and supports both their later realization and the analysis of the results obtained. The tool has been used to carry out both a survey and an experiment on line which have allowed us to obtain some preliminary validation results.

*Keywords: on-line surveys and experiments, empirical tool, empirical software engineering.*

## 1. INTRODUCTION

The need to develop fail-safe and cost-effective computerized systems in the face of their increasing complexity has propelled researchers to advocate the use of empirical studies in software engineering [1-3]. There is a large folklore of failed, late and over budgeted software projects coupled with many well publicized software related disasters. Consequently the demand for improvements considerably outstrips supply [2] and the current competitive market which the software world has become is forcing companies to improve their quality. In many situations this search for quality requires the adoption of new technologies where no evidence about their practical usefulness exists. Other proposals are not considered despite their practical validity. In fact, one major problem of software engineering is that often a great diversity of methods, languages, tools, environments, etc. are proposed, without their usefulness having been demonstrated in practice [4]. Therefore, it is fundamental for company managers to adopt an evidence-based software engineering approach when making decisions if they want to adopt good practice more quickly and with fewer risks, improve the quality of products, and reduce the risk of project failures [5].

The strategies that we have at our disposal in empirical software engineering are the so-called empirical studies, these being the controlled experiments, case studies and surveys which are most commonly used. They allow us to obtain benefits which are centered not only upon research aspects but also upon the academic environment, and the human factor plays a very important role [4, 6, 7]. Traditionally, these studies have been undertaken in laboratories, bringing their experimental subjects together in one time and space. Obviously this issue marks certain limitations in empirical research. Although having personal and direct control over subjects is beneficial, more and more workers are telecommuting, and many organizations have offices in multiple locations, resulting in geographically dispersion [8]. By removing the collocation requirement, the benefits of carrying out surveys and experiments could be enjoyed by a wider audience.

Furthermore, it is a challenge to achieve realism with regard to experimental tasks, subjects and environment [9]. To achieve this realism, Sjöberg et al. propose running the experiments on real tasks, on real systems, with software professionals being representative of the target population of the technology, and using their usual development technology in their usual working environment [10].



In addition, replications that alter key attributes of the experiments are necessary to build up knowledge about whether the results hold under other conditions. Unfortunately, in software engineering, an excessive amount of studies tend to be isolated and are not replicated, either by the same researchers or by others [11]. Also, the automatic storage of such replications is very interesting both for homogeneous comparison and in the later analysis of the results. Furthermore, gathering data electronically has other advantages such as easy storage (paper forms are easy to lose, and bulky to store), fewer errors and less work (data entry introduces errors, takes time, and is tedious), and it is cheaper to send people a URL rather than a paper questionnaire.

It would consequently be very useful to have open and distributed tools at our disposal which would permit the configuration of the parameters of an experiment or survey, depending on the characteristics of the people to whom they are directed and, furthermore resolving the problem of time and space thanks to the possibility of their being carried out via the Web. A tool of this nature would permit the development of experiments in a more realistic environment in which professionals could perform the tasks by using their usual development technology in their usual working environment, and would facilitate empirical study replications, automatic results storage and the analysis of results. We would, therefore, obtain a useful empirical software engineering tool. This context has led to the creation of "**Empirical-WebGen**" as a web-based environment for the automatic generation of surveys and experiments.

This paper describes the characteristics of Empirical-WebGen and shows some significant results of the application of the tool to conduct surveys and experiments. The remainder of this paper is organized as follows. Section 2 gives an overview of related works. Section 3 presents the functionality and architecture of Empirical-WebGen. Section 4 describes the case studies in which the tool has been applied and Section 5 presents our conclusions, and outlines ongoing efforts and future work.

## 2. RELATED WORKS

Many tools are currently related, in one way or another, to the design of surveys or experiments. Some are centered exclusively on survey creation and others are eLearning platforms which have certain specific modules for surveys or exercises. Another group of tools is orientated towards the design of psychology experiments and, to the best of our knowledge, one tool, called SESE [12], is closely related to the current work, and permits the definition of experiments through which to evaluate software engineering technologies.

Among the tools used in survey creation, it is worth mentioning *Free Online Surveys* [13], *Zoomerang* [14] and *Surveyo Survey Software* [15]. The common features of all them are: multiple pages for the survey, e-mail invitation sending, optional image add-in for the survey, an acknowledgment end page, and individual and exportable results management. They differ mainly in the way in which they manage the results (reports, graphics, data exports...). The most significant tools in the eLearning and eWorking field are *Claroline* [16], *LRN* [17] and *Moodle* [18]. These are all open-source and freeware platforms which allow the design of online courses and the management of learning and Web collaboration activities. These tools have a great variety of modules (calendar, forums, task lists, FAQs, file storage, news, etc.), but those which are most closely related to our objectives are the design of surveys and exercises, and the results analysis. The *LRN* tool includes an added value with regard to survey tools, as it incorporates an upload file mode to solve exercises, which is an interesting consideration in on-line experiments. Another web-based tool able to support on-line empirical studies related to programming tasks is that of *Praktomat* [19], which presents tasks to users, and allows users to submit their solutions and to review these solutions on-line. With regard to the tools which permit the design of experiments in the field of psychology, it is worth mentioning *MediaLab* [20], *RiddleMeThis* [21] and *WebExp2* [22]. The new capabilities these tools introduce are task grouping, task randomizing and sound and video embedding. However, they do not permit result analysis.

