ISESE'06

Proceedings of the

5th ACM-IEEE International Symposium on Empirical Software Engineering

Co-located with the 2006 Experimental Software Engineering International Week (ESEIW' 2006)

September 21-22, 2006 Rio de Janeiro, Brazil

Volume II: Short Papers and Posters

The printing of this publication is sponsored by FINEP

Promoted by COPPE UFRJ, UNIRIO

In cooperation with FINEP MCT FNDCT CT-PETRO, SBC Sociedade Brasileira de Computação

Sponsored by The Computer Society, SIGSOFT

Proceedings of the

5th ACM-IEEE International Symposium on Empirical Software Engineering

Co-located with the 2006 Experimental Software Engineering Week (ESEIW' 2006)

September 21 – 22, 2006 Rio de Janeiro, Brazil

Volume II: Short Papers and Posters

The printing of this publication is sponsored by FINEP

5th ACM-IEEE International Symposium on Empirical Software Engineering 2006

Short Papers Sessions: 3A, 3B and 3C

Session 3A: Empirical studies of software processes and products
Chair: Hakan Erdogmus *(NRC, Canada)*

Session 3B: Experiences with research methods and systematic reviews
Chair: Andreas Jedlitschka (*Fraunhofer Institute of Experimental Software Engineering, Germany*)

5th ACM-IEEE International Symposium on Empirical Software Engineering 2006

Session 3C: Evaluation and comparison of techniques and models
Chair: Emilia Mendes (*University of Auckland, New Zealand*)

Posters Session:

Improving the Experimentation for Evaluating the Effect of Composite States on the Understandability of UML Statechart Diagrams

José A. Cruz-Lemus, Marcela Genero, Mario Piattini
Department of Technologies and Information Systems
UCLM-Soluziona Research and Development Institute
University of Castilla-La Mancha
Paseo de la Universidad, 4
13071 Ciudad Real, Spain
+34 926 295300 ext. 3748

Sandro Morasca
Dipartimento di Scienze della Cultura, Politiche e dell'Informazione
Università degli Studi dell'Insubria
Como, Italy
Sandro.Morasca@uninsubria.it

{JoseAntonio.Cruz, Marcela.Genero, Mario.Piattini}@uclm.es

ABSTRACT
Understandability is one of the most important characteristics of UML diagrams. In previous works, we carried out some experiments to investigate the effect of composite states on UML statechart diagrams understandability. As the results of these empirical studies were quite contradictory, we reviewed the experimental process, tasks and materials. In this work, we have used a different approach to establish our work hypotheses and carry out a new experiment. The results we have obtained in this case add further evidence to the results of our previous studies and show that composite states in an UML statechart diagram may improve its understandability. Nevertheless, further replications of the experiment should be carried out in order to strengthen this assertion.

Categories and Subject Descriptors
D.2.2 [**Software** Engineering]: Design Tools and Techniques – *object-oriented design methods, state diagrams.*

General Terms
Measurement, Experimentation.

Keywords
UML statechart diagrams, Understandability, Metrics.

1. INTRODUCTION
UML has become the *de facto* standard for modeling software systems and UML statechart diagrams have become an important technique for describing the dynamic aspects of a software system [5]. In this framework, we have studied the effect that some UML constructs can have on the understandability of UML statechart diagrams [3], and paper addressed the effect of composite states on the understandability of UML statechart diagrams [4]. The results were not conclusive, and indicated that the benefits of using composite states depended on the previous experience of the subjects. We then planned a review of the experimental process that was carried out in that study, focusing especially in the design and the materials that were given to the subjects.

In this work, we present a new approach to the experimental tasks design based on some ideas outlined in [2, 6]. These works assert that models can be viewed as multimedia messages, as they include words and graphical elements. The Cognitive Theory of Multimedia Learning (CTML) by Mayer [8] helps explain how individuals viewing explanative material develop understanding of the multimedia content being presented to them. Based on CTML, we decided to change and extend part of our original design and materials to investigate understandability issues in UML statechart diagrams, and carried out a new empirical study.