Finally, it is important to highlight the SESE tool which was developed in Norway by the professional staff of KompetanseWeb AS in collaboration with the Simula Research Laboratory (SLR) [12]. This tool is that which is most closely tool related to our approach as it is a web-based environment for software engineering experiment design. The most important requirements that SESE covers are real-time monitoring of the experiment, flexibility of defining new kinds of questions and measurement scales, automatic recovery of experiment sessions, automatic backup of experiment data and multi-platform support.

In summary, upon analyzing the related works, we may conclude that these tools do not cover the requirements stated in the current work to support on-line empirical studies. Neither the survey creation tools nor the eLearning and eWorking platforms support the creation of experiments. The tools for the creation of psychological experiments match with our focus very closely; but they do not meet all the requirements, such as the possibility of reshaping images in the models, the upload exercises, the generation of reports, the creation of permissions for the

users, etc. The SESE tool incorporates many of the features which are desirable to the attainment of our goals, but additional features must be considered if our goals are to be accomplished, particularly with regard to experiments concerned with modeling languages and techniques, in which the management of images with diagrams, automatic control of time, randomization in the order of presentation of diagrams and questionnaires, the possibility of assign marks for evaluating tasks, and preliminary report analysis is necessary.

With the aim of supporting the previously mentioned requirements, a flexible and open environment, called Empirical-WebGen, has been developed.

### 3. EMPIRICAL-WEBGEN

Empirical-WebGen has been developed under the Microsoft .NET platform on client/server architecture. The integrated development environment (IDE) chosen has been Visual Studio .NET 2005. Users communicate with the application through a standard web-browser. The web pages have been built using HTML, CSS, ASP.NET 2.0 and AJAX under Framework 2.0. The application/business layer is implemented in VB.NET. The interface with the business layer uses ADO (ActiveX Data Object). All data is managed by the Microsoft SQL Server 2005 DBMS. Report generation is powered by the Visual Studio integrated version of Crystal Reports. Empirical-WebGen is a multi-language tool in which both English and Spanish interfaces are available.

With regard to functionality, Empirical-WebGen mainly supports the management of surveys and experiments by authorized administrators and authorized users, and the on-line experiments and surveys carried out by registered users. Report generation and permission management are also supported. These functionalities are illustrated in Figure 1:

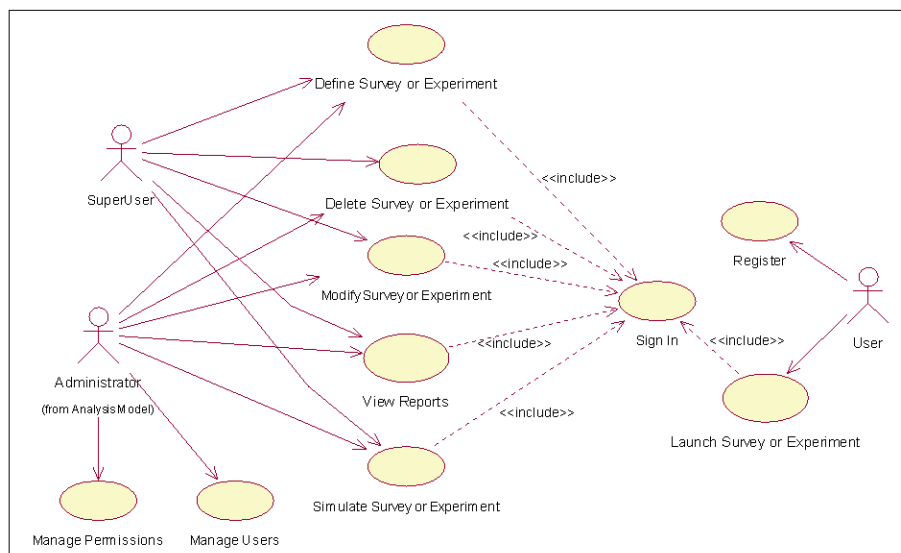


FIGURE 1: Empirical-WebGen use case main diagram

As can be observed in Figure 1, three roles are possible when interacting with the tool: the **Administrator**, who is the main role and can perform all the operations on surveys and experiments (definition, deletion, modification, reports visualization and simulation), along with permissions and users management; the **SuperUser**, who can perform all the functions related to surveys and experiments but who cannot manage permissions or users; the **User** who can perform a survey or an experiment. All of the roles must sign on before using the tool. The following subsections illustrate the typical process to be followed when using Empirical-WebGen and the support that the tool provides.

#### 3.1 Define Survey or Experiment

A survey or experiment is defined in Empirical-WebGen according to a generic structure we have defined to represent typical software engineering experiments and which is made up of the following elements (Figure 2):

- A generic **experiment** has a name and instructions and is made up of one or more models.

- **Models** are made up of one or more task groups and are shown on a web page with their image (optional) and their name. If so desired, it is possible to make the images disappear after a given time. This feature is useful for read-to-recall [23] tasks. Furthermore, a model can include a complexity assessment task to collect the subjective opinion with regard to the complexity of the model from users.
- **Tasks groups** are made up of one or more tasks. Every group has a name. The time taken by the user to perform the tasks included in the group is automatically stored by the system and is later used in the times report (see 3.4).
- **Tasks** have a statement and, depending on their type (Yes/No, True/False, Multiple Choice, Open Answer, Matrix or Exercise), can store the possible answers and the solution. Furthermore, it is possible to specify a positive mark when an answer is correct, and a negative mark in the opposite case. These marks can be used in the final evaluation of the subjects' correctness when solving the task. In fact, this feature permits the design of multiple choice tests through Empirical-WebGen and their automatic marking thanks to the marks report (see 3.4).
- Finally, tasks are associated with the **answers** given by users at a specific moment.

The order in which models, tasks groups and tasks appear may occur in the same order as that in which they were designed or may be random if this has been specified in the experiment design.

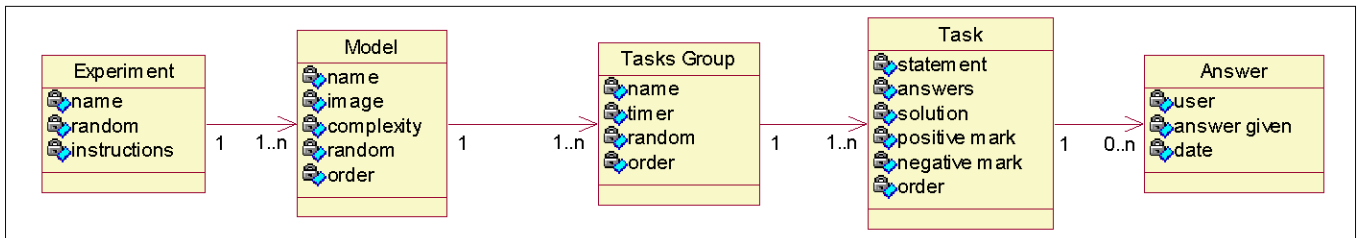


FIGURE 2: General experiment definition scheme

Following this structure, Empirical-WebGen provides an intuitive and usable interface for the design of surveys and experiments by showing one screen for each of the above items (Figures 3 and 4).

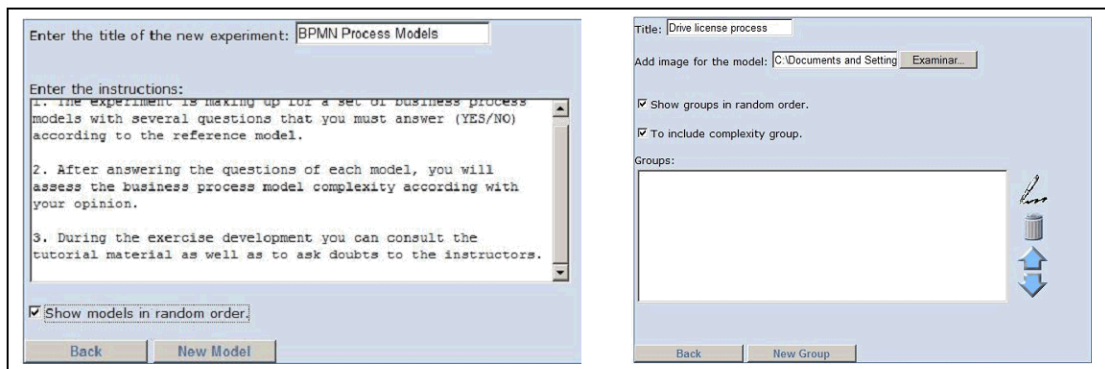


FIGURE 3: Experiment design (left) and Model design (right)