In Section 2, we define our research question. The description of the experimental process is explained in Section 3. Section 4 describes the data analysis and the interpretation of the results. Finally, future work is presented in Section 5.

2. RESEARCH QUESTION
Our research question can be stated as *"Does the use of composite states improve the understandability of UML statechart diagrams?"*

To answer this research question, we have carried out a controlled experiment in which we took into account two variables used by the CTML: *transfer*, i.e., the ability to use knowledge gained from the material to solve related problems not directly answerable from it, and *retention* i.e., the comprehension of the material being presented. The CTML explains that learning occurs when these two variables are high, and this indicates that a high level of understanding of the presented material has been achieved [6]. We have also used another variable, *effectiveness*, defined as the capability for a correct understanding of the presented material.

We used the statistical test of hypotheses to test a set of experimental "alternative" hypotheses with our experiment. Specifically, we studied if using composite states improves the subjects' *effectiveness* (hypothesis H_{1a}), *transfer* (hypothesis H_{1b}), and *retention* (hypothesis H_{1c}). The corresponding "null" hypotheses are the logical negations of these alternative hypotheses.

3. EXPERIMENTAL PROCESS
We carried out the experimental process to test the hypotheses of section 2 based on the guidelines of [11].

3.1 Subjects
14 Computer Science PhD students of the University of Castilla-La Mancha in Spain voluntarily participated in this experiment. The tasks to be performed did not require high levels of industrial experience, so experiments with students could be considered appropriate [1, 7].

3.2 Experimental Design

The dependent variable was the understandability of UML statechart diagrams and we measured this through the previously introduced measures of *retention, transfer* and *effectiveness*.

The independent variables were the Universe of Discourse (UoD) to which the diagrams were related (an *ATM* and an *alarm clock*) and the use or not of composite states (CS) in the diagrams.

For each UoD, we gave the subjects two different diagrams with an identical semantic content, one with and one without composite states. The diagrams of each group were given to the subjects in different orders and the number of subjects in each group was balanced.

Each subject had to fulfill three tasks: the first one (Test A, about the *effectiveness* variable) was a test with 7 questions which were exactly the same for either UoD, independent of the usage of composite states. The questions inquired about navigation between states, variable values, etc. The subjects were allowed to check the diagram to answer the questions. This is a kind of task we had already used in previous studies [3, 4]. The second task (Test B, about the *transfer* variable) was a 5 questions test in which the subjects were asked about how the model worked, i.e., some more specific questions than in the previous test. The subjects were not allowed to use the diagrams to answer the questions. Finally, the third task (Test C, about the *retention* variable) was a 'fill-in-the-blanks' test. The subjects received a text in which the requirements of the model were commented but there were a number of missing words. The subjects had to fill in these blanks without using the diagrams.

The number of correct answers obtained divided by the number of questions, was used for measuring each dependent variable: CEffec (Correctness for *effectiveness*), CTrans (Correctness for *transfer*) and CReten (Correctness for *retention*).

3.3 Experimental Procedure

The experiment started with an introductory session in which we concisely explained the main motivation for the experiment as well as the elements of an UML statechart diagram. The goal of the experiment and the research question were not disclosed. Then we showed two examples in shortened version along with the correct answer to each question.

We divided the subjects into two groups at random. Then each subject received a diagram, depending on the subject's group, and its corresponding Test A sheet. From that moment, the subjects had 20 minutes to look at the diagram, understand how the model worked and answer the questions. Though the time allotted may appear to be quite short, this was done on purpose, as we wanted to investigate both the *transfer* and *retention* variables afterwards. After that time, these materials were collected and each subject received the sheets with the Tests B and C for the diagrams they had been studying. They had 20 minutes to work on both tests. The materials were collected again and the same process was repeated, i.e., first they received a diagram and a Test A sheet and later the tests B and C. In this second diagram, each subject received a different UoD than in the first one (ATM / clock) and also a different usage of composite states (with / without).