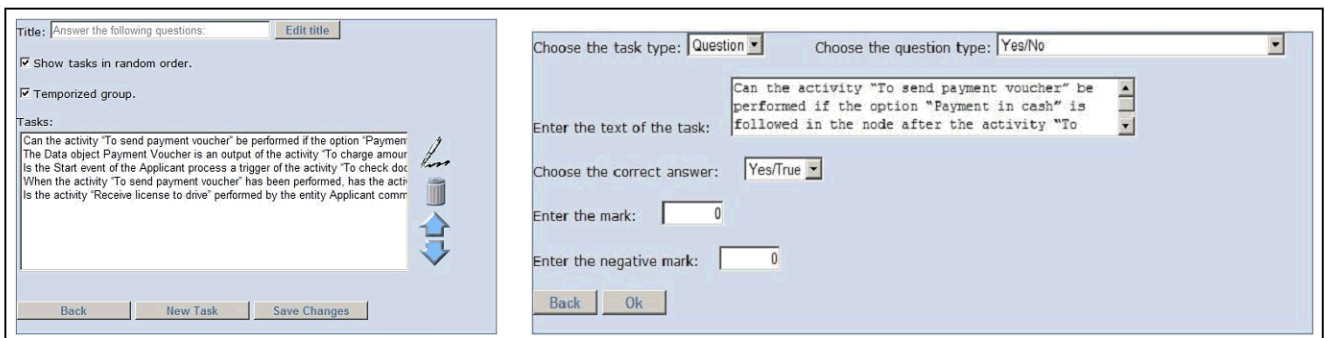


FIGURE 4: Group of tasks design (left) and Task design (right)

Empirical-WebGen supports the creation of two main kinds of tasks:

- **Exercises.** - These are made up of one statement and one optional downloadable file for users. Users can attach a solution file by means of a file upload control.
- **Questions.** - It is possible to create the following kind of questions: Yes/No, True/False, Multiple Choice: Choose one of the following, Multiple Choice: Choose all that apply, Open Answer and Matrix. In the first four cases, it is possible to choose the correct answer/s and the positive mark (in the case of success) and the negative mark (in the case of error) (see Figure 4, right). This is a significant feature in test development which can be used, for example, to create homogeneous subject groups by allocating people to one group or another depending upon their mark in a preliminary test previous to an experiment.

This set of tasks allows Empirical-WebGen to cover the majority of types of tasks according to the taxonomy proposed by Sjoberg et al. [24]. Document comprehension tasks (subcategory of Analyse) can be carried out, for example, by means of Yes/No, True/False or Multiple Choice questions. Modify Tasks can be performed through exercises in which users can download a file and attach it as a solution after its modification. The only requirement is that users must have the appropriate software in their computers. Create tasks can be carried out in a similar manner by attaching the file. Plan tasks are not currently supported by Empirical-WebGen because either interaction among subjects (project planning tasks) or the negotiation of software requirements with a customer (requirements analysis) are necessary.

### 3.2 Modify, Simulate and Delete a Survey of Experiment

The administrators (or superusers) can manage the surveys and/or experiments created, and can carry out various actions such as querying, editing, deleting, generating results reports (see Section 3.4) and simulation (Figure 5), which allows them to preview the experiment but without storing the results.

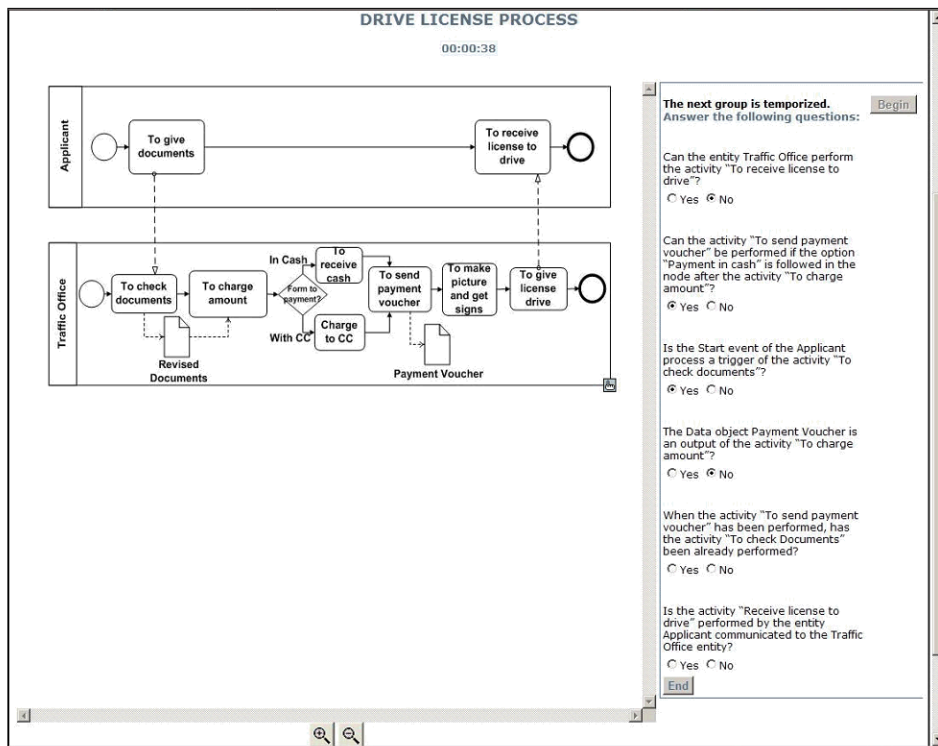


FIGURE 5: Simulation of Experiment

### 3.3 Users and Permissions Management

Once a survey or an experiment has been created, the administrator can create new permissions associated, on one hand with users and on the other hand with surveys or experiments, by means of **permissions managing**. Therefore, one user can carry out only surveys or experiments with the same permission. Furthermore, the administrator can give "superuser" permission to the users of his/her choice and these new "superusers" can then

create new surveys and experiments, modify, delete and simulate them, and view their results. With regard to **user managing**, the administrator can edit any of the registered users and modify any of their personal details, assign a previously created permission, and even remove the users of his/her choice from the system.