To alleviate possible learning effects, in the second part of the experiment the subjects were given 2 minutes less than in the first.

4. EXPERIMENTAL RESULTS

We used SPSS [10] to carry out the data analyses presented here.

4.1 Data Analysis

We first analyzed the descriptive statistics of the data. Table 1 shows the means and standard deviations for the different groups.

Table 1. Means and standard deviations (in parentheses) across groups

UoD	ATM		Alarm clock	
CS usage	With (n=7)	Without (n=7)	With (n=7)	Without (n=7)
CEffec	0.683673 (0.0761)	0.795918 (0.0687)	0.785714 (0.0643)	0.867347 (0.0453)
CTrans	0.642857 (0.0684)	0.634920 (0.0898)	0.690476 (0.0765)	0.577381 (0.0683)
CReten	0.607143 (0.0798)	0.614286 (0.0643)	0.571429 (0.1359)	0.444444 (0.1118)

We excluded several outlier values from the data analysis and proceeded to test the statistical hypotheses of Section 2. We refined our statistical hypotheses into three groups of hypotheses. Each group contained three null-hypotheses and was related to one of the original hypotheses presented in Section 2. As an example we present the group of hypotheses related to the hypothesis H_{1b}, which is about the *transfer*:

- H_{0b-1}: The UoD of the diagrams does not affect the mean value of the Correctness for *transfer*.
- H_{0b-2}: Using composite states in the diagrams (CS) does not affect the mean value of the Correctness for *transfer*.
- H_{0b-3}: The interaction between the UoD and the usage of composite states in the diagrams does not affect the mean value of the Correctness for *transfer*.

Therefore, we obtained 9 different null hypotheses (i=1,2,3): H_{0a-i}, related to the Correctness for *effectiveness* (CEffec); H_{0b-i}, related to the Correctness of *transfer* (CTrans); H_{0c-i}, related to the Correctness of *retention* (CReten). We set a significance threshold value $\alpha=0.05$. We performed an ANOVA, because this type of analysis allows us to analyze the interaction between the independent variables under study when the measurement of the dependent variable is repeated [9]. **Table 2** shows the obtained results. For each of the different test results it shows the value of the F statistic and the level of significance. For instance, testing the hypotheses related to the Correctness for *effectiveness*, we obtained a value F=2.221 for the effect of the UoD and the statistical significance was p=0.151.

Table 2. ANOVA results

	UoD	**CS**	**UoD * CS**
	F (sig.)	**F (sig.)**	**F (sig.)**

CEffec	2.221 (0.151)	4.870 (0.039)	0.047 (0.831)
CTrans	0.002 (0.961)	0.714 (0.407)	2.506 (0.128)
CReten	1.932 (0.179)	0.684 (0.418)	0.155 (0.697)

4.2 Interpretation

We now discuss the results we have obtained for the variables.

CEffec. The results agree with those obtained in previous works [4]. **Table 1** shows that the subjects obtained higher scores in those diagrams that are modeled without using composite states.

CTrans. We obtain completely different results than with the previous variable. The subjects obtained higher scores when they worked with diagrams that used composite states. However, the ANOVA results, shown in **Table 2**, have significance values above the α value, so we cannot reject any of the null hypotheses concerning to the Correctness for *transfer* variable.

CReten. As **Table 1** shows, the subjects obtained higher scores when facing the ATM diagram without composite states but with respect to the other diagram (the alarm clock) they got better results when using the diagram with composite states. In our opinion, this model was slightly more difficult and the difference between the means shown in **Table 1** is more distant here.

Thus, we may affirm that using composite states affects some aspects of the understanding of an UML statechart diagram, although the obtained results are not definitive.

4.3 Threats to validity

First, the statechart diagrams and tasks are not completely representative of real cases. Second, we used PhD students instead of professionals. However, the tasks to be performed did not require high levels of industrial experience, so using students for experimentation could be considered appropriate [1, 7]. So, more empirical studies using 'real cases' from software companies and possibly software professionals would be needed to strengthen the evidence we have obtained.

Besides, to alleviate possible effects of learning and fatigue, we changed the order in which treatment combinations were given to the subjects; furthermore, the subjects were assigned at random to each possible treatment order sequence. Finally, to minimize plagiarism, it was clear to the subjects that they were not going to be judged in any way based on the number of their correct answers and the number of subjects was limited.

5. FUTURE WORK

Further empirical studies need to be performed to acquire more evidence, so we have already planned to perform a replication of this experiment, which will take place at the Università degli Study dell'Insubria in Italy in the Spring semester 2006.

6. ACKNOWLEDGMENTS

This research is partially supported by the DYNAMICA/CALIPO project (TIC2003-07804-C05-03) financed by 'Dirección General de Investigación del Ministerio de Ciencia y Tecnología (Spain)' and the ENIGMAS project (PBI-05-058), supported by 'Consejería de Ciencia y Tecnología de la Junta de Comunidades de Castilla-La Mancha (Spain)', the IST project "QualiPSo", sponsored by the EU in the 6th FP, the FIRB project "ARTDECO," sponsored by the Italian Ministry of Education and University, and the project "La qualità nello sviluppo software", sponsored by the Università degli Studi dell'Insubria and the EU-funded ERASMUS exchange program.

The authors would like to deeply thank to all the PhD students that unselfishly helped us by performing the experiment.

7. REFERENCES

[1] Basili, V., Shull, F., and Lanubile, F. Building Knowledge through Families of Experiments. IEEE Transactions on Software Engineering, 25, (1999), 456-473.

[2] Bodart, F., Patel, A., Sim, M., and Weber, R. Should Optimal Properties Be Used in Conceptual Modelling? A Theory and Three Empirical Tests. Information Systems Research, 12, 4 (2001), 384-405.

[3] Cruz-Lemus, J.A., Genero, M., and Piattini, M., Metrics for UML Statechart Diagrams, in Metrics for Software Conceptual Models, Genero, M., Piattini, M., and Calero, C., Editors. 2005, Imperial College Press, UK.

[4] Cruz-Lemus, J.A., Genero, M., Manso, M.E., and Piattini, M. Evaluating the Effect of Composite States on the Understandability of UML Statechart Diagrams. In Proceedings of 8th International Conference on Model-Driven Engineering, Languages and Systems (MoDELS 2005) (Montego Bay, Jamaica, 2005). Lecture Notes in Computer Science 3713, 113-125.

[5] Denger, C. and Ciolkowski, M. High Quality Statecharts through Tailored, Perspective-Based Inspections. In Proceedings of 29th EUROMICRO Conference "New Waves in System Architecture" (Belek, Turkey, 2003). 316-325.

[6] Gemino, A. and Wand, Y. Complexity and Clarity in Conceptual Modeling: Comparison of Mandatory and Optional Properties. Data and Knowledge Engineering, 55, (2005), 301-326.

[7] Höst, M., Regnell, B., and Wohlin, C. Using Students as Subjects - a Comparative Study of Students & Professionals in Lead-Time Impact Assessment. In Proceedings of 4th Conference on Empirical Assessment & Evaluation in Software Engineering (EASE 2000) (Keele, UK, 2000). 201-214.

[8] Mayer, R.E., Multimedia Learning, Cambridge University Press, 2001.

[9] Reynoso, L., Genero, M., and Piattini, M., Measuring OCL Expressions: An approach based on Cognitive Techniques, in Metrics for Software Conceptual Models, Genero, P.a.C., Editor. 2005, Imperial College Press, UK.

[10] SPSS, SPSS 11.5, Syntax Reference Guide. 2002, SPSS Inc.: Chicago, USA.

[11] Wohlin, C., Runeson, P., Hast, M., Ohlsson, M.C., Regnell, B., and Wesslen, A., Experimentation in Software Engineering: an Introduction., Kluwer Academic Publisher, 2000.