### 3.4 Reports Generation

One of the most important functionalities that Empirical-WebGen covers is that of report generation. Surveys and experiments in Software Engineering are usually carried out by a great number of people. Furthermore, experiments can be replicated several times. Therefore, it is fundamental to have a mechanism which is able to automatically store, display and summarize results. Empirical-WebGen supports the generation of four kinds of reports:

- **Times** - The average times that users have taken in answering group tasks are shown in this kind of report.
- **Answers** - The number of users that have chosen each of the different answers, along with their respective proportion over the total are shown.
- **Marks** - This report shows the marks obtained by each user in a test.
- **Success/Failure** - The number of successes and failures of users in each task along with their respective proportion over the total are shown.

Some examples of these reports can be seen in Figure 6.

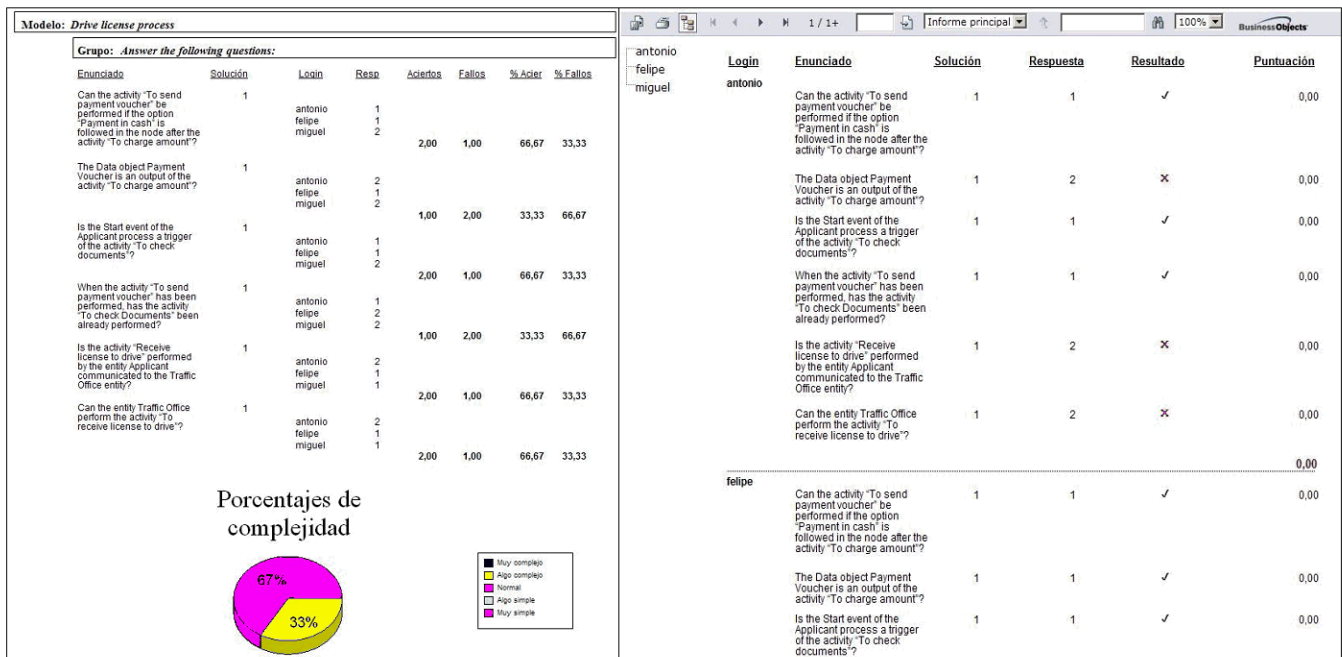


FIGURE 6: Reports of success/failure (left) and scores (right)

In the *times*, *answers* and *success/failure* reports, the administrator (or superusers) can choose between showing the summary of user details (see Figure 6, left) or otherwise. In the *marks* report, it is possible to choose between showing the summary of task details or not. In addition, it is possible to specify one or more of the following kinds of filters to show data:

- **Date filter** allows us to show the results of a survey or experiment, taking only users that have performed them between the dates and times specified.
- **Login filter** selects only users with login starting for the string typed into the text box.

- **IP filter** selects users that have taken part in the survey or experiment from a pc whose IP starts with the number typed into the attached text boxes.

These filter mechanisms enable us to select the results data of only the subjects who have participated in the context of a specific experiment or survey and distinguish them, for example, from subjects who have participated in a replica.

Once the report is produced, the top menu allows users to navigate among the different report pages, to change zoom, perform searches in the report, print, and even to export within Crystal Reports (RPT), Acrobat (PDF), MS Word, Ms Excel and RFT formats.

#### 4. CASE STUDIES

Empirical-WebGen has been applied in two real cases: survey creation, which concerns the assessment of university studies in computer science by professionals; and on-line experiment development to validate measures for BPMN notation (Business Process Modeling Notation) [25] models. The main results obtained are described in the following subsections.

##### 4.1 Survey of professional assessment in computer science university studies.

Certain members of the Alarcos research group were asked to carry out research on the assessment of computer science university studies by professionals in the IT sector in order to modify the current course programmes and fit it, on the one hand to The European Higher Education Area (EHEA)'s new system of courses and credit points [26, 27] and on the other, to what the profession is really asking for. In order to carry out this research, these Alarcos group members believed Empirical-WebGen to be an optimal solution for the performance of such research owing to its usability, to the possibility of online execution and to the generation of reports. The survey has been carried out by more than 120 professionals in the ITC sector who are distributed throughout Spain, both in universities and in IT companies, thus covering a large number of professionals of all kinds of career types. Fortunately, although many of them performed the survey at the same time, we did not experience problems related to network connections, increased network traffic or server load during these sessions. No problems related to tool usability were reported. This research thus assisted us in testing the survey generation functionality of the tool with successful results. The research is still open, and it is expected that many other professionals will carry out this survey by means of Empirical-WebGen.

##### 4.2 An Experiment to Validate Measures for Business Process Models

The Business Process Modeling Notation (BPMN) is a standardized graphical notation for drawing business processes in a workflow. BPMN was developed by the Business Process Management Initiative (BPMI), and is now being maintained by the Object Management Group since the two organizations merged in 2005. BPMN model measurement is the subject of study of Rolón et al. [28], in which the empirical validation of a set of measures for evaluating the usability and maintainability of BPMN models is tackled.

This experimentation has been used in the context of the present work to test the usefulness and usability of Empirical-WebGen when performing experiments. The original experiment required the paper-and-pencil solving of material made up of 15 BPMN models with different degrees of complexity (as evaluated by the proposed measures), and which included three "yes/no" questions about the understandability of the models and a question to subjectively evaluate the overall complexity of the model (from extremely difficult to extremely easy with 5 possible values). The order of models and questions was given to the different subjects randomly. 22 subjects participated in the experiment. All of them had a broad knowledge of the modeling of the product (UML, databases, etc) but they had little experience of the conceptual modeling of business processes, so a preparatory lesson was given before the experiment was carried out. To obtain a preliminary validation of the potential usefulness of the tool, a replica of this experiment with 9 subjects was conducted. The experimental design, material and subject background was the same, with the difference that the subjects took part in the experiment on-line by using Empirical-WebGen, and the preparatory session was replaced with a set of instructions and supporting documents that the subjects had to read in order to be prepared to do the experiment in conditions which were similar to those of the original subjects.

Table 1 summarizes the main results obtained with respect to the measures of the dependent variables: times, number of correct answers, efficiency (correct answers/time) and subjective assessment. To test whether the differences were statistically significant, the Mann-Whitney test was used ( $\alpha=0,05$ ).



	<b>T-STUDENT</b>				<b>MANN-WHITNEY</b>
	<b>PAPER-AND-PENCIL (P)</b>		<b>ONLINE (O)</b>		
	$\chi$	$\sigma$	$\chi$	$\sigma$	<b>Stat. Signif. <math>\alpha</math></b>
<b>TIMES (T)</b>	114,676	63,211	160,852	121,694	0
<b>CORRECT ANSWERS (CA)</b>	2,803	0,442	2,741	0,532	0,266
<b>EFFICIENCY (E)</b>	0,033	0,020	0,029	0,025	0
<b>SUBJECTIVE ASSESSMENT (SA)</b>	3 (median)	1,237	3 (median)	1,146	0,462

TABLE 1: Student's T analysis and Mann-Whitney Test

As we can see in Table 1, the average time is 46 seconds longer for the models in the online experiment than in the paper-and-pencil experiment. The main reason for these results may be due to the fact that the subjects who solved the experiment online had to spend some extra time in moving and resizing the biggest images which could not immediately be seen in their entirety. However this factor equally affected all the subjects who performed the online experiment. These results also seem to influence the efficiency measure results, as its value is inversely proportional to the time. With regards to the other two measures, correct answers and subjective assessment, no significant differences were found.

In order to analyze whether the use of the tool might affect to the results according to the objectives of the experiment a comparison analysis of the measures, which were validated in both experiments, was conducted. To analyze which measures were correlated with the understandability of the models in terms of times, correct answers, efficiency and subjective assessment, a Spearman's correlation analysis was performed. With regard to efficiency, the same set of measures were validated in both experiments. In relation to time and subjective assessment, all measures were correlated except one in both experiments. The only differences between the two experiments were found in relation to the number of correct answers in which from the 29 measures which were correlated in the paper and pencil experiment 14 were also validated in the replica. Further investigation is required to evaluate what the cause of these differences is.

Finally, to complete the analysis of Empirical-WebGen, the subjects who participated in the replica were asked about their impressions when using the tool in order to obtain their degree of satisfaction. These subjects were asked to rate the issues related to image manipulation (zoom controls, resize control and scroll bars), the page load time and their preference between using the tool or performing the experiment with paper-and-pencil.

The most repeated claim among the subjects was that they had to use the scroll bars for the biggest images. This motivated a preference for paper-and-pencil solving amongst some of them, but most of them (67%) showed their preference for the tool. On the other hand, 90% of users assessed the usability of the tool as normal, good or very good. Once these data had been analysed, we decided to solve the problem of image visualization by redistributing the elements shown on the Web and enlarging the image to its maximum size, which can, moreover, be adapted to the user's screen resolution. We thus succeeded in ensuring not only that any image which can be clearly seen on a standard sheet of paper could also be seen through Empirical-WebGen without manipulation on the user's part, but also that larger images could be shown thanks to the controls included in the tool, which constitutes an advantage with regard to paper-and-pencil experiments.

Another issue that was taken into account was that of external threats to the validity of the results [29]. Each subject carried out the experiment in their usual work station and might have been potentially interrupted, so we requested them to limit interruptions to times between each change of model. In addition, subjects were asked about the page load time and 90% of them said that it was optimal. This issue is highly important in discarding page load as a time validity threat. Furthermore, slow response times or interruptions caused by too much network traffic or server load might have threatened the results of the experiment. In particular, it might have caused the subjects to feel frustrated, which could in turn have affected their performance. Fortunately, we did not experience problems related to increased network traffic or server load during the experiment sessions, although new replicas must be conducted with larger samples in future to confirm this assertion.

## 5. CONCLUSIONS AND FUTURE WORK

This paper has presented Empirical-WebGen, a web tool with which to carry out surveys and experiments online which support the creation of a wide range of surveys with different kinds of questions and tasks, along with defining experiments exactly as they are understood in Software Engineering and even the creation of multiple choice tests. Taking the parameters stored in the database, the tool is able to generate automatically a web site in which users can take part in such a survey or experiment. A web tool like this makes surveys and experiments

designed in Empirical-WebGen available to any person at any moment anywhere in the world, with all the advantages that this implies. The electronic storage facilitates later analysis of the results and the comparison of different replications of the same experiment. Furthermore, in comparison with the traditional paper-and-pencil method, data losses and errors are avoided, there is no data entry time, and the distribution of surveys or experiments is cheaper. The automatic report generation functionality permits the analysis of preliminary results and presents the possibility of more exhaustive analysis by exporting reports in several formats. Also, when generating reports, the available filters allow us to distinguish between several user groups.

As a result of the application of the tool to two empirical studies, some preliminary observations have been obtained. The first case allowed us to test the functionality and usability of the tool when supporting on-line surveys, and no serious threats were detected. The experiment with BPMN allowed us to test certain significant aspects, namely:

- **Differences in experimental results.** In general no significant differences were found except in times and efficiency in models with larger images, which may be due to the extra time users needed to manipulate the images. Most of the measures were validated in both the experiment and replica with respect to times, efficiency and subjective rating. Very few differences were found with regard to the number of correct answers.
- **Performance.** The results of the survey and experiment reveal that the use of the tool did not have a negative impact on user performance as a result of network traffic or server overload.
- **Usability.** In general no significant problems were reported. We collected very useful feedback, especially regarding image display, which will be improved in the followings releases of Empirical-WebGen.

One important future work will be to collect new empirical data with larger samples of subjects to validate whether the use of the tool may affect experimental results. Additionally, new features are planned in order to support more exhaustive results analysis, for example the inclusion of certain statistical functions which will allow us to contrast the hypothesis suggested in the experiments. Finally, it will be interesting to provide the necessary functionality of the tool for real-time monitoring of surveys and experiments.

## 6. ACKNOWLEDGMENT

This research has been funded by the ESFINGE (Dirección General de Investigación del Ministerio de Educación y Ciencia, TIN2006-15175-C05-05) and INGENIO (Junta de Comunidades de Castilla-La Mancha, PAC08-0154-9262) projects.

## REFERENCES

- [1] Kallakuri, P. and Elbaum, S. (2001). *Empirical Studies in Empirical Software Engineering*, (<http://www.acm.org/crossroads/xrds7-4/empirical.html>)
- [2] Shepperd, M. (2003) Empirically-based Software Engineering. *UPGrade*. **4**(4), 37-41.
- [3] Anger, F.D. (1999) Directions for the NSF Software Engineering and Languages Program. *Software Engineering Notes*. *ACM SigSoft*. **24**(4), 52.
- [4] García, F., Serrano, M., Cruz-Lemus, J.A., Genero, M., Calero, C., and Piattini, M. (2004) La experimentación en la docencia de ingeniería del software. *X Jornadas de Enseñanza Universitaria de la Informática (JENUI'04)*, 237-245.
- [5] Dyba, T., Kitchenham, B., and Jorgensen, M. (2005) Evidence-Based Software Engineering for Practitioners. *IEEE Software*. 58-65.
- [6] Carver, J., Jaccheri, L., Morasca, S., and Shull, F. (2003) Issues in Using Students in Empirical Studies in Software Engineering Education. *Proceedings of the 9th International Software Metrics Symposium (METRICS'03)*, 239-251.
- [7] Baresi, L., Morasca, S., and Paolini, P. (2003) Estimating the Design Effort of web Applications. *Proceedings of the 9th International Software Metrics Symposium (METRICS'03)*, 62-71.
- [8] F. Hanks, B. (2004) Distributed pair programming: An empirical study. *Extreme Programming and Agile Methods - XP / Agile Universe 2004, Proceedings*. California, 81-91.
- [9] Harrison, W. (2000) An Alternative for Software Engineering Research?, Beg, Borrow, or Steal: Using Multidisciplinary Approaches in Empirical Software Engineering Research. *22nd International Conference on Software Engineering (ICSE)*. Limerick, Ireland, 5 June, 2000, 39-44.
- [10] Sjoberg, D.I.K., Anda, B., Arisholm, E., Dyba, T., Jorgensen, M., Karahasanovic, A., Koren, E.F., and Vokac, M. (2002) Conducting Realistic Experiments in Software Engineering. *ISESE' 2002*. Nara, Japan, 3-4 October 2002, 17-26.

- [11] Basili, V., Shull, F., and Lanubile, F. (1999) Building knowledge through families of experiments. *IEEE Transactions on Software Engineering*. **25**(4), 435-437.
- [12] Arisholm, E., Sjoberg, D.I.K., Carelius, G.J., and Lindsjorn, Y. (2002) A Web-based Support Environment for Software Engineering Experiments. *Nordic Journal of Computing*. **9**(4), 231-247.
- [13] Ltd., P.F. *Free Online Surveys*. (<http://freeonlinesurveys.com/>)
- [14] MarketTools, I. *Zoomerang*. (<http://info.zoomerang.com/>)
- [15] Software, S.S. *Surveyo*. (<http://www.surveyo.com/>)
- [16] Claroline, C. *Claroline*. (<http://www.claroline.net/>)
- [17] Consortium, L. *LRN*. (<http://dotlrn.org/>)
- [18] Dougiamas, M. *Moodle*. (<http://moodle.org/>)
- [19] Zeller, A. (2000) Making students read and review code. *ACM SIGCSE Bulletin*. **32**(3), 89-92.
- [20] Empirisoft. *MediaLab*. (<http://www.empirisoft.com/medialab.aspx>)
- [21] Loewald, T. *RiddleMeThis*. (<http://loewald.com/RiddleMeThis/>)
- [22] Mayo, N., Corley, M., Keller, F., Jaeger, F., and Mather, M. *WebExp2*. ([http://www.hcrc.ed.ac.uk/web\\_exp/](http://www.hcrc.ed.ac.uk/web_exp/))
- [23] Bodart, F., Patel, A., Sim, M., and Weber, R. (2001) Should optional properties be used in conceptual modelling? A theory and three empirical test. *Information Systems Research*. **12**(4), 384-405.
- [24] Sjoberg, D.I.K., Hannay, J.E., Hansen, O., By Kampenes, V., Karahasanovic, A., Liborg, N.-K., and Rekdal, A.C. (2005) A Survey of Controlled Experiments in Software Engineering. *IEEE Transactions on Software Engineering*. **31**(9), 733-753.
- [25] OMG. (2006) Business Process Modeling Notation (BPMN) Specification. OMG, (<http://www.bpmn.org>)
- [26] EHEA. *European Higher Education Area*, (<http://portal.unesco.org/education>)
- [27] ECTS. *European Credit Transfer System*, (<http://www.ects.es/>)
- [28] Rolón, E., García, F., Ruiz, F., and Piattini, M. (2007) An Exploratory Experiment to Validate Measures for Business Process Models. *First IEEE International Conference on Research Challenges in Information Science (RCIS'07)*. Ouarzazate, Marruecos, 23-26 april 2007, 271-280.
- [29] Anda, B. and Sjoberg, D.I.K. (2003) A Controlled Experiment Using a Professional Modelling Tool. *ISESE'2003*. 50-60.